THE IIOAB JOURNAL

VOLUME 7 : NO 7 : DECEMBER 2016 : ISSN 0976-3104

Institute of Integrative Omics and
Applied Biotechnology Journal

Dear Esteemed Readers, Authors, and Colleagues,

I hope this letter finds you in good health and high spirits. It is my distinct pleasure to address you as the Editor-in-Chief of Integrative Omics and Applied Biotechnology (IIOAB) Journal, a multidisciplinary scientific journal that has always placed a profound emphasis on nurturing the involvement of young scientists and championing the significance of an interdisciplinary approach.

At Integrative Omics and Applied Biotechnology (IIOAB) Journal, we firmly believe in the transformative power of science and innovation, and we recognize that it is the vigor and enthusiasm of young minds that often drive the most groundbreaking discoveries. We actively encourage students, early-career researchers, and scientists to submit their work and engage in meaningful discourse within the pages of our journal. We take pride in providing a platform for these emerging researchers to share their novel ideas and findings with the broader scientific community.

In today's rapidly evolving scientific landscape, it is increasingly evident that the challenges we face require a collaborative and interdisciplinary approach. The most complex problems demand a diverse set of perspectives and expertise. Integrative Omics and Applied Biotechnology (IIOAB) Journal has consistently promoted and celebrated this multidisciplinary ethos. We believe that by crossing traditional disciplinary boundaries, we can unlock new avenues for discovery, innovation, and progress. This philosophy has been at the heart of our journal's mission, and we remain dedicated to publishing research that exemplifies the power of interdisciplinary collaboration.

Our journal continues to serve as a hub for knowledge exchange, providing a platform for researchers from various fields to come together and share their insights, experiences, and research outcomes. The collaborative spirit within our community is truly inspiring, and I am immensely proud of the role that IIOAB journal plays in fostering such partnerships.

As we move forward, I encourage each and every one of you to continue supporting our mission. Whether you are a seasoned researcher, a young scientist embarking on your career, or a reader with a thirst for knowledge, your involvement in our journal is invaluable. By working together and embracing interdisciplinary perspectives, we can address the most pressing challenges facing humanity, from climate change and public health to technological advancements and social issues.

I would like to extend my gratitude to our authors, reviewers, editorial board members, and readers for their unwavering support. Your dedication is what makes IIOAB Journal the thriving scientific community it is today. Together, we will continue to explore the frontiers of knowledge and pioneer new approaches to solving the world's most complex problems.

Thank you for being a part of our journey, and for your commitment to advancing science through the pages of IIOAB Journal.

Yours sincerely,

Vasco Azevedo

***Vasco Azevedo**, Editor-in-Chief*
Integrative Omics and Applied Biotechnology
(IIOAB) Journal

ISSN: 0976-3104
REGULAR ISSUE

ARTICLE OPEN ACCESS

FIRST REPORT OF SEABUCKTHORN WILT CAUSED BY FUSARIUM SPOROTRICHOIDES IN INDIA

Nivedita Malik*

Department of Botany and Microbiology, S.G.R.R. (P.G.) College Constituent of S.G.R.R Education Mission, Dehradun-248001 Uttarakhand, INDIA

ABSTRACT

Hippophae salicifolia D.Don (seabuckthorn) is a deciduous tree species that yield highly nutrient- and antioxidant-rich fruits, the health protecting properties of which have been known since time immemorial. Seabuckthorn is also a good fodder, provides strong fencing material, has nitrogen fixing and soil binding properties, and possesses wide edaphic adaptation. The plant is restricted to the Himalayan region, between 1500-3500 m a.m.s.l. found in dry temperate forests of western Himalayas, sloppy areas near river banks and on sandy soil. Research on medicinal properties and other aspects of seabuckthorn has received much attention in recent past, but there is no information regarding pathogenic diseases of Hippophae salicifolia D.Don in India. During this study seabuckthorn wilt caused by Fusarium sporotrichoides was reported for the first time from Chamoli region of Uttarakhand Himalayas in India.

Received on: 22nd-Feb-2016
Revised on: 10th –April-2016
Accepted on: 1st -July, 2016
Published on: 2nd –July-2016

KEY WORDS

Antioxidant; Chamoli; Fusarium sporotrichoides; Himalayas; Uttarakhand.

***Corresponding author: Email:** maliknivedita8730@gmail.com; **Tel:** +91-9634676792

INTRODUCTION

Hippophae salicifolia D.Don commonly known as Seabuckthorn is a multipurpose, deciduous, dioecious thorny and nitrogen fixing shrub-tree growing widely on high altitude regions of Himachal Pradesh, Jammu & Kashmir, Sikkim and Uttarakhand. It is tolerant to extremes of temperature (-43 to +45^0 C), resistant to drought conditions and well adapted to the salinity and alkalinity [1,2]. It is supposed to be a store house of nutrients, vitamins and many items like jams, soft drinks, sauces, and pickles. In Indian Himalayan region, Seabuckthorn plant can offer benefits of nutrition, food, medicine, cosmetics etc. to the rural people for their socio-economic development. Seabuckthorn leaves are used for antioxidant and other properties. During a study on the occurrence of pathogenic diseases of *Hippophae salicifolia* inhabiting Garhwal Himalayas of India, Seabuckthorn wilt caused by *Fusarium sporotrichoides* was reported in this paper as a new record from India.

MATERIAL AND METHODS

In June 2015, samples of seabuckthorn plants (*Hippophae salicifolia* D.Don) showing wilt symptoms were collected during survey of naturally growing seabuckthorn populations from different locations of district Chamoli in Uttarakhand state viz. Rangad, Hanuman Chatti, Govindghat, Pandukeshwar, Badrinath, Niti, Mana. Infected plants were bagged, labelled and brought to the laboratory for further diagnosis and microscopic examination. Symptoms of the disease consisted of chlorosis, stunting, wilting and death. Isolation of the pathogen was made from infected tissues by performing moist chamber incubation method [3] and pure cultures of the isolated mycoflora were prepared using various culture media viz. Potato Dextrose Agar, Malt Yeast Agar and Czapek Dox Agar [4,5]. Identification of the isolated species was done by using standard literature and further confirmed from National Fungal Culture Collection of India-Agharkar Research Institute, Pune, India. Pathogenicity tests were conducted on twenty, 4 month old rooted cuttings under greenhouse conditions. Each plant was planted in a separate pot containing 0.7 litter of sterile soil. Inoculum for artificial infection was prepared with sterilized mixtures of wheat and barley seeds (10g of each). Seeds were inoculated with *F.sporotrichoides* spore suspension (10^6 conidia/ml) and inoculated at 220^0 C for 10 days. Non inoculated seeds served as controls. Ten seeds were placed under the soil surface around the root of each plant. Plants were irrigated and placed in a greenhouse (220^0 C and 12 hr day/night photoperiod).

RESULTS

The species was isolated from Rangad site and was identified as *Fusarium sporotrichoides* Sherbakoff, C.D. 1915 (NFCCI Accession No. 3672). Fusaria of potatoes. Memoirs of the Cornell University, Agricultural Experimental Station 6:87-270. Synonym=*Fusarium sporotrichiella* var. sporotrichoides (Sherb.) Bilais.

Morpho-taxonomic features

Colonies were initially white, but with age became red, and red pigments were produced in agar. Conidia at the tip of conidiophores which branch irregularly or dichotomously, non-, one, two or three septate; globose, ellipsoid, or pear shaped, often with a basal papilla of attachment, scattered dustily in the mycelium. Conidia in the sporodochia and pionnotes more or less three- to five- septate, spindle-sickle- shaped, with larger and smaller spores mixed, both narrow and thick, sometimes with parabolic curvature, both ends tapering, foot cell real or slight, in mass ochre to salmon or orange red **[Figure-1 and -2].**

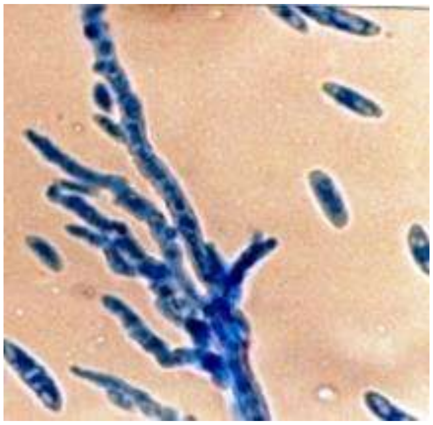


Fig:1. Macroscopic colony of *F. sporotrichoides*

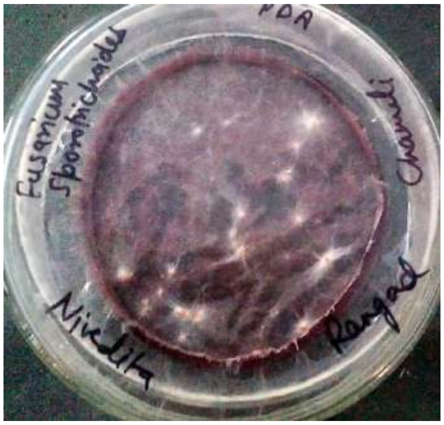


Fig:2. 10 days old colony of *F. sporotrichoides*

Pathogenicity test

Pathogenicity tests also confirmed the presence of *F.sporotrichoides*. After Sixteen days of inoculation, 80% of inoculated plants were wilted. Symptoms on infected plants were similar to those observed in the field. The pathogen was reisolated and confirmed from the infected leaves, thus fulfilling Koch's postulates. **[Figure- 3]**

Fig: 3. Potted plants of seabuckthorn under green house conditions

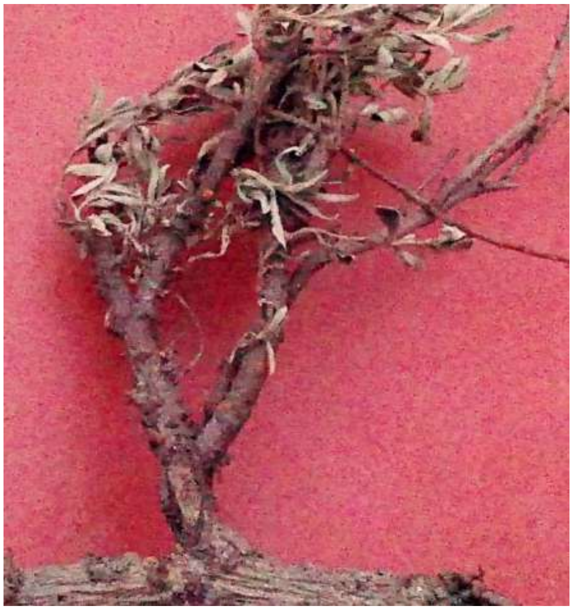


Fig: 4. Seabuckthorn showing Fusarium wilt

DISCUSSION

Diseases and insects/pests which affect almost every stage/part of the seabuckthorn are the factors affecting its cultivation. At present few pests and diseases of seabuckthorn have been reported; however more are likely to be identified as the number of plantations grow [6]. The major fungal disease reported on seabuckthorn includes verticillium wilt, fusarium wilt, damping off, brown rot, scab and dried shrink disease in China. The other common pathogenic fungi include the species of *Fusarium, Alternaria, Pythium, Fomes, Monilia, Stigmina hippophae* and *Valsa* [7]. 47 pathogen species were reported from Russia including *Fusarium sporotrichiella* causing maximum damage. In addition *Monilia altaica, Stigmina hippophae, Alternaria sp., Valsa sp. and Pythium* sp. have reported to cause damage in Russia and Liaoning, Shanxi and Gansu provinces of China [8]. Damping off can be caused by a number of soil-borne fungi such as *Alternaria, Fusarium and Botrytis*. *Fusarium* spp. were previously considered minor pathogens on sunflower [9] but currently in United States of America and Argentina, *Fusarium* is causing serious problems where damage upto 80% has been reported. Among the twelve *Fusarium* species identified in Russia*, F.oxysporum* var orthoceras was the most widespread, and F. sporotrichoides was the most aggressive [10]. In 2009, *F.sporotrichoides* and *F. acuminatum* were reported causing pink discoloration of the sunflower pith in addition to *F.oxysporum* in northern Great Plains [11]. *Fusarium sporotrichoides* has been reported to cause foliar spots on Forage Corn in Chile. Certain graminaceous plants such as Zea mays and *Triticum aestivum* serve as host for *Fusarium sporotrichoides* [12]. *Fusarium sporotrichoides* is a frequent pathogen in corn silage [13] and cereal crops [14, 15].

Very few reports are available regarding the pathological aspect of *Hippophae* spp.in India. Incidence of powdery mildew of Seabuckthorn was recorded in Himachal Pradesh [16]. Three fungal endophytes *Aspergillus niger*, *Mortierella minutissima* and a sterile mycelium and four species of VAM spores (*Glomus albidum, G. fasciculatum, G. macrocarpum and Gigaspora margariata*) have been isolated from different plant parts and soil samples [17]. Root rot caused by *Rhizoctonia solani* is major problem at nursery stage in Uttarakhand [18]. Thus to my knowledge there is no record of occurrence of genus Fusarium in association with *Hippophae* species in India and *Fusarium sporotrichoides* is being reported for the first time from leaves of *Hippophae salicifolia* D. Don causing seabuckthorn wilt.

CONCLUSION

The present study provides comprehensive information on pathological aspect of this wonder plant so that proper disease management of this multipurpose species could occur which favours the development and economic potential of seabuckthorn to improve socio economic status of the people residing in its natural habitat. The study will open up new horizon for local farmers and policy makers to develop effective action plan for sustainable use and conservation management of seabuckthorn in cold desert region in particular and Indian Himalayan region in general.

CONFLICT OF INTEREST

The author declares no competing interest in relation to the work.

ACKNOWLEDGEMENT

Author is thankful to Prof. V.A. Bourai, Principal, Shri Guru Ram Rai College, Dehradun; Uttarakhand, India.The Chairman Shri Guru Ram Rai Education Mission, Honourable Shri Mahant Devendra Dasji Maharaj for providing necessary facilities and Dr. S.K.Singh, Senior Scientist & Coordinator, National Fungal Culture Collection of India, Pune for providing confirmation of the result.

FINANCIAL DISCLOSURE

Nil

REFERENCES

[1] Jen Kumar V.[2003]-Sea buckthorn: A potential bio resource in Himalayas. *Invent. Intell* 159-167.

[2] Jodha NS, Banskota M, Pratap T. [1992]-Sustainable Mountain Agriculture: Perspective and Issue.Jodha N.S., Banskota M &Pratap T.(Eds.), Vol.1 Oxford and IBH Publishing Co. Ltd., New Delhi, pp: 38.

[3] Shutleff MC, Averre CW. [1997] The Plant Disease Clinic and field diagnosis of abiotic diseases,American Phytopathology Society, St.Paul MN.

[4] Agrios GN. [1997] Introductory Plant Pathology4thed. Academic Press,New York, NY.

[5] Waller JM, Ritchie BJ, Holderness M. [1998] Plant Clinic Handbook, CAB International New York, NY

[6] Kalia RK., Singh R, Rai MK., Mishra GP, Singh SR. &Dhawan AK. [2011] Biotechnological intervention in Sea buckthorn (Hippophae L.): Current status and future prospects; *Trees* 25:559-575.

[7] Li TSC. [2003] Taxonomy, natural distribution and Botany In: Sea buckthorn (Hipppohaerhamnoides L.): Production and Utilization. Li, TSC &BeveridgeTHJ (eds), PRC Research Press, *Ontario*, pp 7-12.

[8] Singh V. [2008]Seabuckthorn (Hippophae L.): *A multipurpose wonder plant*, 3: 166, 38. Daya Publishing House, New Delhi, India

[9] Gulya T, Rashid KY, Masirevic SM. [1997] Sunflower diseases: Phoma black stem in: Schmeiter, A.A. (Ed.), Sunflower Technology and Production. Agronomy Monographs no. 35 ASA, CSSA, SSSA, Madison, WI, USA, pp. 319-322.

[10] Orellana RG. [1971] Fusarium wilt of sunflower, Helianthus annus: First report. *Plant Disease* 55:1124-1125.

[11] Leslie JF, Summerell BA. [2006]The Fusarium: Laboratory Manual Blackwell Publishing Professional, Hoboken NJ P-256

[12] Tomoya A, Makoto K, Takumi N. [2012] The defence response in Arabidopsis thaliana against Fusarium sporotrichoides. *Proteome Science* 10:61.

[13] Baath H, Knabe O, Lepom P. [1990] Occurrence of Fusarium species and their mycotoxins in corn silage. 5. Fusarium infectation in corn silage. *Arch. Tiermahr* 40:397-405 (in German, English abstract)

[14] Leslie JF, Summerell BA.[2006]The Fusarium: Laboratory Manual Blackwell Publishing Professional, Hoboken NJ P-256

[15] Vargo RH, Baumer JS. [1986]Fusariumsporotrichioides as a pathogen of spring wheat. *Plant Disease* 70: 629-631.

[16] Bharat NK. [2006] Occurrence of powdery mildew on Sea buckthorn in Himachal Pradesh. *Indian Forester* 132(4):517

[17] Kumar S, Sagar A. [2007] Microbial associates of Hippophaerhamnoides (Sea buckthorn) *Plant Pathology Journal* 6(4):299-305.

[18] Singh K.P, Prasad & Yadav VK. [2007] The first report of Rhizoctoniasolanikuhn on Sea buckthorn (Hippophae salicifolia D.Don) in Uttaranchal Himalayas. *Journal Mycology PlantPathology* 37:126-127.

ABOUT AUTHOR

Ms.Nivedita Malik, *Research Scholar in S.G.R.R (PG) College, Dehradun, Uttarakhand, India, Affliated to Hemwati Nandan Bahuguna University, Srinagar, Garhwal (A Central University), Uttarakhand, India. Research Area- Mycology and Plant Pathology*

ISSN: 0976-3104
REGULAR ISSUE

ARTICLE OPEN ACCESS

ESTIMATION OF PEG TYPES AND THEIR CONCENTRATION DURING PROTEIN CRYSTALLIZATION

Rajneesh K. Gaur
Department of Biotechnology, Ministry of Science and Technology, 814, Block-2, C.G.O. Complex, Lodhi Road, New Delhi -110003, INDIA

ABSTRACT

Polyethylene Glycol (PEG) is a major precipitant in protein crystallization. This study focused on analytical estimation of different PEG types and their concentration on protein crystallization. The results indicate that ~84% of soluble proteins and ~78% of membrane proteins are crystallized with six PEG types (PEG3.35k>4k>8k>0.4k>6k>2kMME) and four PEG types (PEG3.35k>4k>6k>8k) respectively. The ~48% of soluble and ~62% of membrane proteins are crystallized with PEG3.35k & PEG4k only. Therefore, PEG4k may be used as an independent screening agent in PEG only commercial screens. Except PEG0.4k, remaining five PEG types contribute ~15-20% of soluble protein crystallization at 25% w/v concentration. The various classes of soluble protein i.e. All Alpha, All Beta and Alpha & Beta (a/b & a+b) does not show distinct preference for different PEG types. These results can be used to improve the PEG based protein crystallization commercial screens.

Received on: 27th-May-2016
Revised on: 08th-June-2016
Accepted on: 08th– June-2016
Published on: 08th– July-2016

KEY WORDS

Polyethylene Glycol (PEG) type, Polyethylene Glycol (PEG) concentration, Protein Crystallization

***Corresponding author: Email:** meetgaur@gmail.com; **Tel:** 011-24360295 **Fax:** 011-24360295

INTRODUCTION

Protein crystallization is a complex process, which is influenced by multiple parameters [1]. Empirical knowledge facilitated development of several screening methods to enhance the overall crystallization hits [2]. Efforts are continuing to improve the protein crystallization screens [3, 4]. Despite the remarkable progress made, protein crystallization is still a major bottleneck [5].

The most successful precipitants for protein crystallization are Polyethylene Glycol (PEG) and Ammonium Sulfate [6]. PEG is widely used in the crystallization of proteins, which are either crystallized alone or in complex. The various PEG related parameters such as different PEGs types, their concentration and molecular mass as well as influence of pH and salt concentration of the crystallizing solution have been studied for improving the crystallization process and introducing new commercial screens [7 ,8]. However, the influence of different PEG type and their concentration on crystallization of major group of proteins and their classes needs further exploration. With increasing number of structures available in the Protein Data Bank (PDB; till date ~98720 proteins structures solved through X-ray), the influence of different PEG types & concentration on protein crystallization can be easily ascertained.

METHODOS

Protein sequences having 30% sequence identity and crystallized with PEG are downloaded from Protein Data Bank (PDB) [9]. Out of the downloaded 9410 X-ray diffracted protein entries; two separate experimental dataset comprised of 1441 soluble protein and 43 membrane protein entries has been manually curated. The protein entries are curated after excluding the entries crystallized in complex with any type of ligand including protein/peptide/any chemical entity such as ATP etc. and also the entries possess inadequate and insufficient crystallographic information. The experimental dataset for soluble proteins is further divided for analytical purpose into four sub-datasets of 'All Alpha (423)', 'All Beta (323)', 'Alpha and Beta [a/b (165); a+b (530)]' proteins as per the Structural

Classification of Protein (SCOP) [10]. The percentage of proteins crystallized by different PEG types & their concentration profile was manually calculated and analyzed.

RESULTS AND DISCUSSION

PEG is known as one of the main precipitant for protein crystallization. PEG parameters directly influence the protein crystallization is determined either through data mining [11] or surveys [12]. Considering the tremendous growth of PDB database, two experimental datasets of protein (soluble and membrane) entries having 30% sequence identity and crystallized with PEG is prepared to determine the relationship between PEG types and their concentration on the crystallization of Proteins. The dataset of soluble proteins is further subdivided in to various classes of proteins to assess the occurrence of preference(s) towards PEG types and their concentration by any particular class. Considering the statistically less number of entries, the membrane protein dataset is not subdivided in to various classes. The 30% sequence identity criteria selected to incorporate only the unique sequences or sequences of least similarity in the experimental dataset [13]. The protein-protein complexes crystallized with PEG is not included in the dataset as they are studied earlier in detail [12]. Both the datasets were checked manually and found to possess the non-redundant crystallization conditions. In addition to PEG as a precipitant, ~20% of experimental dataset entries also possess Ammonium Sulfate or Sodium Chloride or MPD or combination of them as an additive.

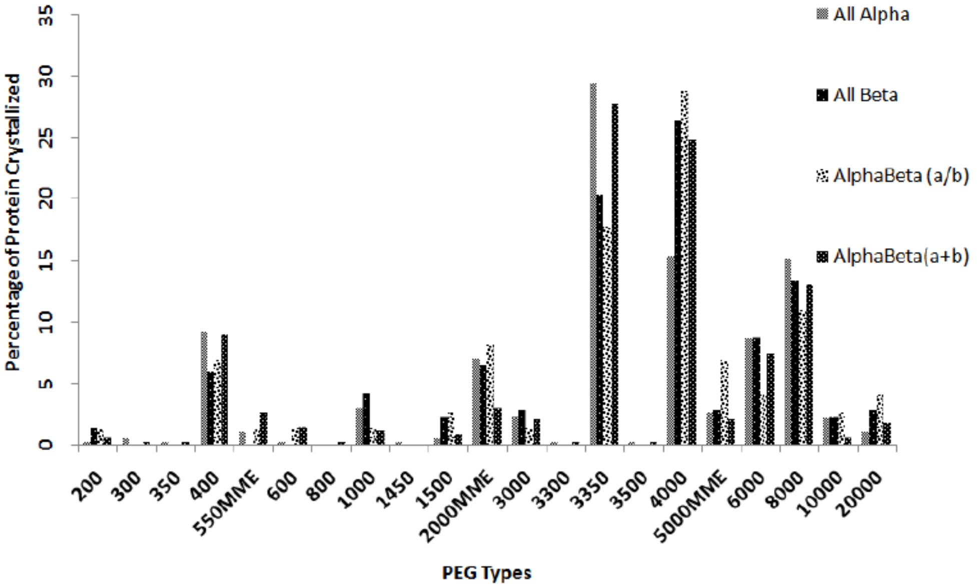


Fig: 1 shows the percentage of various classes of soluble proteins (alone) crystallized with different PEGs. The classes of protein are All Alpha (Grey), All Beta (Black) and Alpha & Beta (a/b - Starred & a+b – Black and White points)) types. The PEG types used based on the crystallizing conditions reported in PDB database.

In total, ~84% of soluble proteins are crystallized by just six PEG types [PEG3.35k (26.08%) > PEG4k (21.95%) > PEG8k (13.8%) > PEG0.4K (8.26%) > PEG6k (7.96%) > PEG2kMME (5.64%)] [**Figure-1]**. These results are in agreement with earlier studies, which shows that most of the proteins are crystallized with few PEG types mainly 3.35k, 4k, 6k & 8k [14, 15]. The results obtained here imply that only six PEG types substantially cover protein crystallization space. The percentage of soluble proteins (~62%) crystallized with three PEG types i.e. PEG3.35k, 4k & 8k shows that nearly double the percentage of proteins crystallized with these three PEG types in comparison to earlier studies. The difference is due to the inclusion of proteins having unique sequences or sequences of least similarity alone in this study. Out of the six PEG types, the two PEGs i.e. PEG3.35k and PEG4k contribute in the crystallization of ~48% soluble proteins. PEG3.35K used alone as a screening agent in commercial PEG screens such as PEG/ion screen (Hampton Research) and PEG suite (Qiagen). These results suggest that PEG4k, like PEG3.35k, can also be used independently as a screening agent for preliminary screening of proteins in commercial crystallization screens especially when the protein sample is precious. The commercial availability of an independent PEG3.35k & PEG4k screen will certainly enhance the overall efficiency of protein crystallization.

THE IIOAB JOURNAL ISSN: 0976-3104

In total, ~78% of the membrane proteins are crystallized by just four PEG types [PEG3.35k (35.1%) > PEG4k (27.0%) > PEG6k (8.1%) > PEG8k (8.1%)] [Figure-2]. Like soluble proteins, PEG3.35k and PEG4k also contributes remarkably in the crystallization of membrane proteins (~62%). These results are in contrast to a study where PEG4k is reported as most successful PEG precipitant [16]. The difference is due to the inclusion of all the membrane proteins in this study rather than only the outer membrane proteins. The crystallization of various membrane protein classes is not studied as a result of less number of entries.

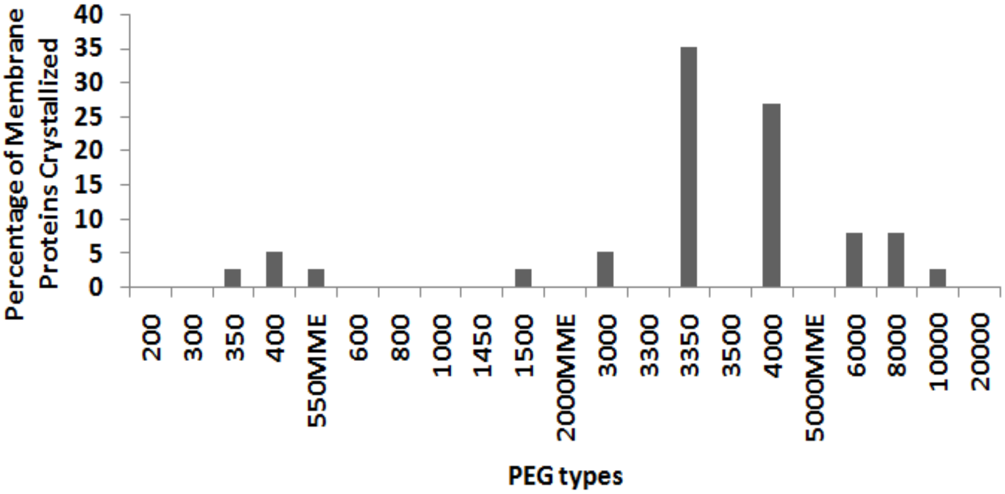


Fig: 2 Shows the percentage of membrane proteins (alone) crystallized with different PEGs. The PEG types used based on the crystallizing conditions reported in PDB database.

To determine the occurrences of preference(s) of various classes of soluble proteins towards six PEG types, the experimental dataset of soluble proteins is divided into sub-dataset of 'All Alpha, All Beta, Alpha and Beta (a/b and a+b) proteins as per the SCOP classification. It is revealed that the four classes of soluble proteins are crystallized with six PEG types with minor variation [Figure-1]. The PEG types i.e. PEG3.35k and PEG4k cumulatively leads to the crystallization of All Alpha (44.87%), All Beta (46.76%), Alpha and Beta [(a/b - 46.58%) & (a+b – 52.54%)] proteins, which is near to the total value (~48%) observed for the soluble proteins. 'Alpha and Beta (a/b)' proteins (17.81%) show ~9% (from the total) less percentage crystallization with PEG3.35k in comparison to other protein classes. Similarly, 'All Alpha' proteins (15.41%) show ~6% (from the total) less percentage crystallization with PEG4k in comparison to other protein classes. Furthermore, the percentage of 'Alpha and Beta (a/b)' protein crystallized with PEG6k is nearly half as compared to other protein types, while crystallized two and half times more with PEG5kMME. Similarly, the percentage of 'Alpha and Beta (a+b)' proteins crystallized with PEG2kMME is nearly one third as compared to 'Alpha and Beta (a/b)' protein type. These differences can be attributed to the influence of secondary structure composition on protein packing [17].

The data is further analyzed to determine the concentration profile of six PEG types resulted in the highest percentage of soluble protein crystallization [Figure-3]. The concentration profile of six PEG types shows distinct concentration peaks. PEG0.4K shows peaks at 2% & 35% v/v (31.88% of total protein crystallized); PEG2k at 15%, 20%, 25% & 30% w/v (49.12% of total protein crystallized); PEG3.35k at 20% & 25% w/v (60.31% of total protein crystallized); PEG4k at 25%, 30% & 35% w/v (43.81% of total protein crystallized); PEG6k at 15% & 25% (46.15% of total protein crystallized) and PEG8k at 25% w/v (21.04% of total protein crystallized) respectively. Similar studies regarding the PEG concentration and their relative frequency are also carried out by Peat etal. 2005 [11]. The 'relative frequency' was the count of the number of occurrences of a PEG at a given concentration divided by the total number of conditions containing that PEG. Their results indicate that the Low molecular weight (LMW) PEGs show no clear preferred concentration. However, our results clearly show a sharp peak for percentage of proteins crystallized at 2% v/v for PEG0.4k, which suggests that PEG0.4k acts as a precipitant as well as an additive at low concentration. Further, these results justify the usage of PEG0.4k in the concentration range of 30-35% in commercial screens. Peat et al., 2005 [11] reported that Medium Molecular Weight (MMW) PEGs (1k-8k) shows their concentration peak between 20-25%, however, our results shows wider range for MMW PEGs. In our study, it is observed that MMW PEGs facilitate crystallization of ~15-20% of protein at 25% w/v concentration.

ISSN: 0976-3104

Furthermore, the four PEG types also shows other major concentration (% w/v) peaks i.e. PEG2kMME (15%, 20% & 30%), PEG3.35k (20%), PEG4k (30% & 35%), PEG6k (15%). The doubling of protein crystallization percentage with PEG2k from 20% to 25% w/v is due to higher volume exclusion effect leading to more protein aggregation. The concentration profile for membrane proteins is not mapped due to less number of entries in the dataset. PEG3.35k & 4k facilitate at 20% & 30% w/v concentration the crystallization of Membrane proteins.

The occurrence of few concentration peaks for six PEG types complies with the concentration range found in most of the commercial kits. However, currently none of the commercially available PEG based screens possess this combination of six PEG types and the concentrations revealed out of this study. PEG suite (Qiagen) is the only screen using various PEG types at 25% w/v concentration, as observed in our results. These results indicates that five PEGs (i.e.2k, 3.35k, 4k, 6k & 8k) can be incorporated at the concentrations, observed out of this analysis, in commercial screens for augmenting the chances of crystallization.

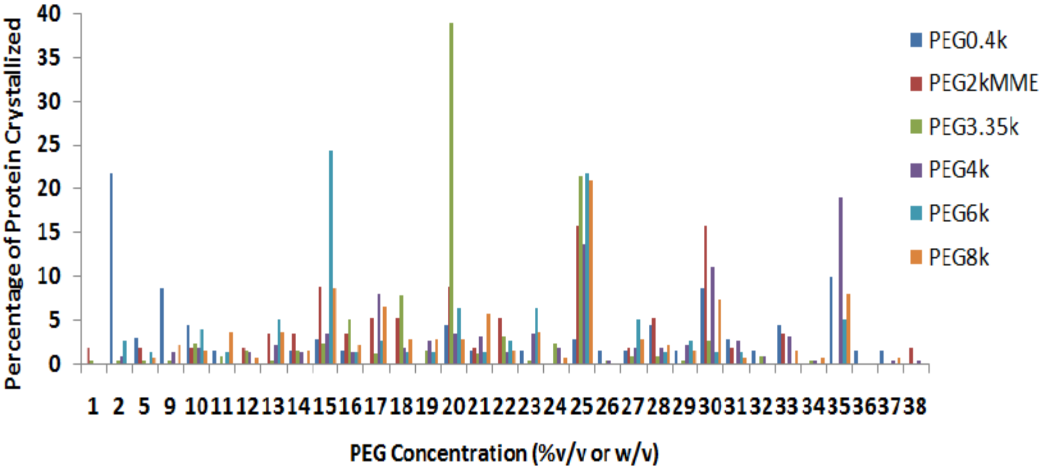


Fig:3 shows the percentage of different PEG concentrations and percentage of proteins crystallized. The six PEG types, whose concentration plotted, are PEG0.4k (dark blue), PEG2kMME (Red), PEG3.35k (Green), PEG4k (violet), PEG6k (light blue) and PEG8k (organge).

……

In this study, no clear correlation observed between the concentration profile and molecular weight of different PEGs as opposed to a generalized relationship of decreasing the median PEG concentration with increasing molecular weight of PEGs as reported for protein-protein complexes [12]. This type of diffused pattern may be due to the effect of multiple parameters such as viscosity, interplay of attractive and repulsive forces, ionic strength etc. on protein crystallization [18].

SYNOPSIS

Medium molecular weight (2k-8k) PEG types contribute significant percentage of protein crystallization. A major fraction (~50-60%) of proteins is crystallized with only PEG3.35k & PEG4k types. The results obtained can be used to improve the commercial crystallization screens.

CONCLUSION

In conclusion, a similar study as carried out earlier by Fazio et al., 2014 [15] is desired to evaluate the success of PEG based screens in crystallization space to validate the results obtained. Alternatively, a PEG screen prepared on the basis of the results obtained in this study may be used in high throughput protein crystallization for validating the success of a new recipe for improving the efficiency of protein crystallization.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENT

None.

FINANCIAL DISCLOSURE

No financial support was received for this study.

REFERENCES

[1] Kluser Rosenberger F. [1996] Protein Crystallization *J Cryst Growth* 166: 40-54.

[2] Shieh HS, Stallings WC, Stevens AM, Stegeman RA. [1995] Using sampling techniques in protein crystallization Acta Cryst D51: 305-310.

[3] Rupp B, Wang J. [2004] Predictive models for protein crystallization Methods 34: 390-407.

[4] Gorrec F. [2009] The MORPHEUS protein crystallization screen *J Appl Cryst* 42: 1035-1042.

[5] Luft JR, Newman J, Snell EH. [2014]. Crystallization screening: the influence of history on current practice Acta Cryst F, *Struc Biol Commun* 70: 835-853.

[6] Dumetz AC, Chockla AM, Kaler EW, Lenhoff AM. [2009] Comparative Effects of Salt, Organic, and Polymer Precipitants on Protein Phase Behavior and Implications for Vapor Diffusion Growth Des 9: 682-691.

[7] Page R, Stevens RC. [2004] Crystallization data mining in structural genomics: using positive and negative results to optimize protein crystallization screens Methods 34: 373-389.

[8] Newman J, Egan D, Walter TS, Meged R, Berry I, Jelloul MB, Sussman JE, Stuart DI, Perrakis A. [2005] Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy Acta Cryst D61: 1426-1431.

[9] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. [2000] The protein Data *Bank Nucleic Acids Res* 28: 235–242.

[10] Murzin AG, Brenner SE, Hubbard T, Chothia C. [1995] SCOP: a structural classification of proteins database for the investigation of sequences and structures *J Mol Biol* 247: 536-540.

[11] Peat TS, Christopher JA, Newman J. [2005] Tapping the Protein Data Bank for crystallization Information Acta Cryst D61: 1662-1669.

[12] Radaev S, Sean Li, Sun PD. [2006] A survey of protein-protein complex crystallizations Acta Cryst D 62: 605-612.

[13] Pearson WR. [2013] An Introduction to Sequence Similarity ("Homology") Searching Current Protocol in Bioinformatics, 42:3.1.1-3.1.8.

[14] Li C, Kirkwood KL, Brayer GD. [2007] The Biological Crystallization Resource: Facilitating Knowledge-Based Protein Crystallizations Crystal Growth Des 7: 2147-2152.

[15] Fazio VJ, Peat TS, Newman J. [2014] A drunken search in crystallization space Acta Cryst F 70: 1303-1311.

[16] Newstead S, Hobbs J, Jordan D, Carpenter EP, Iwata S. [2008] Insights into outer membrane protein crystallisation Mol Membr Biol 25: 631-638.

[17] Fleming PJ, Richards FM. [2000] Protein packing: dependence on protein size, secondary structure and amino acid composition ,*J Mol Biol* 299: 487-498.

[18] Kozer N, Kuttner YY, Haran G, Schreiber G. [2007] Protein-Protein Association in Polymer Solutions: From Dilute to Semidilute to Concentrated *Biophysical J* 92: 2139-2149.

ABOUT AUTHOR

__Dr. Rajneesh Kumar Gaur__ is currently associated with Department of Biotechnology, Ministry of Science and Technology, New Delhi, India. His interest is the analyses of biological data available in public domain databases.

ISSN: 0976-3104
REGULAR ISSUE

CASE STUDY OPEN ACCESS

OCCURRENCE OF SCHISTOSOMA NASALE INFECTION IN CROSSBRED CATTLE: A CASE STUDY

Kausar Qadri[1] and Subha Ganguly[2]*

[1]Department of Veterinary Medicine, Arawali Veterinary College, Rajasthan University of Veterinary and Animal Sciences, Bikaner, INDIA

[2]Department of Microbiology, Arawali Veterinary College, Rajasthan University of Veterinary and Animal Sciences, Bikaner, INDIA

ABSTRACT

A 8 years old cross bred Holstein Friesian (H.F.) cow was presented at Arawali Veterinary College, Sikar with history of sneezing, ocular discharge, nasal discharge, difficulty in breathing with snoring sound, there were presence of cauliflower like growth on one side of the nasal septum. The disease was diagnosed as nasal schistosomiasis and was successfully treated with Anthiomaline (Lithium Antimony Thiomalate) @ 15 ml intramuscularly three doses at weekly intervals.

Received on: 03rd -June-2016
Revised on: 15th-June-2016
Accepted on: 20th-June-2016
Published on: 10th-July-2016

KEY WORDS

Schistosoma nasale, cross H.F cow, Anthiomaline

***Corresponding author:** Email: ganguly38@gmail.com;**Tel:** +91 9231812539

INTRODUCTION

Nasal schistosomiasis is caused by the blood fluke Schistosoma nasale adversely affects the health and production of domestic livestock in various parts of India. S. nasale is a species of digenetic trematode in the family Schistosomatidae. It was identified first in 1933 by Dr. M. Anant Narayanan Rao at Madras Veterinary College, Tamil Nadu, India. The freshwater snail Indoplanorbis exustus acts as intermediate host [1]. Affected cattle shows rhinitis, profuse mucopurulent nasal discharge manifested clinically by sneezing, dyspnoea and snoring. Chronic infections show proliferation of nasal epithelium as granuloma and small abscesses containing eggs.

CASE HISTORY AND OBSERVATION

A 8 years old cross bred Holstein Friesian cow was presented at the Teaching Veterinary Clinical Complex (T.V.C.C.) of Arawali Veterinary College, Sikar, Rajasthan with history of sneezing, ocular discharge, nasal discharge, difficulty in breathing with snoring sound. On clinical examination it was found that there was presence of cauliflower like growth on one side of the nasal septum. Laboratory examination of nasal mucous reveals the presence of boomerang shaped eggs of Schistosoma nasalis. On the basis of history, clinical signs and laboratory examination, disease was confirmed as Nasal Schistosimiasis.

TREATMENT AND DISCUSSION

Treatment started with Inj Anthiomaline (Lithium Antimony Thiomalate) @ 15 ml intramuscularly. The cow responded after first dose of Anthiomaline and there was reduction in the size of nasal granuloma. The Inj Anthiomaline @ 15 ml intramuscular was repeated after weekly interval. Further reduction in the size of nasal granuloma was recorded and snoring sound was also reduced to slight sound that was audible, animal was breathing normally. Then the third injection of Anthiomaline was given after one week, reports complete recovery of the animal. Thus complete recovery of the animal was reported with three doses of Anthiomaline at weekly intervals.

Anthiomaline was the drug of choice.[2] Physical examination of animals revealed sneezing, thick mucus nasal discharge, congestion of nasal mucosa. The above findings are in conformation with Soulsby [3]. Antimony attached itself to sulphur atoms in trypanothione reductase (the putative enzyme targeted by antimonial compounds) which was used by the parasites. High incidence of schistosomiasis was seen in older animals as reported by Banerjee and Agrawal [4] and Sumanth et al. [5].

CONCLUSION

The present study reports on the successful treatment of clinical case of nasal schistosomiasis with the administration of three consecutive doses of Inj Anthiomaline (Lithium Antimony Thiomalate) @ 15 ml intramuscularly at regular weekly intervals.

.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENT

The authors acknowledge Hon’ble Dean and Management (Hony. Chairman and Secretray, Aastha Society, Sikar) of Arawali Veterinary College, Sikar for providing the necessary facilities to carry out this clinical case study.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] Kluser Liu L. [2010] The phylogeography of Indoplanorbis exustus (Gastropoda: Planorbidae) in Asia. Parasites Vectors, 3:57–58.

[2] Agrawal MC and Alwar VS. [1992] Nasal schistosomiasis : A review. Helminthological Abstract, 61: 373–384.

[3] Soulsby EJL. [1982] Helminths, arthropods and protozoa of domesticated animals. 7. London: Bailliere Tindall Ltd.

[4] Banerjee PS, Agrawal MC. [1992] Epizootiological studies in bovines on fluke infections with special reference to schistosomiasis. *Indian Vet. J* 69:215–220.

[5] Sumanth S, D’Souza PE and Jagannath MS. [2004] A study of nasal and visceral schistosomiasis in cattle slaughtered at an abattoir in Bangalore, South India. Rev. *Sci. Tech. Off. Int. Epiz.,* 23(3):937–942.

ISSN: 0976-3104
REGULAR ISSUE

THE IIOAB JOURNAL
ARTICLE OPEN ACCESS

INNOVATIVE GEN PRACTICE ANALYSIS TOWARDS CULPABILITY RECOGNITION IN EARTHING STRUCTURE

T. Yuvaraja* and K. Ramya

Department of EEE, Sri Sai Ram College of Engineering, Bangalore City, INDIA

ABSTRACT

Single wire earth return (SWER) frameworks are broadly utilized as a part of inadequately populated zones because of their ease. Nonetheless, identifying issues, particularly the open conductor issues, for SWER can be entirely troublesome. This paper proposes a novel answer for this issue. It includes the era, transmission and location of the force unsettling influence signals on the SWER conductor. Possibility of the proposed plan was confirmed through PC re-enactments. The effects of line length, stacking condition and establishing resistances are explored. The outcomes demonstrate that the proposed strategy is a straightforward and viable plan for recognizing open conductor shortcomings in SWER frameworks.

Received on: 08th -June-2016
Revised on: 14h-June-2016
Accepted on: 16th-June-2016
Published on: 15th-July-2016

KEY WORDS

Open conductor fault, single wire earth return system, power system protection.

***Corresponding author: Email:** yuvarajastr@gmail.com; **Tel:** +91-9043255408

INTRODUCTION

Single wire earth return (SWER) frameworks are broadly utilized worldwide as a part of New Zealand, Australia, Brazil, India and North America for appropriating power to rustic zones, where the end clients are scattered inadequately with low populace thickness and light load necessities [1], [2]. Typical SWER system makes a solitary stage circuit by using only one current conveying transmitter and the earth as the present return way. This topology drives a noteworthy cost sparing contrasted with the ordinarily utilized three-stage and two-stage dispersion frameworks, at all the line equipment (conductor, shafts, separators, and so on.) or the insurance gadgets (wires, re-nearer, and so forth.) According to the joint report issued by the World Bank and NRECA universal, measurement results have demonstrated that almost 200,000 kilometers of SWER lines are in administration among the provincial zone of Australia, which have been bringing about 30% capital cost sparing contrasted with the traditional dissemination framework [3]. Therefore, SWER frameworks are exceptionally appealing in the provincial dispersion markets.

Disregarding the financial focal points of SWER framework, as the length of the single stage conductor could degree to many kilometers, open conductor shortcomings can be a genuine concern since they are not really discernible by customary insurance hardware [4]. What's more, such occurrence may take quite a while to wind up saw to service organizations since it is troublesome and tedious for the support staff to identify flaws among a boundless country range [5]. Besides, an open conveyor shortcoming does not just purpose electric blackout issue to the downstream client, it is likewise one of the pivotal dangers prompting bushfires as the broken transmitter may light the close-by venetians [6]. The disaster happened in Australia, specifically the Black Saturday, on Aug. 29, 2009 took 190 lives and cause 400 individuals harms. The most genuine bushfire that day, the Kilmore East fire, was created by a broken conductor from the nearby SWER framework [7]. In the event that a viable and solid open conductor flaw location plan have been produced and connected, it is conceivable to evade such deadly occurrence, or decline its misfortune. Be that as it may, little work has been done around there. Reference [4] has introduced a model open conductor discovery framework. The framework triggers a review by identifying a voltage hang occasion when the conductor is brought down. The identification sign is a high recurrence current infused into the framework by the coupled current transformer. This plan is substantial for a solitary or three stage appropriation framework where an unbiased conductor exists as the impartial goes about as the sign return way. Reference [8] proposed a broken conductor recognition plan taking into account the framework uneven level, which is not pertinent in the SWER framework. The electrical cable correspondence (PLC) technique is utilized to screen the SWER frameworks continuously [9]. Nonetheless, because of the critical reactance and induction of the SWER conductor, down to earth experience has demonstrated that the normal reverberation recurrence is framework subordinate [13]. The PLC sign is perhaps to be constricted drastically over separation. Another worry of this plan is the changing commotion on the PLC channels, which is additionally demonstrated by [10]. In outline, there is still no agreeable plan to tackle the open conductor shortcoming recognition trouble in SWER frameworks.

This paper shows a novel open conductor shortcoming location system by infusing two-way recognition signals through the SWEN line. The SWER conductor itself goes about as the sign transmission media. After distinguishing the infused signals, the SWER respectability can be perceived. The proposed system requires no additional specialized gadgets, which is essential in the rustic range where the general population correspondence system may not be accessible. Our discoveries have demonstrated that the proposed method is an exquisite and successful methodology for identifying open conductor deficiencies in various SWER frameworks.

DEPICTION ON PROJECTED SYSTEM

The structure of the proposed technique is depicted in **Figure-1**. A master unit is connected to the SWER line through a step down transformer (*T*2) near the re-closer, and one slave unit is installed at the secondary side of each customer service transformers (*T*3) as a regular residential appliance. The master unit monitors the SWER conductor current, and each slave unit monitors the customer side system voltage, respectively.

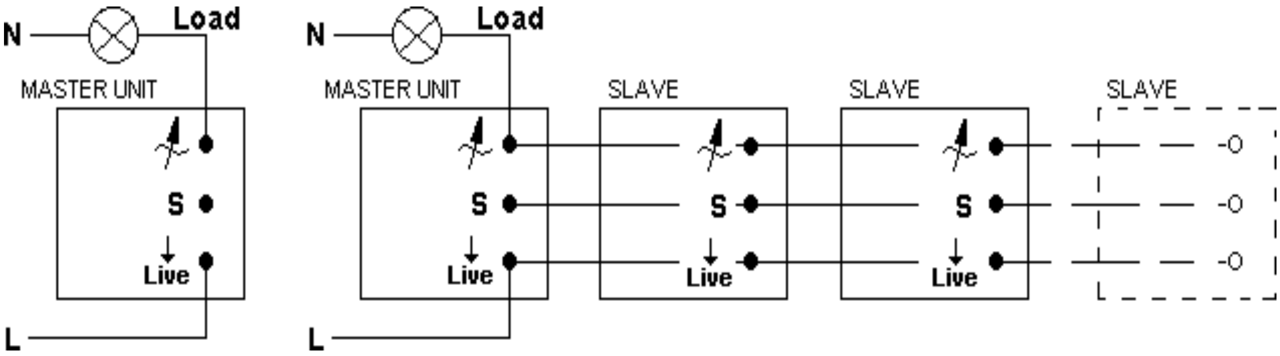


Fig: 1. Proposed scheme for the installation of master and slave units

In this work, we propose to utilize a waveform unsettling influence going on the SWER line as the open conductor deficiency finding signal. Such power unsettling influence flagging procedure has been utilized effectively for programmed meter perusing [11] and DG islanding recognition [12]. The fundamental thought behind is to frame a two-way correspondence circle so that the expert unit occasionally asks for all slave units to report their status. Every slave unit sends back its input catching up a pre-decided correspondence convention, each one in turn sub-successively. In particular, in **Figure-1**, a thyristor portrayed as SCR1 in the expert unit is set off a few degrees before its voltage zero intersection, making a flashing short out heartbeat inside one cycle. SCR1 consequently kills when its current contrarily one-sided. The SWER transport voltage is misshaped by SCR1 terminating occasion. Such voltage contortion is used as a sign. Generally there is no waveform twisting and subsequently no sign. Since the SWER appropriation framework is an outspread system, such flag is capable shown outbound to all the downstream client branches, where the slave units are associated with. The proposed flagging plan contains an exceptional example: it is made each other cycle. The waveform contrast between two sequential cycles is used to recognize the presence of the sign. This plan enormously encourages the sign location process as the obstruction from the foundation enduring state sounds, which are exceptionally normal in the conveyance framework, is dispensed with. The fruitful business sector experience of TWACS [11] has demonstrated that the proposed flagging plan has insignificant effect on force quality.

Every slave unit continues "listening to" the framework voltage at the client site. By contrasting any two sequential cycles, the solicitation signal from the expert unit can be recognized. Slave unit's reaction sign is produced in a comparative way by terminating the thyristor SCR2, which makes an inbound current unsettling influence recognizable upstream. To show such correspondence circle, case flagging waveforms recorded at the expert unit and at one of the slave units required in are exhibited in **Figure-2** The terminating plots for the thyristors SCR1 and SCR2 in this illustration are both designed as 150 degrees. Notice that when SCR1 is let go, the framework voltage V_{bus} detected by the slave units encounters a brief voltage droop, appeared as ΔV_{bus}. Thus, a present aggravation

ΔI_{bus} happens and can be observed by the expert unit if any slave unit works. Distinctive terminating edge might be set in the real execution to tune the sign quality legitimately to fit the particular SWER framework. The outbound and inbound sign attributes will be clarified in Section II.

Along these lines, the two-way correspondence circle is set up. In real usage, a dreary correspondence stages running for settled time interims are characterized as standard report exercises. Contingent upon the size of SWER framework, one open conductor investigation procedure could take up to 30 minutes or even less, which is much quicker than the time spent by utility work force to recognize this sort of issue.

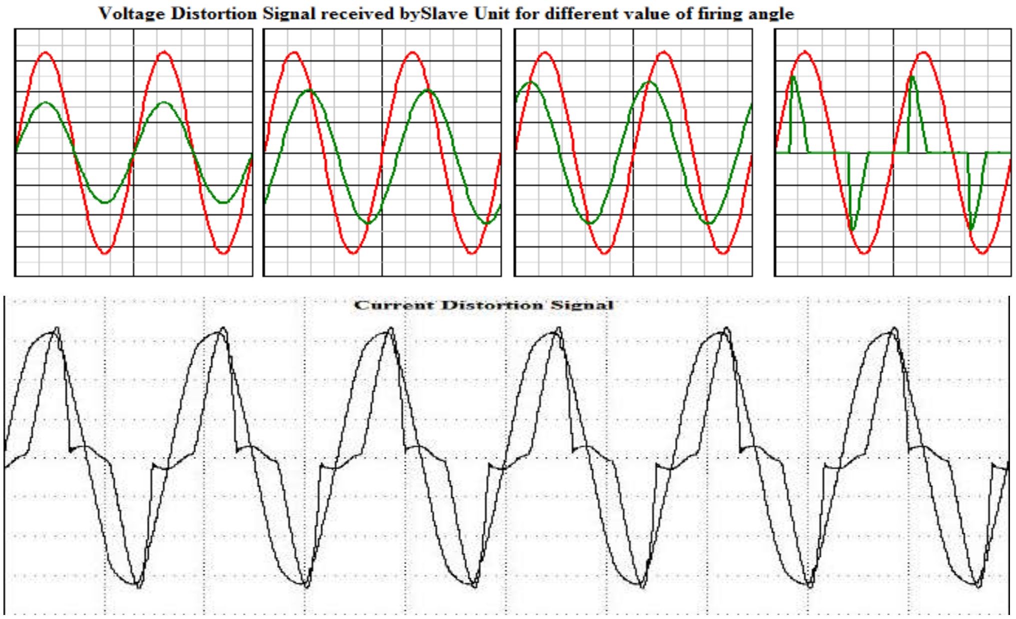


Fig: 2. Example outbound and inbound signals at the master and slave units

SIGNAL GENERATION AND DETECTION SCHEME

The proposed open conductor location strategy incorporates two sorts of gadgets: one expert unit associated with the transport that shows outbound signs to all conveyed client branches intermittently, and a few slave units at the client branches that react such request utilizing inbound signs.

Outbound Gesture Cohort

Outbound sign is created by SCR1 terminating occasion at the expert unit. The outline of the outbound sign requires appropriate choice of the sign transformer T2 and the thyristor SCR1, to fulfill the sign quality and guarantee the hardware warming anxiety. The situation where the heaps are lumped at the remote end of the SWER line is appeared in **Figure-3**. This improved circuit is utilized to break down the outbound sign qualities and to decide the outline elements said above.

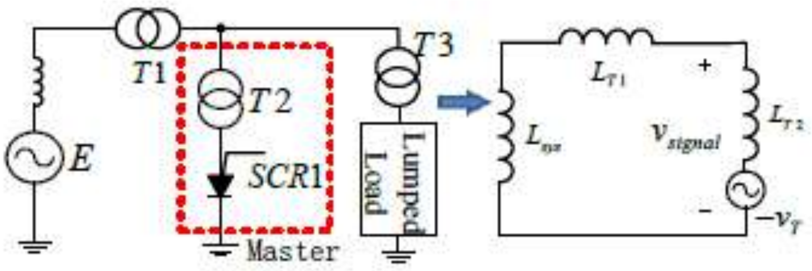


Fig: 3. Circuit analysis of outbound signal characteristics

The steady-state voltage at the secondary side of the signal transformer *T*2 (phase-ground) is expressed as

$$V_{T2}(t) = -\sqrt{2V_N}\sin wt \qquad (1)$$

where V_N is the rated voltage at the secondary side of *T*2. During the firing event, it is equivalent to injecting a negative voltage source –*vT*2(*t*) between the secondary side of *T*2 and the ground, as depicted in the circuit analysis of **Figure-3**. In this figure, *LT*1 represents the isolation transformer inductance, *L*sys is the system inductance upstream *T*1, and *LT*2 is the signal transformer inductance.

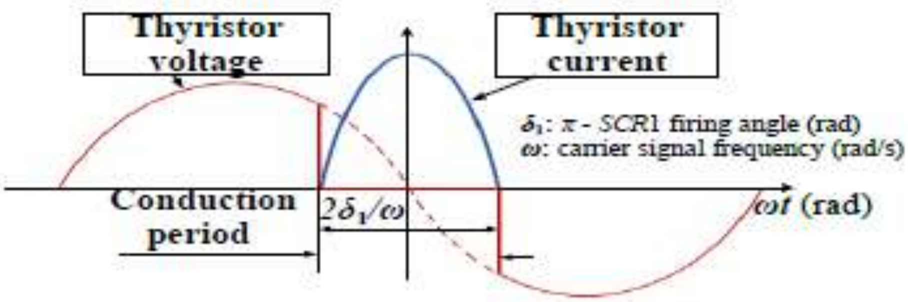


Fig: 4. Sample thyristor waveform for generating the outbound signal

During *SCR*1 firing event, all the loads downstream the master unit are neglected in the circuit analysis as their impacts on the outbound signal are minor, which is same as the loads are ignored in the regular single-phase-to-ground short-circuit calculation. The corresponding waveform illustrating the *SCR*1 operation is shown in **Figure-4**. Defining *δ*1 is the angle ahead the zero-crossing point of the carrier voltage (*vT*2) in the after *SCR*1 turning on, the thyristor conduction period will be 2*δ*1. In this way, the outbound signal can be derived as

$$V_{signal}(t) = \sqrt{2V_N}\,\frac{L_T}{L_T + L_{T2}}\sin wt,\ wt \in [-\delta_1, \delta_1] \qquad (2)$$

where *LT=Lsys+LT*1.

The signal transformer *T*2 limits the short circuit current and reduces the bus voltage distortion level during the *SCR*1 firing period. The peak of *vsignal*(*t*) represents the strength of the outbound signal, which can be determined as:

$$V_{signal\_peak} = \sqrt{2V_N}\,\frac{L_T}{L_T + L_{T2}}\sin\delta_1 \qquad (3)$$

The ratio of the outbound signal peak to its carrier voltage peak *koutbound* is another useful index to estimate the signal strength relatively, which is derived as the follows:

$$K_{outbound} = \frac{V_{signal\_peak}}{V_{PG\_pea}} = \frac{L_T Sin\delta_1}{L_T + L_{T2}} = \frac{X_T\ Sin\delta_1}{X_T + X_{T2}} \quad (4)$$

where X_{T1} and X_{T2} denote the transformer leaking reactance as:

X_{T1} = ωL, X_{T2} = ωL, $X_{T1,Tsys} = X_{T1} + X_{T2}$

When designing the outbound signal, Eqn(4) can be used to adjust a proper signal strength. Other two important factors that need to be considered are the ratings of the thyristor *SCR*1 and the signal transformer *T*2. Based on the analysis above, the peak, average and RMS values of the thyristor current is calculated to guide the equipment selection as:

$$I_{peak} = \frac{\sqrt{2V_N}}{X_T + X_{T2}}(1 - \cos\delta_1) \qquad (5)$$

$$I_{rms} = \frac{I_{peak}}{\sqrt{N_1}}\,\frac{1}{1 - \cos\delta_1}\sqrt{\frac{1}{2\pi}[\delta_1(2 + \cos 2\delta_1) - 1.5\sin 2\delta_1]} \quad (6)$$

$$I_{mean} = \frac{I_{peak}}{N_1}\,\frac{\sin\delta_1 - \delta_1\cos\delta_1}{\pi(1 - \cos\delta_1)} \qquad (7)$$

where *N*1 stands for the number of cycles that the outbound signal is generated one time. **Figure-2** illustrates the situation that *SCR*1 is operated once every two cycles, i.e. *N*1=2. A larger value of *N*1 will release the heat dissipation stress both on the signal transformer *T*2 and the thyristor *SCR*1, with the tradeoff of the communication speed reduction.

Incoming Gesture Cohort

The inbound sign is created by the downstream slave unit thyristor SCR2 terminating occasion. In the genuine usage, the slave unit is a reduced and convenient gadget that can be connected to the low voltage framework as a general family machine. As presented in Section II, an individual slave unit sends back its status report just once per assessment arranges, the warming anxiety brought about by the thyristor SCR2 operation would not be a major concern. There is impedance associated in arrangement with SCR2 whose reason for existing is to constrain the inbound current.

Our viable experience has reasoned that an inductor (LSG in **Figure-3**) is the best choice from various points of view, for example, the size, the force rating and the expense so as to deliver sufficiently solid current heartbeat. The rearranged equal circuit when SCR2 is terminated can be found in **Figure-5**. Amid a terminating occasion that goes on for a few cycles (contingent upon the correspondence convention), the SWER channel current comprises of the inbound current sign drawn by a slave unit and the current from whatever remains of electrical loads additionally associated in the circuit. Since the operation of one slave unit has unimportant effect on different burdens, the heaps in parallel with the slave unit are dismissed.

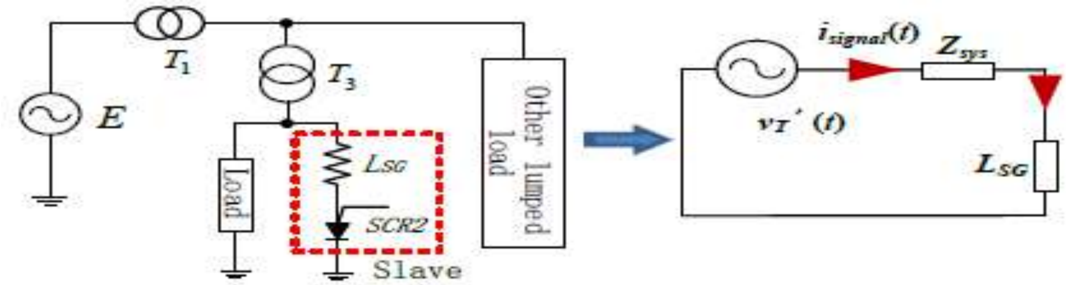


Fig: 5. Circuit analysis of inbound signal characteristics

Similarly, the circuit analysis when the slave unit is fired is shown in **Figure-5**. δ_2 is defined as the angle between the trigger instant and the carrier voltage (*vT'*) zero-crossing, the inbound current is determined by

$$i_{signal}(t)=\frac{\sqrt{2V_N}}{Z_{sys}+X_{SG}}\left(\cos wt-\cos\delta_2\right), wt\in\left[-\delta_{2,}\delta_2\right] \quad (8)$$

where *V'N* is the rated voltage at the secondary of the service transformer T_3. *Zsys* represents the system impedance upstream the slave unit. X_{SG} = ωL_{SG}. The peak, average and RMS values of the inbound current signal are calculated as the follows to guide the selection of thyristor SCR_2 and inductor L_{SG} as:

$$I_{peak\_inbound}=\frac{\sqrt{2}V_N}{Z_{sys}+X_{SG}}\left(1-\cos\delta_2\right) \quad (9)$$

$$I_{rms\_inbound}=\frac{I_{peak\_inbound}}{1-\cos\delta_2}\sqrt{\frac{\delta_2\left(2+\cos2\delta_2\right)-1.5\sin2\delta_2}{2\pi N_2^2}} \quad (10)$$

$$I_{mean\_inbound}=\frac{I_{peak\_inbound}}{N_2}\frac{\sin\delta_2-\delta_2\cos\delta_2}{2\pi N_2^2} \quad (11)$$

Where N_2 stands for the number of cycles that the inbound signal is generated once. Since the slave unit is deployed at the low voltage side with relatively less firing frequency, we normally set N_2=2 to speed up the report time.

Signal Detection Procedure

In Section II, the inbound and outbound sign era and extraction plans have been presented quickly. **Figure-6** demonstrates a case outbound sign waveform separated by the two-back to back cycle subtraction technique (signify as ΔVsignal). Such waveform subtraction is controlled constantly at the expert unit. The test is to recognize if any of the Δvsignal waveform undoubtedly contain the outbound sign. Hence, a solid and simple to-use signal location calculation is required. After one of expert unit terminating occasions, the outbound sign spreads along the SWER line. At the location side, a progression of motions can be found. Such oscillatory example is valuable for the recognition calculation. In real usage, ΔVsignal is handled by the Discrete Fast Fourier Transform (DFFT). The DFFT window, or the recognition window, is the length of half of 60 Hz cycle. It begins at 150° and closures at 330° of the bearer waveform. The identification window utilizes the zero intersection purpose of the bearer voltage as an including reference. As the location window width is 180°, just even request music are acquired.

In **Figure-6**, the SWER transport voltage goes about as the transporter. Then again, for the inbound sign, the conductor current is the bearer. As the terminating point in the proposed plot ordinarily equivalents to 150°, the greater part of the sign waveform (principle segment and the motions) will be caught by the recognition window.

Preferably, the consonant segments from the FFT consequences of ΔVsignal could be utilized to register a record to demonstrate the sign presence. The sounds whose frequencies near the outbound sign recurrence are chosen. In any case, it is hard to foresee such frequencies as they are exceptionally framework subordinate. Existing field test results demonstrate that the characteristic frequencies for the North American circulation frameworks are in the scope of 200 Hz to 600 Hz [13].

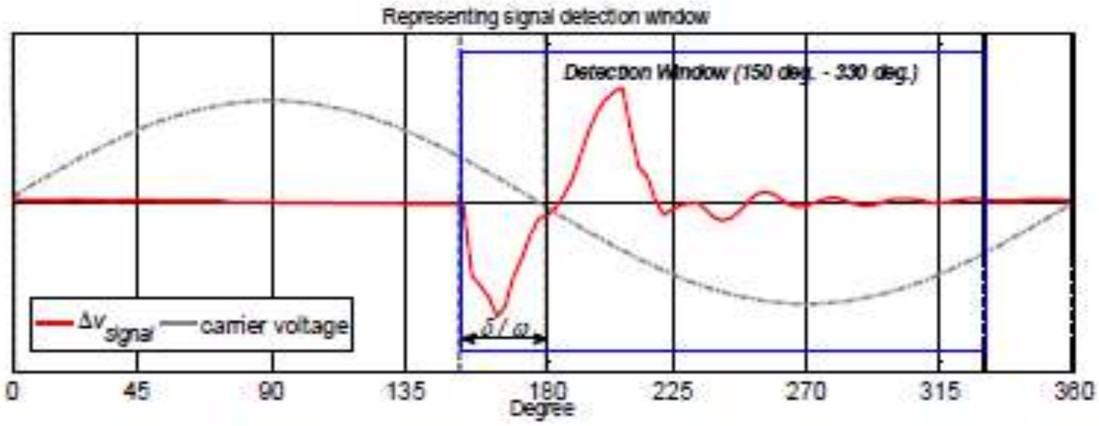


Fig: 6. Signal detection and extraction scheme based on 180° window

Δ*Iinbound_h* is the *hth* harmonic current magnitude of the FFT results of Δ*isignal* from the inbound signal detection side. To illustrate the proposed detection algorithm, a representing outbound signal captured by the proposed signal detection window is presented in **Figure-7**. The SWER conductor length in this case is 100 km, the loads are with 0.95 power factor and 50% service transformer capacity (sparse load condition). The harmonic components of the signals up to 1080Hz computed by the 180° window FFT are also shown. It can be seen that the outbound signal exist in the detection window. The dominant harmonic components are included in the detection spectrum proposed detection algorithm as well. Similarly, one representing case demonstrating the inbound signal situation with the same system parameters is shown in **Figure-8**.

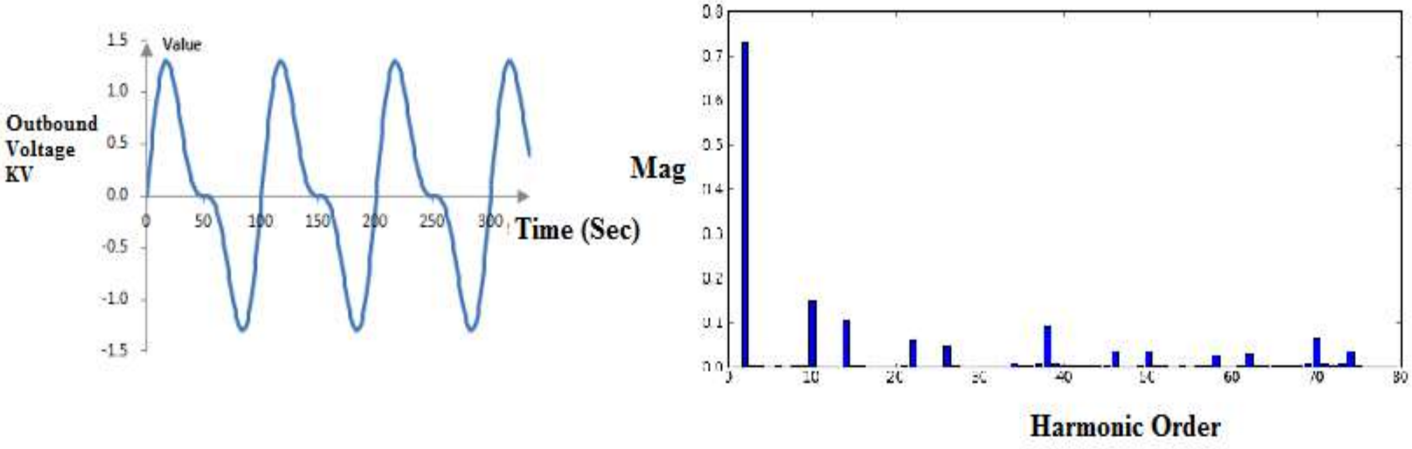


Fig: 7. Spectrum based outbound signal detection algorithm

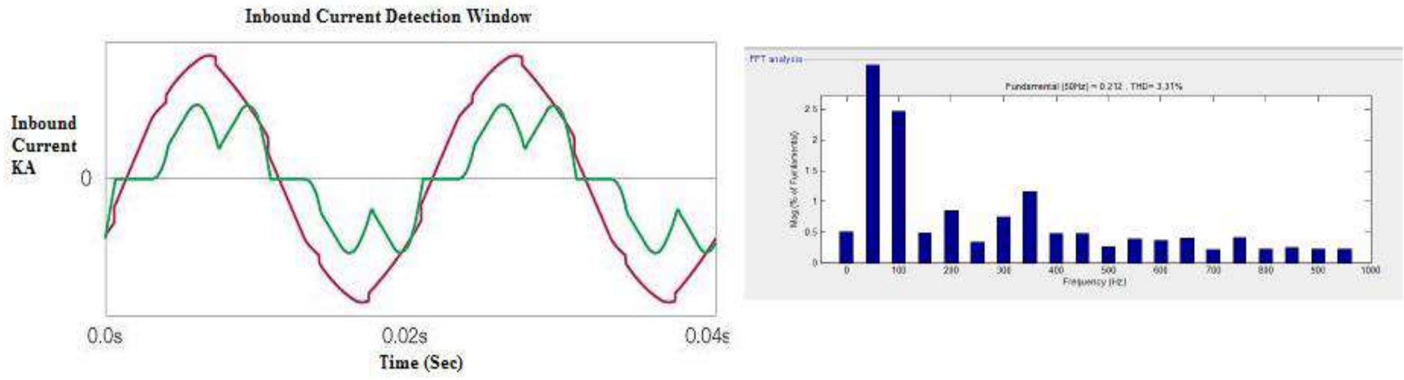


Fig: 8. Spectrum based inbound signal detection algorithm

The outbound signal or inbound signal is considered as existing only if the computed $DT_{outbound}$ or $DT_{inbound}$ index from Eqn (12) or Eqn (13) is higher than the preset threshold value, respectively. These two thresholds are independent and need to be determined in the field tests.

COMMUNIQUÉ SYSTEM

The solicitation summon from the expert unit and the criticism from the ointment units are known not other in view of a preset correspondence convention. A progression of force aggravations are utilized to speak to a grouping of computerized data. A case in **Figure-9** demonstrates the utilization of thyristor SCR1 terminating in a succession to speak to computerized bits. Two cycles are expected to speak to one advanced piece, i.e. in the event that the subtraction of the two back to back cycles incorporates the infused voltage unsettling influence; it speaks to a bit '1'. The unmoving status, generally, speaks to '0'.

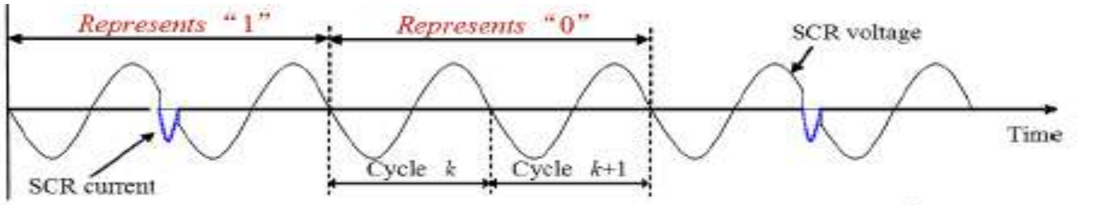


Fig: 9. Representing communication bits by a sequence of *SCR*1 firing pulses

An ordinary correspondence exchange is depicted as the accompanying strides: ordinarily, the slave units stay as the unmoving status, i.e. the open conductor shortcoming assessment is not empowered. They remain by unit accepting the summon from the expert unit. Commonly, three sorts of charge can be transmitted to the slave units: the investigation empower outline, the status demand outline, and the examination cripple outline. The proposed outline structure utilized by the expert unit to send order and by the ointment unit to report is shown in **Figure-10**. The aggregate number of the slave IDs relies on upon the size of the SWER framework. Taking a system containing 250 circulated load branches (250 slave unit establishments) as a case; it devours 40 cycles (60 Hz framework recurrence, N1=N2=2) to finish one correspondence outline (found in **Figure-10**). We characterize the time interim for producing a complete popularity as tF, where tF equivalents to roughly 667 ms.

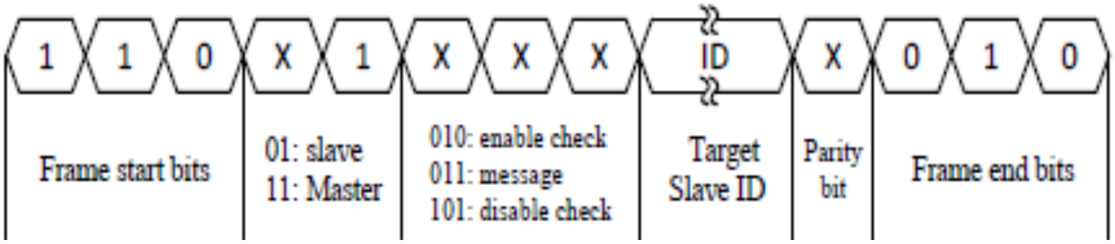


Fig: 10. The communication frame structur

A correspondence circle characterized by the proposed correspondence convention is exhibited in **Figure-11**. The expert unit actuates the status check by producing an examination empowers edge, to educate the slave units to exit from the unmoving status and get prepared for the approaching status check process. After a period interim of td (500 ms in current usage), the expert unit begins to send demand request downstream. Just the objective slave unit who's ID coordinate the ID portion contained in the casing, reacts the expert unit's request. Such ace slave handshakes are rehashed in a steady progression, one set at every time. The time interim of td exists between each two handshakes to guarantee the gadget operation completely finished. The status check procedure is done until all the slave units are recognized. Toward the end of a correspondence circle, one edge of review cripple summon is sent by the expert unit to recuperate all slave units to the unmoving status.

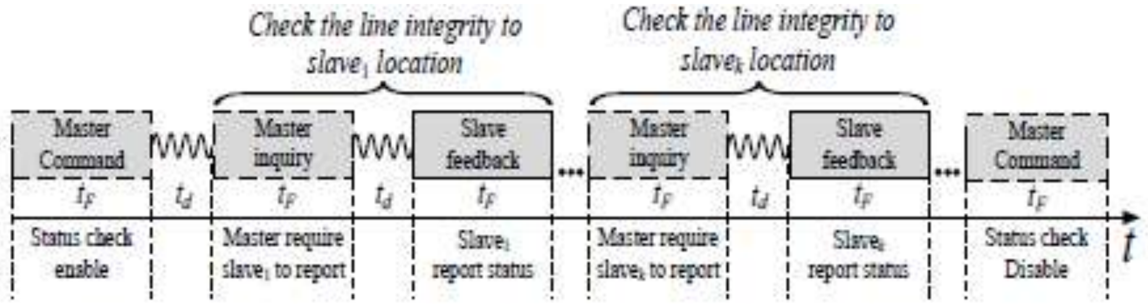


Fig: 11. The normal communication protocol for SWER line inspection.

As appeared in **Figure-11**, for a SWER system with M slave units introduced, the aggregate time expended to finish one effective correspondence circle will be:

As per Eqn (14), it is anything but difficult to figure that the aggregate time expected to run a 250 branch SWER framework conductor respectability review is roughly 10 minutes. Once the expert unit demand one of the slave units to report its status yet neglects to get its input inside te, a hazardous slave area is found. te is in any event as td+ tF. In current plan, te is set to 2 second. Two extra status check endeavors requiring this particular slave unit to input are produced to affirm the correspondence uprightness at this area (**Figure-12**). A conceivable open conductor deficiency area is distinguished after three endeavour disappointments and the expert unit will caution the operation focus promptly with respect to this matter.

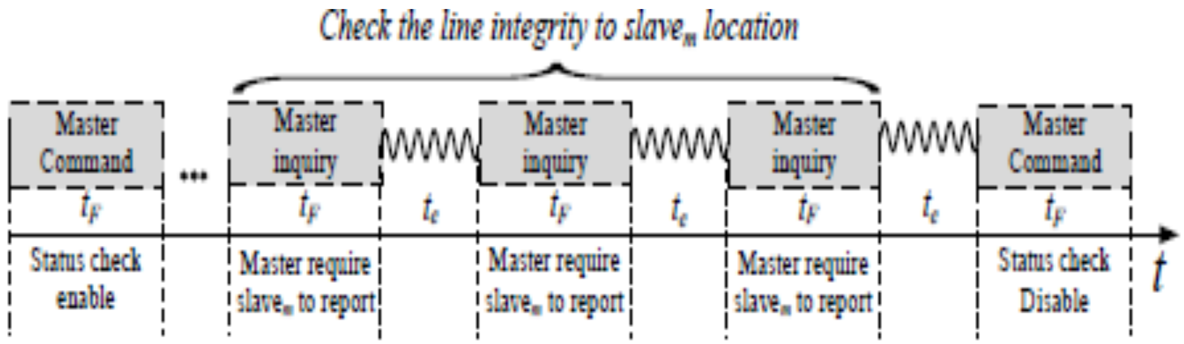


Fig: 12. The communication protocol for suspicious open conductor fault.

SIMULATION AND ENDORSEMENT

To approve the possibility of the proposed procedure and assess its execution in various SWER frameworks, broad PC reproductions are performed on a delegate test framework by utilizing PSCAD/EMTDC. The SWER framework utilized as a part of the reproductions is from a real SWER framework given in [14]. Its topology is appeared in **Figure-13** and the primary parameters of this framework are recorded as the takes after.

- The input of the test SWER system is coming from one phase out of a 22kV three-phase feeder. The conductor used is #6 ACSR; its default length is 150km.
- SWER overhead line is modelled by Carson's line model, with its self-impedance and shunt admittance computed according to the reference [15].

- $T1$'s capacity is 150kVA, step-down ratio is 22/12.7kV with 2.5% impedance and is grounded through a resistor $R1=2\ \Omega$. $T2$'s capacity is 5kVA, step-down ratio is 12.7/0.6kV, solid grounded. $T3$ is 25kVA, step-down ratio is 12.7/0.12kV with 2.5% impedance and grounded through a resistor $R2=10\Omega$.
- The firing angles of master unit and slave unit thyristors are all set to 150 degree ($\delta1=\ \delta2=30°$). A 10mH inductor is embedded in the slave unit, limiting the inbound signal pulse peak.
- SWER system is a radial network where lumped loads are connected at the remote end to represent the worst scenario for detecting the signal. The default loads consume half of the service transformer capacity with $R=2.4\Omega$ and $L=19.6$mH.

SWER overhead line's influence on signal broadcast

Numerous PC reenactments were done to assess the effect of single conductor length on the signs quality. Both the outbound sign and the inbound sign exhibitions are analyzed. The primary finding is that as the SWER line augments, both the outbound sign and the inbound sign don't weaken at the recognition side. The proposed calculation (as clarified in Section V) in light of the sign qualities saw at the era side can adequately perceive the signs. Variable line lengths are tried; the reproduction results for the situations when SWER line is 50 km and 200 km are displayed in **Figure-14**and **Figure-15**. In this investigation, the normal full cycle discovery window (180°) is utilized. The intention is to look at the time area signal waveforms and the recurrence space consonant segment extents between the era side and location side. In **Figure-1**4, the outbound sign is measured at the sign transformer T2 essential ('era side' in **Figure-15**) and the administration transformer T3 essential ('discovery side' in **Figure-14**) where the voltage levels are same (transport voltage). In **Figure-15**, the inbound sign is measured at the essential of T3 ('era side' in **Figure-15**) and from the SWER line where the expert unit is associated ('discovery side' in **Figure-15**).

At the point when the SWER line is short, e. g. SWER is 50 km, the line length has minor effect to the sign quality. As the SWER line turns out to be longer, huge motions accompanies the sign era are found. It is a result of the expanded inductance and permission connected with the coupled overhead single stage line. In any case, the signs don't constrict, yet stays sufficiently solid to be effortlessly recognized. Such flag improvement is because of the common resounding of the SWER framework at a specific line length [2].

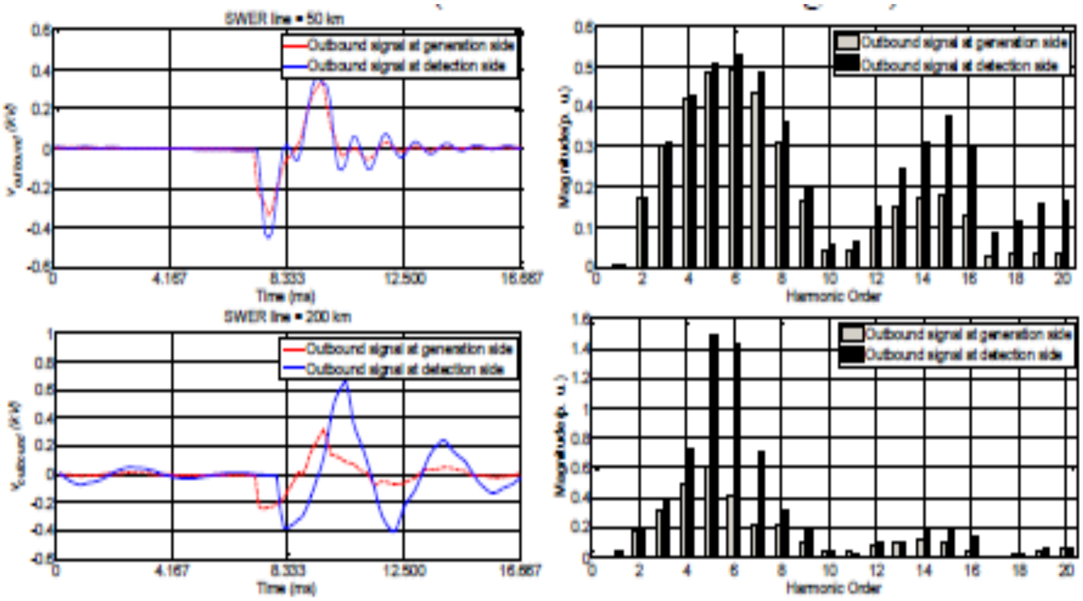


Fig: 14. Outbound signal performance affected by SWER line length

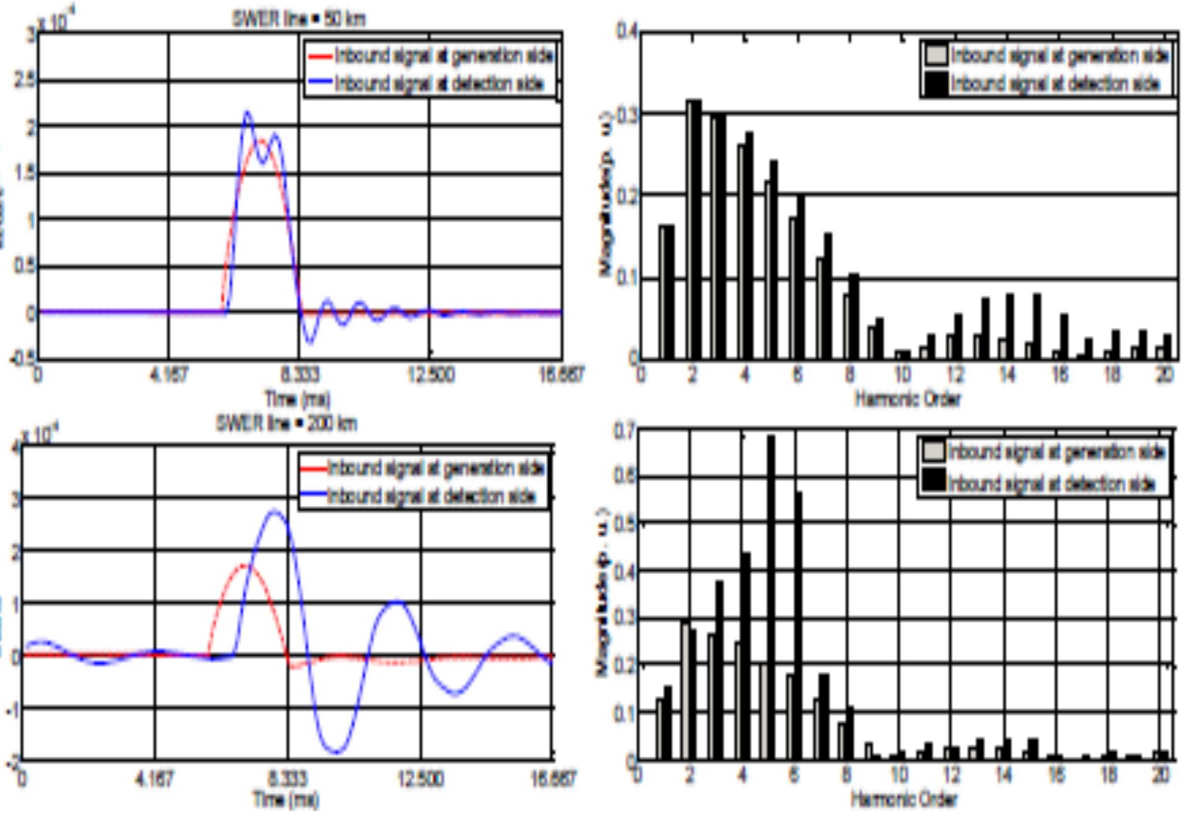


Fig: 15. Inbound signal performance affected by SWER line length

Service transformer loading impact

In Section III, we assumed the customer side loads connected in parallel with the slave unit has negligible impact on the inbound signal. In this way, the slave unit at the customer side generates a constant current signal described by Eqn (9)-(11).

To verify the consistency of the inbound current at the two different locations, at the thyristor *SCR*2 branch and at the service transformer *T*3 branch, different loading conditions are tested. The inbound signals when *T*3 is loaded 70% are shown in **Figure-16** where the power factor remains at 0.95. The inbound signal is measured from the secondary of *T*3 to ensure comparison is made at the same current level. It can be seen in **Figure-16** that the agreement is quite good. The small difference of the waveform is caused by the system impedance *Zsys* upstream the slave unit as Eqn(9), which drops the *SCR*2 branch voltage slightly.

Grounding resistance influence

The impacts of establishing resistance of administration transformer at the client side is observed to be negligible, regardless of the extensive variety of qualities utilized as a part of the study (R2=10*Ω*, 20*Ω* and 50*Ω*). The transient reactions of the outbound sign at the recognition side (essential of the administration transformer) connected with the tried establishing resistance varieties are portrayed in **Figure-17**. Touchy studies in regards to the sign transient reaction in light of these reproduction results turn out that with the exception of the ling length, other delicate parameters have insignificant effect on the sign quality, yet the line length does not endanger but rather improve the viability of the proposed recognition calculation.

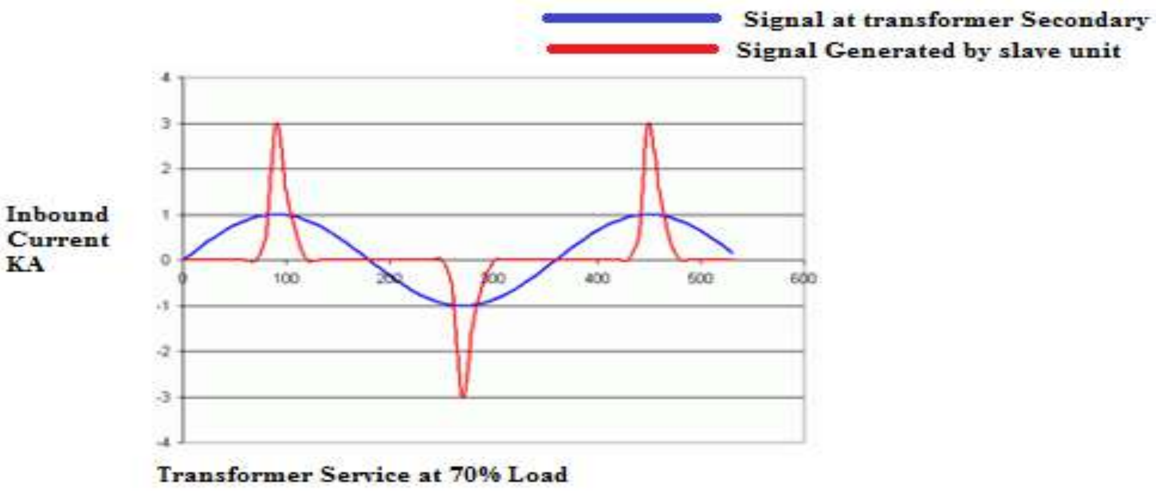


Fig: 16. Inbound signal affected by service transformer loading condition

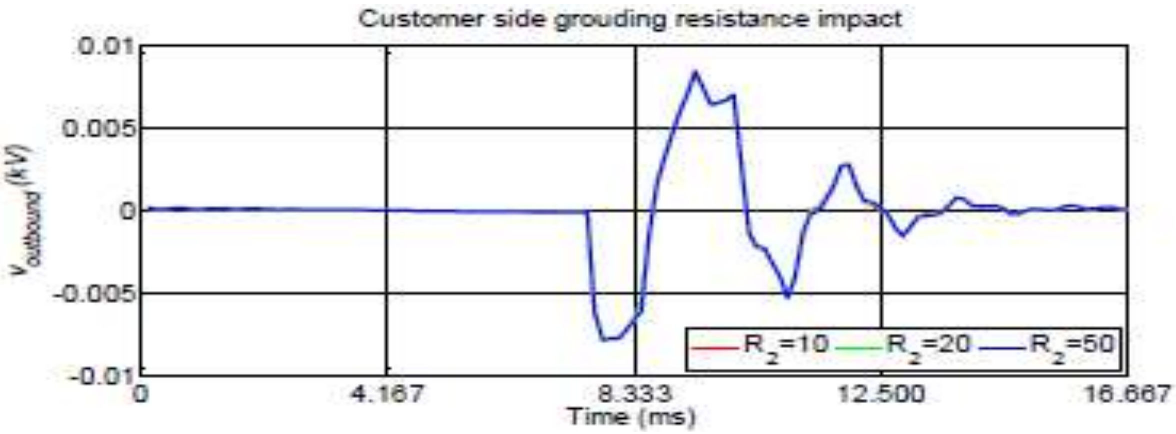


Fig: 17. Outbound signal affected by customer side grounding resistance

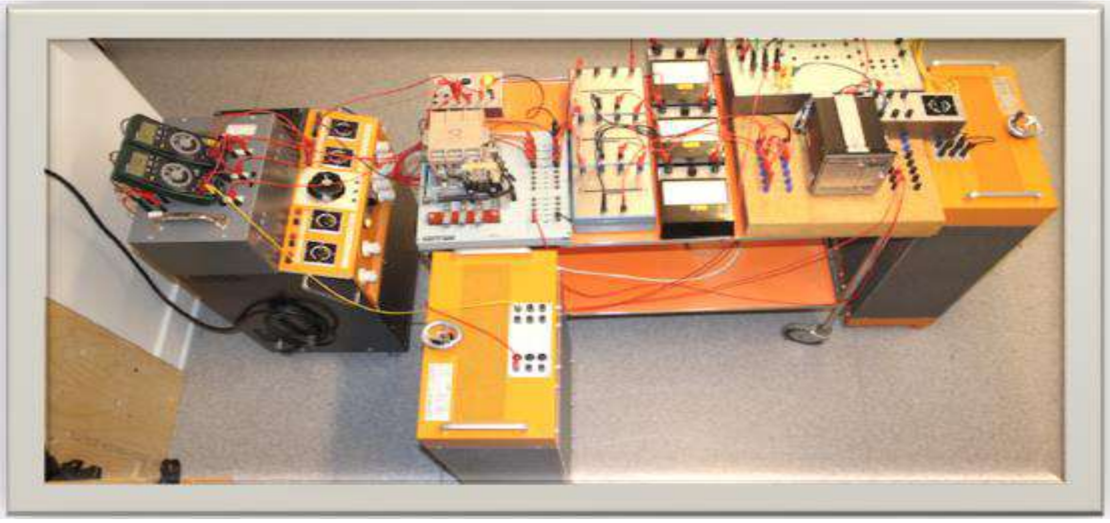


Fig: 18. Phase to Earth Network with resistance grounded neutral (Earth Fault Circuit)

CONCLUSION

This paper proposed a novel and appealing approach solely intended to recognize open conduit flaw in the single wire-earth-return framework, which is a major test and a dire need to utility designers. The proposed plan depends on two-way control flagging method, the outline contemplations for the sign attributes and the gear determination are introduced. A succinct and simple to-use signal recognition calculation is presented. The proposed strategy has been connected to a test SWER framework and broad reproduction results have demonstrated its viability for identifying the open conductor shortcomings, in any case the single conductor length, the stacking conditions and the client establishing resistance. Such acceptances underscore the strength of the proposed plan, so it can be utilized with certainty to various SWER frameworks.

CONFLICT OF INTEREST

The author declares having no competing interests.

ACKNOWLEDGEMENT

None.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] PJ Wolfs. “Capacity improvements for rural single wire earth return systems,Proc. the 7th international Power Engineering Conference, IPEC 2005, pp. 1–8, Dec..

[2] F de La Rosa, and S. Mak, “A look into steady state and transient performance of power lines integrating single line wire earth return circuits,” in Proc. 2007 Power Eng. Society General Meeting, pp. 1-6.

[3] AR Inversin. [2000]“Reducing the cost of grid extension for rural electrification,” NRECA International, Ltd. and World Bank Energy Sector Management Assistance Program, Feb..

[4] AC Westrom, A PS. Meliopoulos, GJ Cokkinides, and AH Ayoub "Open conductor detector system," IEEE Trans. Power Del 7@43, pp.1643-1651, Jul..

[5] Parsons Brinckerhoff Australia Pty Ltd., “Indicative costs for replacing SWER lines” Department of Primary Industries of Australia, August 2009, Available at: http://www.dpi.vic.gov.au/energy/safety-andemergencies/ powerline-bushfire-safety-program/indicative-costsreplacing- swer-lines

[6] Energy Safety Department of Commerce, Government of Western Australia, “Final electrical incident report bushfire near river and Folewood roads Toodyay,” Aug. 2010.

[7] Victorian Bushfires Royal Commission.[2010.] “The 2009 Victorian Bushfires Royal Commission final report ,” I, Jul.

[8] EC Senger, W Kaiser, JC. Santos PMS Burt and CS. Malagodi, "Broken conductors protection system using carrier communication," IEEE Trans. Power Del., vol.15, no.2, pp. 525-530, Apr. 2000.

[9] Gay D, Thompson A, Amanulla, M.T.O.; Wolfs, P.[. 2009] Monitoring of Single Wire Earth Return systems using Power Line Communication," Power Engineering Conference, 2009. AUPEC 2009. Australasian Universities, vol., no., pp.1,5, 27-30 Sept.

[10] Kikkert CJ, Reid GD.[2008] Radiation Losses from a Single Wire Earth Return Power Line with Bends," Telecommunication Networks and Applications Conference, 2008. ATNAC 2008. Australasian , vol., no., pp.158,162, 7-10 Dec

[11] S. T. Mak, and D. L.kReed, "TWACS, A new viable two-way Automatic communication system for distribution networks. Part I: outbound communication," IEEE Trans. Power Apparatus and Systems, vol.PAS- 101, no.8, pp. 2941-2949, Aug. 1982.

[12] W Xu, G Zhang, C Li, W Wang, G Wang, and J Kliber, "A power line signaling based technique for anti-islanding protection of distributed generators - Part I: scheme and analysis," IEEE Trans. Power Del., vol. 22, no. 3, pp. 1758-1766, Jul. 2007.

[13] ST Mak and RL Maginnis, “Power frequency communication on long feeders and high levels of harmonic distortion,” IEEE Trans. Power Del., vol. 10, no. 4, pp. 1731–1736, Oct.

[14] N Chapman, “When One Wire is Enough,” Transmission and Distribution World, vol. 53, no. 4, pp. 56, Apr. 2001.

[15] JR Carson. Wave propagation in overhead wires with earth return. Bell Syst. Tech. J., vol. 5, pp. 539–554, Oct. 1926.

[16] Please rearrange reference section as per our journal format. Intead of “ST Mak and RL Maginnis, “Power frequency communication on long feeders and high levels of harmonic distortion,” IEEE Trans. Power Del., vol. 10, no. 4, pp. 1731–1736, Oct.

ARTICLE OPEN ACCESS

A DISTRIBUTED APPROACH FOR PREDICTING MALICIOUS ACTIVITIES IN A NETWORK FROM A STREAMING DATA WITH SUPPORT VECTOR MACHINE AND EXPLICIT RANDOM FEATURE MAPPING

Prabaharan Poornachandran [1], Premjith B[2], Soman K. P[2]

[1]*Amrita Center for Cyber Security Systems and Networks, Amrita Vishwa Vidyapeetham, Kollam, INDIA*

[2]*Centre for Computational Engineering and Networking, Amrita Vishwa Vidyapeetha,, Coimbatore, INDIA*

ABSTRACT

Technology reduces human effort. However technological advancements always bring threat to personal as well as organizational security, mainly because we all are connected to the internet. Therefore, ensuring cyber security becomes the major topic of discussion. As the magnitude of activities over the internet is unimaginable, envisioning the characteristics of network activities whether it is malicious or good, coming from a stream of data in real time is really a tough task. To tackle this problem, in this paper, we propose a distributive approach based on Support Vector Machine (SVM) with explicit random feature mapping and features mapping is obtained using Compact random feature maps (CRAFTMaps) algorithm. Distributing the job achieves notable improvement in the total prediction time.

Received on: 2nd-April-2016
Revised on: 27h-April-2016
Accepted on: 28th-April-2016
Published on: 26th-July-2016

KEY WORDS

Cyber security, Streaming data, Support Vector Machine, Compact random feature maps, Distribute processing

***Corresponding author: Email:** prem.jb@gmail.com **Tel:** +91-9597141816

INTRODUCTION

As the new technological innovations are emerged, cyber-attacks are also changing the colours. The ubiquity of technology leads to the exponential growth in the cyber threats. Now cyber security has become one of the primary concerns of governments as well as private organizations. An important problem of cyber security is how to effectively monitor and predict threats in real time, i.e. detecting the threats from streaming data. A streaming data is nothing but a massive volume of data coming from various sources, such as videos, images, text etc. and may not be stored in a disk for analysis. Streaming data are considered as data in motion and the analysis is always single-pass, i.e. the data cannot be reanalyzed once it is streamed.

With the increase in network traffic, network logs have become huge and the detection of cyber-threats from these massive streaming logs is now a tedious task. Conventional machine learning algorithms are not suitable for the real time prediction of malicious activities in a network as they require storage of the data to predict whether it is a threat or not. But this is not possible for streaming data. As the data streams are one-pass and highly non-static, the decision has to be taken in quick time. This necessitates the requirement of a fast and scalable mechanism for real time detection of cyber threats from a data stream.

In cyber security, one of the problems is classification of network logs into malicious and benign. In machine learning, classification [1], a supervised learning approach, is used for detecting the malicious activities in a network. Support vector machine [2], Regularized least squares [3] etc. are common classification algorithms. Generally classification algorithms are linear in the sense; they are able to classify data which are linearly separable. But a streaming data or network traffic logs are never linearly separable and are often non-stationary. So offline storage and analysis is quite impossible [4]. Conventional classification algorithms are designed to work with offline data. There are certain other issues which make the classification of data streams makes tougher which are high speed nature of data streams, unbounded memory and hardware requirements, concept drifting, data visualization, challenges in distributed applications, modeling of mining results in real time, tradeoff between accuracy and efficiency etc. [9].

Nonlinear classification algorithms such as non-linear SVM with kernel methods give state-of-the-art accuracy in detecting cyber threats when the prediction is done offline and come with a huge computational cost for real time prediction. Generally kernel methods transform input data to a finite higher dimensional space where a linear separation is obtained. Kernel methods were widely used for classifying data when the data size is relatively small. But if the input data size is massive and if the data are not linearly separable, the number of support vectors (Special subset of input data) to be stored becomes large. Therefore as the size of the data increased, so does the computation time and storage. Explicit feature mapping methods alleviate this curse of support problem and thereby make the classification algorithms appropriate to deal with streaming data. An explicit feature mapping algorithm projects the input feature vectors to a higher dimensional vectors which are randomly generated from a standard normal distribution and then compute the dot products.

Several explicit feature mapping algorithms such as Random Kitchen Sink [5-6], [13], Fastfood [7], Compact random feature maps [8] etc. are generally used for mapping input data explicitly to a higher dimensional space. In this paper we investigate a recently introduced explicit mapping method - Compact random feature maps. Compact random feature map algorithm is a polynomial kernel approximation which resolves the rank deficiency (underutilization of projected space) problem [8] that commonly appears in random mapping.

In this paper we investigate the feasibility of distributing the abnormal activity detection in a streaming data with compact random feature mapping algorithm and Support Vector Machine classification algorithm. Experimental study showed that, like other explicit feature mapping algorithms, compact random feature map algorithm proposed by Hamid, et.al [8] computes only one dot product in the higher dimensional space and hence the storage requirement is very less. This algorithm also manages the underutilized space in the featured space by down projecting the obtained features. Utilizing this advantage of explicit feature mapping algorithm, this paper proposes a parallel implementation of real time prediction of streaming data. Four machines run in parallel implements the decision function and result from all the machines combined to attain the overall prediction time and accuracy.

METHODS

Even though kernel methods are successful for offline prediction, it is found to be difficult to use typical kernel methods for the prediction at real time. This is because of the fact that, the storage and offline analysis of data streams is impractical. So classification algorithms like Support Vector Machine works poorly for streaming data. In order to fix this issue, we utilize compact random feature map algorithm, an explicit way of projecting feature vectors to a higher dimensional space, in a distributed way. Parallelism is achieved by dividing the projected features uniformly and each subset of features is passed to different processors. For each processor, define weight vectors ω . Linear combination of features and weight is computed at each processor and final prediction is computed by combining the results obtained at each processor. **Figure-1** shows the block diagram of how prediction is implemented distributive after projecting the input feature vectors to a finite higher dimensional space.
Mapping of features to another dimension is achieved by kernel trick and is discussed in the next subsection.

Kernel trick

The heart of a classification algorithm is kernel trick. Using kernel trick, the decision function can be computed as,

$$g(x) = \omega^{T}\psi(x) \qquad (1)$$

Here $\psi(.)$ is a feature mapping operator. Unfortunately this feature mapping may lead to infinite dimensionality and make the computation very expensive. In order to nullify this problem, a dual representation was introduced to compute the decision function,

$$g(x) = \omega^{T}\psi(x) = \sum_{i=1}^{d} a_{i}\langle\psi(x_{i}),\psi(x)\rangle \qquad (2)$$

Where $H(x,x') = \langle\psi(x),\psi(x')\rangle$

This representation is called kernel function. Classification algorithms such as Support Vector Machine, a linear classifier in nature, uncover the non-linear relationship among data using the dual representation. This kernel trick reduces computational cost of evaluating the feature mapping function $\psi(.)$.

One disadvantage of the typical kernel methods is scaling problem. That is, these methods often fail with large data set or real time prediction. A fast learning approach is required to deal with this problem. Explicit random feature mapping algorithms are one of the solutions to this problem. Numerous algorithms such as Random Kitchen Sink, Fastfood etc. have been devised for the explicit mapping of input feature vectors to a manageable higher dimensional space. Compact random feature map is one such algorithm which approximates polynomial kernels. The background of all explicit random feature mapping is the method of random Fourier features proposed by Rahimi et al. [5].

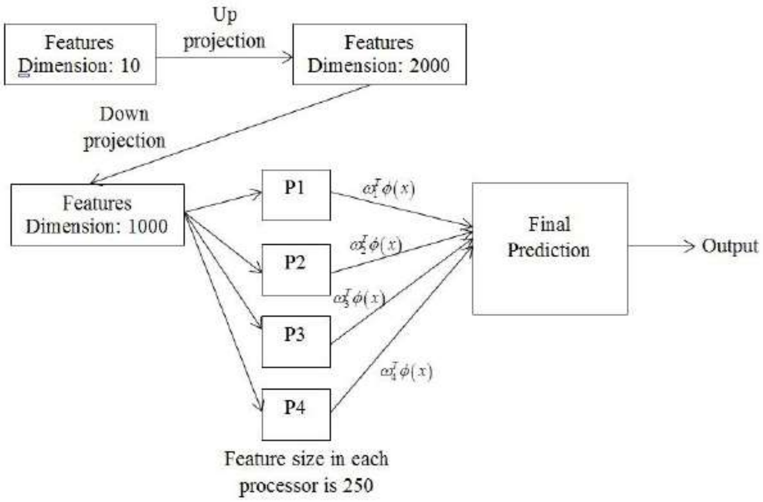


Fig: 1. Block diagram for how the prediction is computed parallel with explicit feature mapping

………

Random Fourier Features

Rahimi et.al proposed an alternate approach, Random Kitchen Sink, to the kernel trick. This approach accelerates the training as well as testing by projecting the feature vector to a manageable random higher dimensional space. And the obtained inner product of the transformed features is approximately equal to the inner product in feature space.

The key theorem behind Random Kitchen Sink (RKS) is Bochner's theorem [10]. According to Bochner's theorem, any shift invariant, continuous function is positive definite if and only if it is the Fourier transform of a positive measure. Mathematically, we can write this theorem as,

$$P(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}H(z)e^{-jz^T\omega}dz \tag{3}$$

Where $z = x - y$.

So as a consequence of Bochner's theorem, the inverse Fourier transform is interpreted as the expectation in probability theory.

$$H(\omega)=\int_{\mathbb{R}^d}P(\omega)\,e^{j\langle\omega,(x-y)\rangle}d\omega \tag{4}$$

Where $P(\omega)$ is a probability distribution if the kernel is properly scaled [11].

Therefore we can write as,

$$H(x-y)=E\left[e^{j\omega^Tx}e^{-j\omega^Ty}\right] \tag{5}$$

Where ω is sampled from $P(\omega)$.

The estimate can be improved by drawing random samples from the distribution $\omega_1, \omega_2, \ldots, \omega_D \sim P(\omega)$ [11]. Now the estimate is computed as the expectation of mean of samples.

$$H(x-y) = E\left[\frac{1}{D}\sum_{i=1}^{D} e^{j\langle \omega_i, (x-y)\rangle}\right] \tag{6}$$

Since both the probability distribution and kernel function are real, we can omit the imaginary part in the expansion of $e^{j\langle \omega_i, (x-y)\rangle}$. Hence, the kernel function can now be written as,

$$H(x-y) = E\left[\frac{1}{D}\sum_{i=1}^{D} \cos\left(\omega_i^T (x-y)\right)\right] \tag{7}$$

Since $\cos\left(\omega_i^T (x-y)\right) = \cos\left(\omega_i^T x\right)\cos\left(\omega_i^T y\right) + \sin\left(\omega_i^T x\right)\sin\left(\omega_i^T y\right)$

$$\cos\left(\omega_i^T (x-y)\right) = \begin{bmatrix}\cos\left(\omega_i^T x\right)\\ \sin\left(\omega_i^T x\right)\end{bmatrix} \bullet \begin{bmatrix}\cos\left(\omega_i^T y\right)\\ \sin\left(\omega_i^T y\right)\end{bmatrix} \tag{8}$$

And now the kernel function can be written as,

$$H(x-y) = \frac{1}{D}\begin{bmatrix}\cos\left(\omega_i^T x\right)\\ \sin\left(\omega_i^T x\right)\end{bmatrix} \bullet \begin{bmatrix}\cos\left(\omega_i^T y\right)\\ \sin\left(\omega_i^T y\right)\end{bmatrix} = \frac{1}{D} J(x) \bullet J(y) \tag{9}$$

Where, $J(x) = \frac{1}{\sqrt{D}}\begin{bmatrix}\cos\left(\omega_i^T x\right)\\ \sin\left(\omega_i^T x\right)\end{bmatrix}$ and $J(y) = \frac{1}{\sqrt{D}}\begin{bmatrix}\cos\left(\omega_i^T y\right)\\ \sin\left(\omega_i^T y\right)\end{bmatrix}$

RKS algorithm defines the maps feature vector to a higher dimensional space with this feature mapping operator.

Compact random feature map (CRAFTMaps)

CRAFTMaps is another explicit random feature mapping algorithm whose key idea is to apprehend the Eigen structure of the exact kernel space completely, and then represent it in a more compact form. Unlike RKS, CRAFTMaps algorithm approximates polynomial kernels of the form $\left(x^T y + q\right)^r$ with $q \in \mathbb{R}^+$ and $r \in \mathbb{N}_0$. CRAFTMaps algorithm computes the feature mapping in the following two steps: up projection and down projection.

In the up projection step, feature mapping function is defined as, $J : \mathbb{R}^d \to \mathbb{R}^D$, where d is input feature dimension and D is the dimension of projected space, $d < D$ and for all $x, y \in \mathbb{R}^D$, $\langle J(x), J(y)\rangle = H(x, y)$. Feature mapping is obtained by projecting x onto a set of random d dimensional vectors from standard Gaussian distribution, and then compute the dot product at the projected space. This up projection can be achieved using any of the well-known explicit feature mapping algorithms.

One of the main disadvantages of random feature mapping is rank deficiency. In order to nullify this effect, CRAFTMaps down projects the resultant feature vectors to a relatively lower dimension. Now the feature mapping operation is defined as, $F : \mathbb{R}^D \to \mathbb{R}^E$, $E < D$ and $\langle F(J(x)), F(J(y))\rangle \approx \langle J(x), J(y)\rangle$.

RESULTS

We conducted our experiment with 1999 KDD cup data set [12]. Data set contains 494021 data samples and 41 features. Among these data, 449785 are abnormal and remaining data is normal. The objective of our experiment is to predict whether the incoming data is abnormal or not. Since the distribution of data is not uniform, (i.e. 91.05% data are abnormal) we can employ One-class SVM algorithm for the classification. In the preprocessing step, all data are normalized to the range [-3, +3] and also we assume that training data are not linearly separable. Therefore, input feature vectors are mapped to a higher dimension where we can draw a classifier which can separate data linearly.

After normalizing the training data, CRAFTMaps algorithm is used for the explicit random projection and features are projected to a dimension 4 times the input feature dimension. Random projection of features is obtained by multiplying each feature vector with the product of a Hadamard matrix and a diagonal matrix whose elements are coming from {+1, -1} with equal probability. Dimension of space to which the features are projected, input feature dimension and the degree of the polynomial kernel determines the number of such multiplications for each feature. Here, the input feature size is $d = 41$, the dimension to which features are projected is taken as $D = 164$ and the degree of the polynomial kernel is fixed to $r = 2$. So 8 multiplications are required in this case ($T = (r * D / d)$).

In order to avoid the rank deficiency problem, resultant feature vectors are down projected to a lower dimension. This dimension depends on the degree of the polynomial kernel. After shuffling elements in resultant feature vectors randomly, we combine two adjacent elements from each feature vectors to generate new feature vectors. Now the dimension of features has reduced to 82. These obtained features are given as input to SVM for classification. Training model is created by adjusting certain parameters in the toolbox. Test data also first normalized and then mapped to a relatively manageable higher dimension.

These computations are performed on 4 parallel machines and the final output is obtained by assembling the results from the parallel machines. Total data (80000 samples) were divided into 4 batches and each batch of 20000 data was given to each processor for prediction. Prediction times from all four machines are observed and total prediction time is fixed as the maximum time taken by a single processor. Even though four machines are of same configuration, speed of prediction depends on many external as well as internal factors.

Despite implicit mapping of features gives 100% prediction accuracy, the time taken to predict the malicious activities in the network is huge. Explicit random feature mapping improves the time complexity to a great extent by allowing a small percentage of error. Application of CRAFTMaps algorithm also reduces computational complexity by avoiding the underutilized spaces in the featured space.

[Table-1] explains the time taken by each machine for the prediction. Four machines gave different prediction time and the time taken for detecting the abnormal activities in the network is taken as 0.32 sec, which is the maximum time taken by a single processor. Total prediction accuracy is taken as the average of accuracies obtained from each machine. Here we get 98% accuracy in detecting malicious activities from a set of 80000 data, which is equivalent to the state of the art accuracy.

Table: 1. Prediction time of parallel machines and prediction accuracy

Machine	Prediction time	Prediction accuracy
P1	0.29 sec	98%
P2	0.32 sec	98%
P3	0.29 sec	100%
P4	0.31 sec	96%

CONCLUSIONS

Identification and detection of malicious activities in a network in real time is a difficult task because time and storage requirement for real time prediction is huge. So detection of abnormal activities from a streaming network data with SVM and explicit random feature mapping algorithms (say CRAFTMaps) reduces the time requirement. Also when the prediction is done in parallel, the speed can be improved significantly. CRAFTMaps algorithm was used for explicit mapping and obtained 98% accuracy in 0.32 seconds for 80000 data. Distributive processing of data improves the prediction of malicious activities from streaming network logs by a great margin.

CONFLICT OF INTEREST

None

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

None

REFERENCES

[1] Hand, David J, Heikki Mannila, and Padhraic Smyth. [2001]Principles of data mining. MIT press.

[2] Cortes, Corinna, and Vladimir Vapnik. [1995]Support-vector networks. Machine learning 20(3): 273-297.

[3] Li Wenye, Kin-Hong Lee, and Kwong-Sak Leung. [2006] Generalized regularized least-squares learning with predefined features in a Hilbert space, Advances in neural information processing systems.

[4] Angelov, Plamen P, and Xiaowei Zhou. [2008] Evolving fuzzy-rule-based classifiers from data streams." Fuzzy Systems, *IEEE Transactions on* 16.6: 1462-1475.

[5] Rahimi, Ali, and Benjamin Recht.[2007] Random features for large-scale kernel machines. Advances in neural information processing systems.

[6] Rahimi Ali, and Benjamin Recht.[2009] Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. Advances in neural information processing systems.

[7] Le, Quoc Viet, Tamas Sarlos, and Alexander Johannes Smola. [2014] Fastfood: Approximate Kernel Expansions in Loglinear Time. arXiv preprint arXiv:1408.3060 .

[8] Hamid, Raffay, et al. [2013]Compact random feature maps." arXiv preprint arXiv:1312.4626 .

[9] Charu C Aggarwal. [2006] Data Streams: Models and Algorithms (Advances in Database Systems). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[10] W Rudin. [1994]Fourier analysis on Groups. Wiley Classics Library. Wiley-Interscience, New York, reprint edition.

[11] von Tangen Sivertsen, Johan.[2014] Scalable learning through linearithmic time kernel approximation techniques.

[12] Lichman M. [2013] UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[13] Kumar S, Sachin B, Premjith M, Anand Kumar, and KP Soman.[2015] AMRITA_CEN-NLP@ SAIL2015: Sentiment Analysis in Indian Language Using Regularized Least Square Approach with Randomized Feature Learning." In Mining Intelligence and Knowledge Exploration, pp. 671-683. Springer International Publishing.

ABOUT AUTHORS

<b>Prabaharan Poornachandran</b> has two decades of experience in Academia and Industry, Currently serves as Assistant Professor at Amrita University and Principal Investigator for large scale Security projects. His current area of research includes, Big-data Security Intelligence, Cyber-Physical systems security, Machine learning for Security, Complex Binary analysis, IoT, SCADA and Hardware security, Application & Network security, Advanced Forensics and Incident handling etc.

***Premjith B** is currently pursuing PhD in Center for Computational Engineering and Networking, Amrita University. His research area includes Natural Language Processing and Machine Learning.*

***K P Soman** currently serves as Head and Professor at Center for Computational Engineering and Networking (CEN), Amrita Vishwa Vidyapeetham, Coimbatore Campus. Further info on his homepage: https://www.amrita.edu/faculty/soman.*

ISSN: 0976-3104
REGULAR ISSUE

THE IIOAB JOURNAL

CASE STUDY OPEN ACCESS

SURGICAL MANAGEMENT OF CHRONIC BUCCAL FISTULA IN A CAMEL (*CAMELUS DROMEDARIUS*): A CASE STUDY

Bajrang Lal Kaswan[1], **Saraswat Sahoo**[1], **Subha Ganguly**[2]*

[1]*Teaching Veterinary Clinical Complex, Arawali Veterinary College, Rajasthan, INDIA*

[2]*Department of Veterinary Microbiology, Arawali Veterinary College, Rajasthan, INDIA*

ABSTRACT

A 6 years old female camel was presented at Arawali Veterinary College, Sikar with history of escape of partially masticated feed materials from the opening at right side cheek from last one year. It was diagnosed as buccal fistula and it was surgically repaired with a rectangular hard leather piece under xylazine sedation. It took nearly 5 weeks for complete recovery under monitoring.

Received on: 6th June 2016
Revised on: 7th June 2016
Accepted on: 8th June 2016
Published on: 26th-July-2016

KEY WORDS

Buccal fistula, Camel, Xylazine

***Corresponding author: Email:** ganguly38@gmail.com **Tel:** +91 9231812539

INTRODUCTION

Camel suffers from various surgical affection of head region such as mandible fracture, soft palate injury, buccal fistula, lacerated eyelid, ruptured eyeball, corneal opacity and lacerated nostril because camel browsing the upper storey tree vegetation and the thorny vegetation from shrubs and bushes, which often inflict injuries to the head region mostly eyes and lips are involved [1]. Buccal fistula is also reported by Gahlot [2], Patel et al. [3] and Purohit et al. [4] in camel. Gahlot [5] reported that a careful per oral examination revealed absence of 3rd cheek tooth where the feed straw got accumulated during mastication and these straws repeatedly injured the oral mucosa at this level leading to buccal fistula and or buccal cum salivary fistula. Therefore feed straw should be removed from the wounds of buccal fistula.

CASE HISTORY AND OBSERVATION

A 6-year old female camel was presented to the Teaching Veterinary Clinical Complex (T.V.C.C.) of Arawali Veterinary College, Sikar with a history of escape of partially masticated feed materials from the opening at right side cheek from last one year. The camels had normal appetite, but showed some irritation at the time of feeding. Owner reported that it was previously treated and sutured by a local para-veterinarian but suture broke down the same day and wound healing did not occur.

The clinical examination revealed 3 cm chronic wound was found about 2 cm below the lower eye lid on right side of cheek. Partially masticated feed and watery fluid was coming out through the openings at right side cheek [Figure- 1].

Careful clinical examination revealed that small pocket was present inside oral cavity and partially masticated feed material was stored inside pocket. These partially masticated feed such as straw repeatedly injured the oral mucosa at this level.

TREATMENT AND DISCUSSION

Camel was secured in sternal recumbence and fistula was repaired under xylazine sedation. Xylazine @ 0.3 mg/kg body weight administered intramuscularly and local infiltration of 2% Lignocaine was made. Partially masticated feed material was recovered from the buccal fistula along with pocket by help of allies forceps and pocket was emptied. The fistula was debrided. One soft circular leather piece of size slightly greater than diameter of fistula was placed on inner oral mucosal opening along with thread which was come out through buccal fistula opening **[Figure- 2].** The wound edge was freshened with B.P. blade to improve vascularity. Buccal fistula was repaired with catgut no. 2 and skin was sutured with silk thread. Another rectangular hard leather piece of size slightly greater than diameter of fistula was placed on outer skin opening of fistula and knot was secured on the outer hard leather piece **[Figure- 3]**.

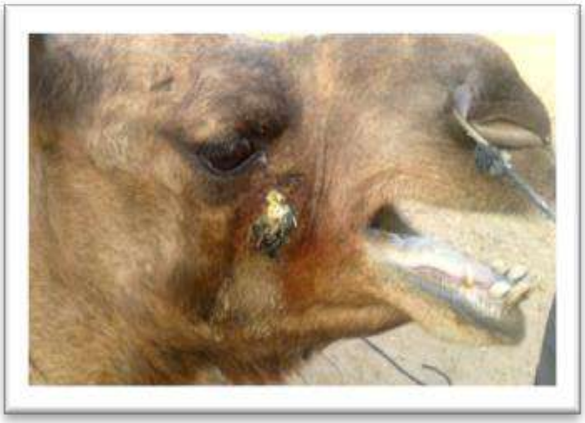


Fig:1. Buccal fistula and partially masticated feed coming out through the opening in camel.

………

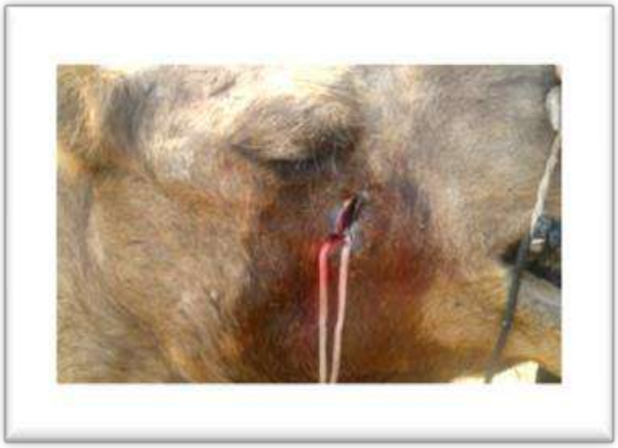


Fig:2. One soft circular leather piece of size slightly greater than diameter of the fistula.

………

Oral cavity was irrigated daily with light potassium permanganate solution. On the day of surgery animal was kept on fluid therapy 5 liters of Dextrose Normal Saline and 5 l Ringer's Lactate. Oxytetracycline 2500 mg administered intravenously for 7 days and Phenylbutazone 3000 mg, intramuscularly for 3 days, postoperatively.The animal was offered soft dry leaves of KHEJARI tree with minimum wet straws from the second day of surgery. The diameter of fistula reduced remarkably in three week time and leather pieces removed after 4 week. The fistula dressed with 5% Povidone iodine and Charmil spray. Complete healing took place in 5 weeks.

Peculiarities of anatomy and physiology of head of camels are such as mandible, soft palate and long neck predisposes the animal. Gahlot and Chouhan [5] reported that camels suffer from variety of surgical affections of head and neck region which markedly affect the value of the animal, draft capacity and overall performance. The results of the present study were in agreement with the observations of Kumar [1], Gahlot [2] and Purohit et al. [4].

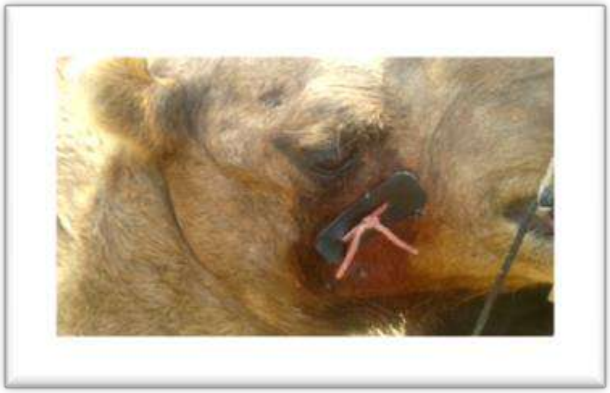


Fig:3. Rectangular hard leather piece of size slightly greater than diameter of fistula was placed on outer skin opening of fistula and knot was secured on the outer hard leather piece.

....

CONCLUSION

Careful clinical judgement, early surgical management with gentle handling of tissue, aqua therapy or through wound irrigation, effective topical medication and sufficient rest bring quick and better recovery in clinical and surgical wounds in the camels

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENTS

The authors are thankful to Hon’ble Dean and Management (Hony. Chairman and Secretary, Aastha Society, Sikar) of Arawali Veterinary College for providing the necessary facilities to conduct the case study.

FINANCIAL DISCLOSURE

None declared.

REFERENCES

[1] Kumar P. [2013] A Clinical Study on Surgical Affections of Head and Neck Region of Camels (Camelus dromedarius). M.V.Sc. Thesis submitted to Rajasthan University of Veterinary and Animal Sciences, Bikaner.

[2] Gahlot TK. [2000] Surgery of the dromedary camel. In: Selected Topics on Camelids. 1st ed., The Camelid Publishers, Bikaner, India. pp. 378-430.

[3] Patel SS, Parikh PV, Patil DB, Kelawala NH, Patil VN, Jhala SK. [2007] Survey of surgical affections in camels 1996–2007. Camel Conf-Book, International Camel Conference. Feb. 16–17, College of Veterinary and Animal Science, Bikaner. pp. 73.

[4] Purohit S, Chaudhary SR, Mistry J, Patel PB Siddiquee GM, Patel JS. [2011] Surgical management of buccal fistula in a camel (Camelus dromedarius). *Journal of Camel Practice and Research*, 18(2):345-346.

[5] Gahlot K and Chouhan DS. [1992] Camel Surgery (1st ed.), Gyan Prakashan Mandir, Bikaner, India, 30-155.

ISSN: 0976-3104
REGULAR ISSUE

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

PREDICTION OF CREDIT RISK EVALUATION USING NAIVE BAYES, ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE

Lohit Mittal*, **Tarang Gupta, Arun Kumar Sangaiah**
School of Computing Science and Engineering, VIT University, Vellore, TN-110003, INDIA

ABSTRACT

An accurate prediction of credit risk evaluation is very useful for the banking and financial industry in minimizing the risk in lending credit to the customer and decreasing the chances of making wrong decision. As the increase in customers and emergence of different trade, it is difficult for bank management to analyze individual physically hence data mining algorithms are implemented in order to reduce the work effort by the bank management. This study attempted to implement three data mining model and compared their performances in predicting the risk in giving credit to the customer. In this study neural network based on back propagation, naive Bayes algorithm and support vector machine were implemented. Eight technical indicators were used as input for the above models. Data preprocessing were done to increase the performance of their prediction. A comparative analysis of the models was carried out and experimental results showed that the performance of Support Vector Machine (92%) was higher than the other two. Naïve Bayes (87%) performance was found little better than that of Artificial Neural Network (85%). Findings from this study can help in improvement of the existing system.

.

Received on: 5th-May-2016
Revised on: 22nd -June-2016
Accepted on: 29th -June-2016
Published on: 28th -July-2016

KEY WORDS

Support Vector Machine; Artificial Neural Network; Binary Classification; Prediction; Credit

***Corresponding author: Email:** lohitmittal22@gmail.com, **Tel:** +91-9159065928

INTRODUCTION

The credit provided by the bank is one of the important features of the banking industry. It is the main source of income for the banks. A credit risk is defined as a failure in returning the loan by the borrower to the lender. Due to increase in globalization and opening of various business opportunities, many small scales or large scale firms require credit to invest in their business or sometimes credit is required in situations like purchasing new house, vehicle or any other necessary expensive stuff. Sometimes a loan is needed for education. As the foreign study rates are very high, students tend to apply for educational loans. In the recent years, banks are facing crises in financial sector as the risk in lending credit to the borrowers is increasing at alarming rates.

As borrower credit risks can not only be determined by the assets the borrower posses during the application of loan, there are several other factors also responsible which can determine the credit risk. As some of the bank handle customers overseas and exponential increase in borrowers it is very difficult for the bank management to analyze the borrowers. Hence banking industry requires a credit risk evaluation system to filter the borrowers' base upon some of their characteristics and with precise accuracy to reject their loan application at preliminary stage to reduce the risk in lending credit .In this study the factors taken under consideration are age, years in service, total income (Per annum), number of dependent, tenor, cash flow, collateral and credit history. A credit score is a formulated expression used to analysis of a person credit file. Banks and lenders uses credit score to determine potential risk possess by creditor while taking credit or loan from lender. The study done is to analysis, compare and study different data mining algorithms involved in classification of the above mentioned factors to determine whether there is a credit risk or not. In order to achieve this goal we have undertaken an assessment, we have chosen three methodologies.

Artificial Neural Network (ANN)

The study on ANN dates back to 1970s by Frank Rosenblatt(1958) [1] who worked on first perceptron algorithms and the outcome of the study was used to develop smart automated software and systems. ANN has been found to work smartly and give promising results [2-7]. Neural network received a boost in its processing after the publication of machine learning by Marvin Minsky and Seymour Papert [8]. Odom and Sharda (1990) [9] were the first to apply NNs in the credit risk evaluation. The early network was based on Hebb network which aims at updating the input vector then perceptron comes into picture which increase the accuracy and was base for many models. After perceptron neural backpropogation comes into picture which was developed by David E.Rumelhart and James McClelland [10]. Backpropogation has the feature to update its weights by keeping the history. It is known as neural processing.But in late 2000 a deep learning created more interest. The first study on credit risk evaluation was done by Angelini et al. 2008 [11] in which the banking management could calculate capital requirement using key risk drivers. An application of ANN by Mohsen et al. 2013 [12] was presented for calculating the variables required for credit risk evaluation.

Support Vector Machine (SVM)

Study using Support Vector Machine has been proven to be promising and efficient in determining credit risk evaluation when compare to other algorithms such as neural network, particle swarm optimization (PSO), and other machine learning classification algorithms. Several SVM based studies for credit risk evaluation problems have been proposed by Ahn et al., Wu et al. Zhang et al and others [13]. The results have been promising and their main work was to compare with genetic algorithms. Others important approaches include combining SVM and fuzzy logic. This fuzzy logic work was attemted by Hao et al [14] who used fuzzy sets whereas Huang et al implemented least square SVM. Xuchuan et al who used particle swarm optimization for optimal parameter selection for SVM. A comparative research by these authors proved that SVM classifier has the potential to replace other algorithms in complexity, speed and reliability.

Naïve Bayesian

The study Naïve Bayesian has been based on Appling Bayes theorem for assuming between the features. When compare to other algorithms such as neural network, particle swarm optimization (PSO), and other machine learning classification algorithms, Naïve Bayesian is proven to be promising and efficient in determining credit risk evaluationIt has more advanced methods as compared to Support Vector Machine. Naïve Bayesian has used in medical diagnosis.Russell and Norvig was the first to study about Naïve Bayesian and they have mentioned in their first book. Rish, Irina in 2001who worked on an empirical study of the Naïve Bayesian. In 2003 Rennie, J.; Shih, L.; Teevan, J.; Karger, D. worked on tackling the poor assumption of Naïve Bayesian classifiers [15-20].

MATERIALS AND METHODS

Research Data

This section describe about the data set on which the research has been carries out and the attributes used for prediction. The data set was collection from Resort Savings and Loan Plc, Lagos Island, Lagos, Nigeria. A total of 200 records are taken and accessed. As there were some missing values, the dataset required some preprocessing. So replacing the missing value with the global mean of the same attribute method was adopted. The data set was again preprocessed as it needed to be converted to binary data set where the output can be either 1 or 2. This 1 or 2 represent the label rather than their usual value, where 1 represents credit risk and 2 represents credit no risk. The output given by is in range of 0 to 1. As given by , the output is marked 2 if value is greater than 0.75 and 1 if the output is lower than 0.75, this result in a binary data set as shown in **Table-2.**

There are eight predictor attributes in the data set that are age, years in service, total income (Per annum), number of dependent, tenor, cash flow, collateral and credit history. To increase the accuracy of the output the eight data sets are converted to three categories as shown in **Table-1**. The classification is a necessary preprocessing as to increase the accuracy of all the three algorithms. Many financial officials approved those eight to be the important factors to be responsible for prediction of credit risk evaluation process.

Here Low is assigned class label 1, Moderate is assigned class label 2 and High is assigned class label 3.Hence the preprocessing of data set is done. Further there is normalization of data is done for artificial neural network as the output of neural network is in the range of 0 to 1, hence we need to normalize the value between 0-1 of the output variable.0.2 is chosen as value corresponding to class risky and 0.8 is chosen for class non risky.

Table :1.input variable numerical value range

CODE	Variable Description	Low	Medium	High
C1	AGE(years)	45-60	33-45	18-33
C2	LENGTH OF SERVICE	18-30	10-18	1-10
C3	TOTAL INCOME	1-1.2x10^7	1.2x10^7-2.2x10^7	>2.2x10^7
C4	NO. OF DEPENDENT	>5	3-4	1-2
C5	TENOR	240-360	120-240	1-120
C6	CASH FLOW	1	2	3
C7	COLLATERAL	1	-	2
C8	CREDIT HISTORY	1	-	2

Table: 2. Conversion into binary data set

S.no	Variable Description	Class	Credit
1	>=0.75	1	Not-Risky
2	<=0.75	2	Risky

Table: 3. Summary statistics for the selected indicators

	Max	min	Medium	S.D
AGE(years)	57	27	36.6	6.928
LENGTH OF SERVICE	31	2	9.235	6.65
TOTAL INCOME	9076832	56789	1544842	1252815.63
NO. OF DEPENDENT	6	1	3.145	1.369
TENOR	360	45	224.7	83.6
CASH FLOW	3	1	2.13	0.829
COLLATERAL	2	1	1.75	0.434
CREDIT HISTORY	2	1	1.9	0.280

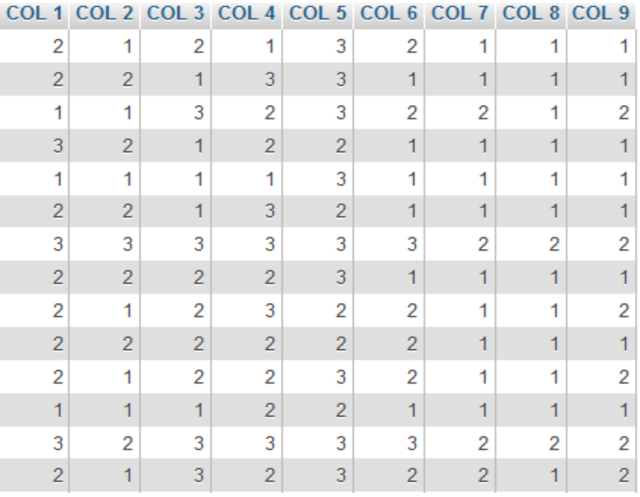
The given data set was normalized and converted. The data was stored in a CSV (comma delimiter) and was stored in mysql database using PHP Myadmin for naïve Bayes and artificial neural network. The data set is derived using java JDBC driver and loaded for further functions. The data set have been divided into set of four on the bases of age. 60 % of each data set is taken for training and the rest is for the holdout. The weights are calculated using training and tested on the holdout datasets.

PREDICTION MODEL

Bayesian Network

The naive Bayesian classification gives the class label of a data which is their in the table, the values of the entity and their attribute are predicted to be conditionally independent of one another. Bayesian classification is the statistical classifiers and every new data which we get has belongs to a class. The method used in Bayesian classifier are joint conditional probability distributions, which allows class conditional independencies to be work between subsets of variables and also it have a graphical model of relationships, by which various interpretation data is performed. The two categories of belief network are first one is a directed acyclic graph and second one is a set of conditional probability tables. Each and every node in directed acyclic graph shows a variable and it is a random variable which is either in the form of discrete or in the form of continues. The attribute can be real which are given in the data or they can be invisible variables and are believed to form a relationship. In this directed acyclic graph each and every arc represents dependence probability. An arc is made from a point C to a point D, then C is a parent D, and D is a descendant of C. Every variable in the graph is independent of its non-descendants, gives its parents.

Bayes theorem formula is given by,



COL 1	COL 2	COL 3	COL 4	COL 5	COL 6	COL 7	COL 8	COL 9
2	1	2	1	3	2	1	1	1
2	2	1	3	3	1	1	1	1
1	1	3	2	3	2	2	1	2
3	2	1	2	2	1	1	1	1
1	1	1	1	3	1	1	1	1
2	2	1	3	2	1	1	1	1
3	3	3	3	3	3	2	2	2
2	2	2	2	3	1	1	1	1
2	1	2	3	2	2	1	1	2
2	2	2	2	2	2	1	1	1
2	1	2	2	3	2	1	1	2
1	1	1	2	2	1	1	1	1
3	2	3	3	3	3	2	2	2
2	1	3	2	3	2	2	1	2

Fig:1.Sample data after conversion

Let assumed that there is a sample called A, the probability of all the possible events h, P(h|A) follows the Bayes theorem stated mathematically as the following equation

$$\text{P (A|B)} = \frac{P(B \mid A)P(A)}{P(B)}$$

Bayesian classification has been develop which is most optimal one. When the network data and its topology is given in th data the various variables used in the sample is known then training the network is candid. It is used to find the continuous probability table (CBT) entries. This is analog to the way of finding the computing probability which is their in naιve Bayesian classification.
In Bayesian network, Naïve Bayesian classifier is also used in which we assume that attributes are conditionally independent.

$$p(C_K \mid x_1 ......x_n) = \frac{1}{Z} P(C_K) \prod_{i=1}^{n} p(x_i \mid C_K)$$

This is naïve hypothesis. Naive Bayesian classifier is very effective in case of cost; it reduces the calculation cost It is best for the problems where there is strong relation between the variables and the data given.

Support Vector Machine

SVM was developed in COLT-92 by Boser, Guyon, Vapnik[10] and having several application like bioinformatics, text handwriting recognition etc. Support vector machines (SVMs) are a way of classification of both linear and nonlinear data, that is, an SVM is an algorithm that works as follows.

1. It uses a nonlinear mapping to convert the given training dataset into a poly-dimensional.
2. In the same dimension, it identify for the linear optimal separating hyperplane (i.e., a “decision boundary” separating the tuples of one class from another). Having an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane.
3. The SVM finds this hyperplane using support vectors (“essential” training tuples) and margins (defined by the support vectors).The task of this type of algorithm is detect complex pattern in data by various data mining approaches like clustering, classifying, ranking, cleaning etc.
4. SVM is a particular instance of kernel machine has large class of learning algorithms. The class of possible patterns are defined by various classes of kernel methods implicitly by introducing a notion of similarity between data and this methods exploits the information about the inner products between the data items.
5. There are two cases in SVM which are case when the data is linearly separable and the other is when the data is non-linearly separable.
6. Although the training time of even the fastest SVMs is slow, but they are highly accurate, have the power to solve complex nonlinear decision boundaries. They are much less prone to over fitting than other methods.
7. SVMs can be used for numeric prediction as well as classification. They have been applied to a number of areas, including handwritten digit recognition, object recognition, and speaker identification, as well as benchmark time-series prediction tests.

ISSN: 0976-3104

The various properties used by the SVM algorithm are as follows:-

1. Duality is the first property of Support Vector Machines in which SVM is represented as Linear Learning Machines in a dual fashion and the data appear only within dot products. But there are many limitation in the first feature of SVM so to minimize this limitation, one theorem was introduced called as Mercer's theorem which solve the symmetric positive definite.

   Linear learning Machines the second feature of SVM that uses dual representation concept which operates in a kernel induced space as a linear function. Various other algorithms like clustering, PCA can also be used by which dual representation often possible (in optimization problems, by Representer's theorem).The various generalization problems which can be solved in SVM are:

a. The effect of dimensionality in which it is very easy to over fit in high dimensional space.

b. The SVM problem of finding one hyperplane that separates the data: many such hyperplanes exist)

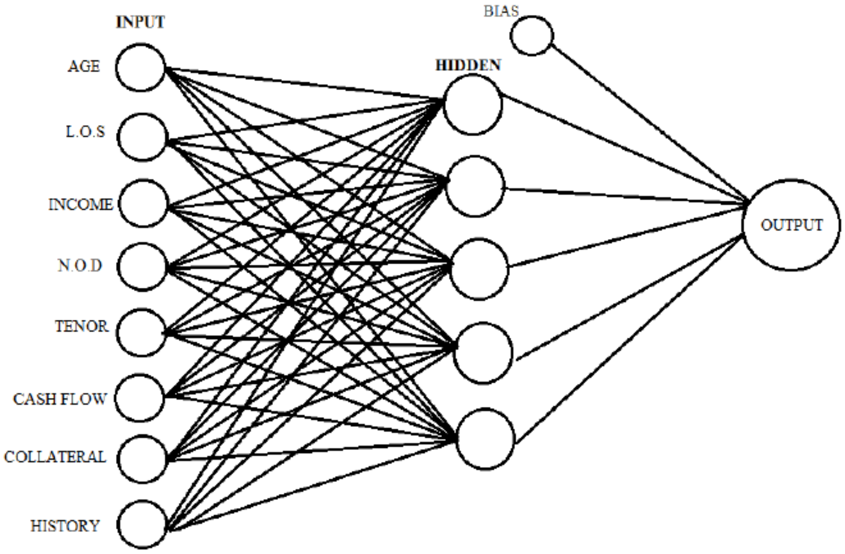
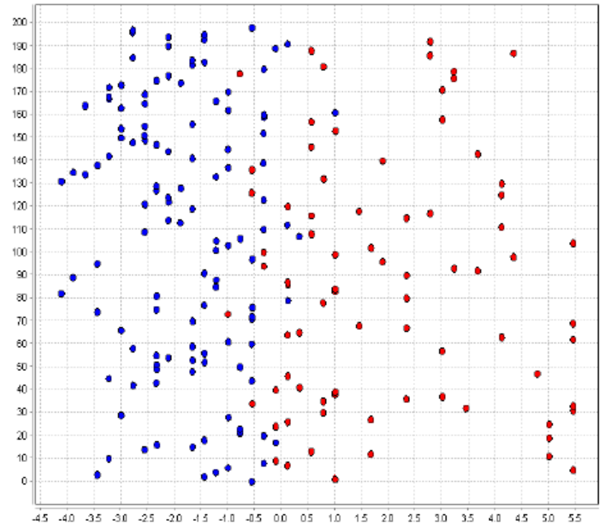


Fig: 2. Architechture of ANN

……..



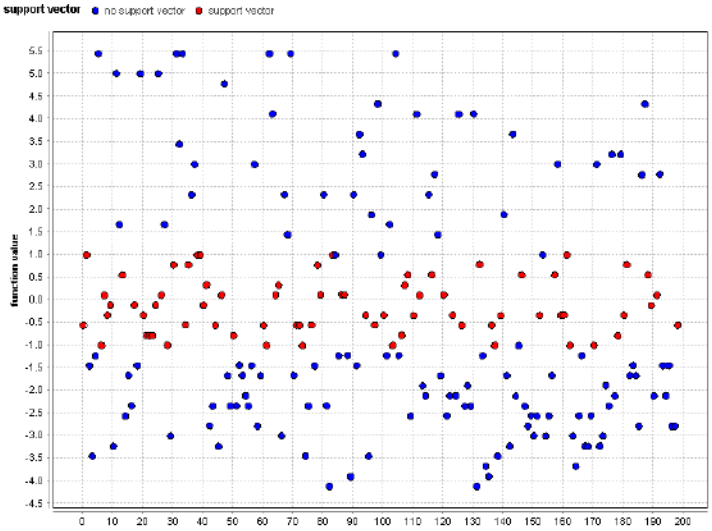


Fig: 3. SVM polynomial output

Fig: 4. SVM radial output

……..

Artificial Neural Network

Artificial Neural Network is an algorithm or a system used to calculation purpose, work on various methods, and to study the various problems related to biological field. Artificial Neural Networks are details processing algorithms which are built and to model the human brain. The main objective of the neural network research is to make a calculation system for designing the experts system to perform calculation faster.

Various tasks performed by ANN such as pattern-matching, classification, optimization of function, approximation, vector quantization, data clustering. ANN was introduced by biological nervous system such as brain processes information. The net input to the neuron u is given by the formula:

$$v = \sum_{j=1}^{m} w_j x_j$$

The activation function is applied to yin to compute the output and the weight represents the strength of synapse connecting the input and the output neurons. The weights may be positive or negative. The positive weight means the synapse is excitatory and the negative weight means the synapse is inhibitory.

Properties of ANN are:-

1. The speed or cycle time of execution of ANN is of few nanoseconds.
2. The processing time of ANN is very fast and it can perform several parallel operations simultaneously.
3. The size and complexity of ANN depends upon the chosen application and the network designer
4. The ANN stores the data in its contiguous memory locations and in ANN the loading may sometimes overload the memory
5. The main property of ANN is learning and learning. There are two kinds of learning in ANN
    1. Parameter learning: It modernize the weight which are linked in the network.
    2. Structure Learning: It concentrate on the topology of the network and check whether any change in the network.
6. learning can be :
    1. Supervised learning: Learning through a teacher/ supervisor
    2. Unsupervised learning: Learning in the absence of supervisor.
    3. Reinforcement learning: Learning basing upon a critic information and it is similar to supervised learning

ANN also has activation function which is used to calculate the exact output. The activation function is applied over the overall input to find the overall output of an ANN. There are many activation function like identity function, binary step function, sigmoidal function, bipolar sigmoid function, hyperbolic tangent function etc

RESULTS AND DISCUSSION

Experimental Results

The Naïve Bayes algorithm and Artificial Neural Network with back propagation was implemented in Java Servlet Package, JDK 8.0 and the data set was uploaded to MySQL as a CSV file whereas for the Support Vector Machine Rapid Miner Tool was used .At initial stage the data set was divided into four sets. The age was chosen to determine the dividing factor as young (18-32), middle (32-35), semi-old (36-44), and old (45-60).At initial stage of ANN the best combination of the four parameters needs to be determined that is number of epoch, learning rate, number of neurons and mc. The best combination is observed and collected results are presented in **Table-4** and **Table-5**. It should be taken care that not only parameter combination but they need to be applied and checked for best accuracy. Now with fixation of three combinations (ep; mc ;n) as shown in **Table-4** , learning rate is iterated and a plot of different accuracy is shown in**[Figure-1]**as a non-linear graph is obtained. As seen from figure the best performance was obtained at lr = 0.2. The average accuracy observed is 85 percent. Next is Support Vector Machine, it is also applied on four sets and average performance is calculated. First SVM polynomial is applied and then radial SVM is applied. In polynomial SVM is iterated for combination of three parameter (d; y; c) as shown in **Table-6**. In polynomial SVM the results obtained were 87 percent whereas in case of radial basis the results obtained were around 76 percent. In radial basis two parameter combinations were iterated as shown in **Table-7**. The average accuracy of polynomial is taken for comparison. Next is Naïve Bayes algorithm. In this algorithm is applied to four sets in one process as seen in **Table-8** and then the algorithm is applied to the whole dataset as a lump sum and a confusion matrix is created , as it can be seen that the best iteration results have been taken for comparison.

TABLE: 4. ANN Holdout Accuracy Parameter combination (ep;mc;n)

AGE(years)	(5000;0.7; 5)		(5000;0.7; 5)		(5000;0.7; 5)	
	Training	Holdout	Training	Holdout	Training	Holdout
18-32	98	86.3	98	83.4	98	83.2
32-35	99	85.4	99	84.1	99	84.2
36-44	98	85.5	98	83.2	98	80.2
45-60	98	86.3	98	85	98	85.5
AVERAGE	98.5	85.4	98.5	83.9	98.5	83.27

Table: 5. Best three combination of ANN model

No.	Lr	Ep	Mc	n	Training	Holdout	Average
1	0.1	5000	0.7	5	98.18	86	93
2	0.1	7000	0.1	6	98.54	85	92
3	0.1	6000^7	0.4	7	98.18	85.6	92.3

As naïve Bayes algorithm doesn't have any parameters hence there is no adjusting need to be done. The Algorithm's efficiency increases with the increase of the data set tuples. Finally all the algorithms accuracy has been compared as shown in **Table-10**. It can be seen that all the algorithms give a promising result in prediction of given problem.

Table: 6. Prediction performance (%) of polynomial SVM model

AGE(years)	(3;2.6;100)		(3;2.5; 100)		(3;3.2; 100)	
	Training	Holdout	Training	Holdout	Training	Holdout
18-32	99	92.3	99	90.5	99	90.1
32-35	98	91.6	98	91.6	98	90.3
36-44	97	92.0	97	91.0	97	92.1
45-60	99	91.8	99	90.3	99	92.2
AVERAGE	98.2	92.1	98.2	91.1	98.2	92.8

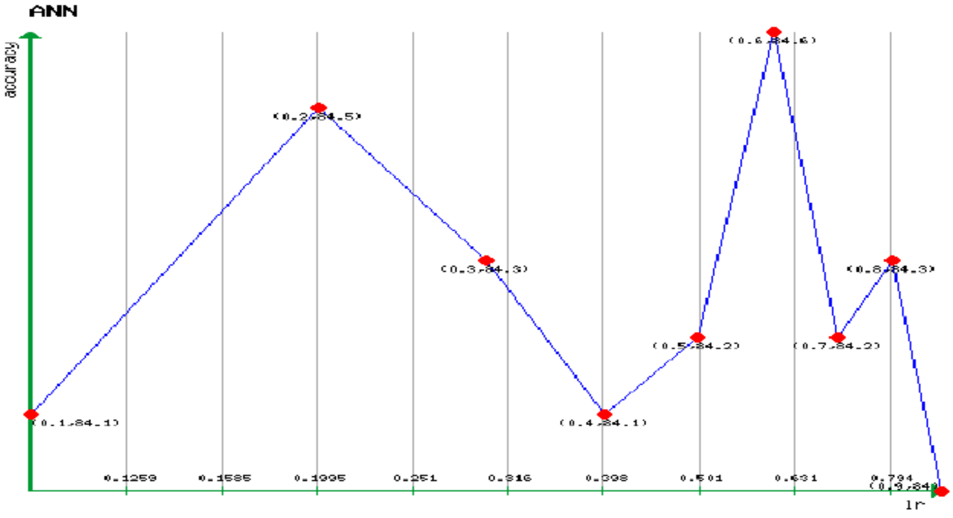


Fig: 5. accuracy vs. lr

Table :7 .Prediction performance (%) of radial basis SVM model

AGE(years)	(2.5;100)		(3.0; 100)		(3.1; 100)	
	Training	Holdout	Training	Holdout	Training	Holdout
18-32	99	78.2	99	75	99	77.2
32-35	98.3	76.4	98.3	76.5	98	77.1
36-44	97.8	79.0	97.8	77.3	97	78.3
45-60	98.4	77.3	99	75.2	99	76.1
AVERAGE	98.3	77.7	98.5	76	98.2	77.2

Table: 8. Naïve Bayes Accuracy

Age	Training	Holdout
18-32	100	87.2
32-35	100	89.1
36-44	100	86.4^7
45-60	100	87.5
AVERAGE	100	87.5

Table: 9. Naïve Bayes Confusion Matrix

Confusion Matrix	risky	non risky
risky	78	5
Non risky	10	87

Table: 10. Final Comparison

Methodology	Accuracy
ANN	**83%**
SVM	**92%**
Naïve Bayesian	**87%**

The novelty of the work is to make comparison among machine learning algorithm and to determine the best which we obtained was SVM as it more efficient in classification than the other two, although the other algorithm were slightly less they can also be used to determine the credit risk evaluation. The results obtained can be base for implementation of a system for credit risk evaluation.

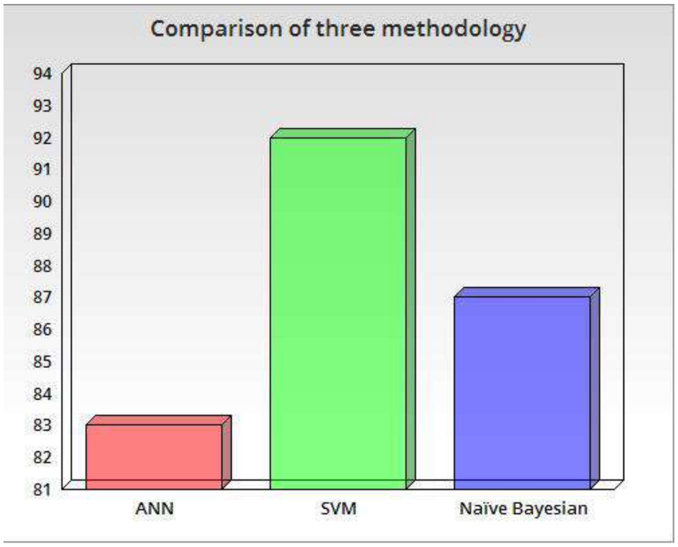


Fig: 6. Results of ANN, SVM, and Native Bayesian

………..

CONCLUSION

Prediction of credit evaluation decision is necessary, it helps the bank management to enhance their decision making and reduce their losses by taking right decision and also filtering loan application at preliminary stage in order to reduce the work overhead. This task is difficult and requires high skills. The study attempted to predict the credit risk of an individual and compare and analyze three algorithms in doing so. Based on the experimental results obtained, some important results can be drawn such as all the three algorithms are capable of producing significant accuracy during classification. Thus we can say all the three are useful tool for prediction for other problems also. The best prediction was found by SVM (92%) followed by Naïve Bayes (87%) and last ANN(83%) **Figure-6**, however ANN performance can be increased by performing fuzzy ANN mentioned in the literature. The efficiency of the Naïve Bayes algorithm can be increased by the grouping of elements. The study aims at comparison of the above three algorithm in terms of their accuracy and the best can be used for implementation of credit risk evaluation as shown above.

CONFLICT OF INTEREST

The authors declare no competing interest.

ACKNOWLEDGEMENT

We are thankful to VIT University for providing necessary resources for successfully implementation of this system

FINANCIAL DISCLOSURE

None declared.

REFERENCES

[1] Rosenblatt F. [1958] The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain". *Psychological Review* 65 (6): 386–408.doi:10.1037/h0042519. PMID 13602029.

[2] Sangaiah AK, Thangavelu AK., Gao XZ, Anbazhagan N, Durai MS. [2015a]An ANFIS approach for evaluation of team-level service climate in GSD projects using Taguchi genetic learning algorithm. *Applied Soft Computing* 30: 628-635.

[3] Sangaiah AK., Gao XZ, Ramachandran M, Zheng X.[2015b] A fuzzy DEMATEL approach based on intuitionistic fuzzy information for evaluating knowledge transfer effectiveness in GSD projects. International *Journal of Innovative Computing and Applications* 6(3-4):203-215.

[4] Huang Z, Chen H, Hsu CJ, Chen WH, Wu S. [2004] Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems* 37(4): 543-558.

[5] Huang W, Lai K., Nakamori Y, Wang S. [2004] Forecasting foreign exchange rates with artificial neural networks: A review. *International Journal of Information Technology*, 3: 145-165. DOI: 10.1.1.121.8174

[6] Jang JS, Sun CT. [1993] Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE Transactions on Neural Networks* 4 (1): 156-158

[7] Hawley DD, Johnson JD, Raina D. [1990] Artificial neural systems: A new tool for financial decision-making. *Fin Anal J* 46: 63-72. DOI: 10.2307/2F4479380 Huang

[8] Minsky M, S Papert. [1969] An Introduction to Computational Geometry. MIT Press.ISBN 0-262-63022-2

[9] Odom M, Sharda R. [1990] A neural network model for bankruptcy prediction. Proceedings of the International Joint Conference on Neural networks, 163-168.

[10] Rumelhart DE, James McClelland. [1986] Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge: MIT Press.

[11] Angelini E. di Tollo, G Roli A. [2008]. A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, 48 (4): 733-755

[12] Mohsen N, Mojtaba A. [2013] Measuring credit risk of bank customers using artificial neural network. *Journal of Management Research* 5(2):17. Mulhim

[13] Ahn H, Lee K, Kim K.[2006] Global Optimization of Support Vector Machines Using Genetic Algorithms for Bankruptcy. *Lecture Notes in Computer Science* 4234(5):420-429. Springer.

[14] Wu C, Tzeng G, Goo Y, Fang W.[2007] A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Systems with Applications* 32: 397-408.

[15] Zhang D, Chen Q, Wei L.[2007] Building Behavior Scoring Model Using Genetic Algorithm and Support Vector Machines. *Lecture Notes in Computer Science* 4488:482-485.

[16] Hao PY, Lin MS, Tsai, LB.[2008] A New Support Vector Machine with Fuzzy Hyper-Plane and Its Application to Evaluate Credit Risk. *2008 Eighth International Conference on Intelligent Systems Design and Applications*, 83-88.

[17] Russell Stuart, Norvig Peter. [1995] Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955.

[18] Rish Irina. [2001] An empirical study of the naive Bayes classifier (PDF). IJCAI Workshop on Empirical Methods in AI.

[19] Zhang, Harry. The Optimality of Naive Bayes (PDF). FLAIRS2004 conference.

[20] Rennie J, Shih L,Teevan J, Karger D. [2003] Tackling the poor assumptions of Naive Bayes classifiers (PDF). ICML.

ABOUT AUTHORS

Lohit Mittal *is currently studying B.tech (computer science and technology) at VIT University, Vellore.*
Tarang Gupta *is currently studying B.tech (computer science and technology) at VIT University, Vellore.*
Dr. Arun Kumar Sangaiah *is a Associate Professor at VIT university and PhD in computer science.*

ISSN: 0976-3104
REGULAR ISSUE
Gholami et al.

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

PRAGMATIC FAILURE AMONG IRANIAN EFL STUDENTS

Morteza Gholami[1], Narjes Nikookar[2]*, Naser Salehi[3] , Mehdi Abbasi[4]

[1]*Dept of foreign languages, Jahrom University of Medical Sciences, IRAN*

[2] *Dept of Speech and Language Pathology, Jahrom University of Medical Sciences, IRAN*

[3]*Dept of foreign languages, Darab Branch Islamic Azad University, Darab, IRAN*

[4]*Dept of foreign languages, Azad University of Shiraz, IRAN*

ABSTRACT

Objectives: *The development of L2 learners' pragmatic competence plays a key role in the achievement of communicative competence. The purpose of this study was to investigate the relationship between language proficiency and pragmatics.* ***Materials and methods:*** *The data were collected from 60 Iranian EFL students (33 females and 27 males) for a period of 2 weeks. The instruments used in this study were Oxford Placement Test and Pragmatic Knowledge Quiz. The data were analyzed qualitatively and the obtained results were analyzed by the software SPSS 16.* ***Results:*** *The results indicate that Iranian EFL students are poor in pragmatic issues and they have a lot of problems in this field. Regarding the relationship between language proficiency and pragmatics, there is a significant relationship between them but it is moderate. Sex difference regarding the use of pragmatic features of English reveals that there is no significant relationship between two sexes and two groups did almost the same job. And finally pragmatic features of English is predictable by a language proficiency test; namely as the students linguistic knowledge increases, their pragmatic awareness rises as well.*

Received on: 20th-June-2016
Revised on: 2 nd-July-2016
Accepted on: 16th– July-2016
Published on: 1st –Aug-2016

KEY WORDS

pragmatics; pragmatic failure; Sociopragmatics; linguistic competence; pragmatic competence

***Corresponding author: Email:** n.nikou1@yahoo.com; **Tel:** +98-9171908539

INTRODUCTION

Learning a language is too often viewed as simply a matter of mastering a distinct system of signs, without reference to the context in which a particular language is used. Recently, research in cross-cultural pragmatics has, however, clearly illustrated that different cultures use language in culturally distinctive ways.

People do not always or even usually say what they mean. Speakers frequently mean more than their words actually say. People can mean something quite different from what their words say, or even just the opposite. As a domain within L2 studies, pragmatics is one of the terrains deserving more attention. Pragmatic knowledge is an inseparable area of language proficiency as defined by Bachman [1]. Pragmatics is concerned with the study of meaning as communicated by speaker (or writer) and interpreted by a listener (or reader). It is the study of speaker's meaning [2].

Language learners may be competent in linguistic forms of the target language, but they may not be aware of the different functions and meanings of those forms in the target language. Pragmatic failure may not only cause ineffective communication, but may also cause native speakers to form misjudgments or misperceptions about the personality, beliefs and attitudes of the learner.

It has, consequently, more to do with the analysis of what the words or phrases in utterances might mean by themselves. This type of study deals with the interpretation of the people's utterance in a specific context and how context influences the speaker's utterance. Pragmatics also explores how listeners of the language interpret what is said in order to reach an interpretation of the speaker's intention. It also deals with what is unsaid by the speaker.
Kasper and Schmidt note that although pragmatics has played a considerable role in approaches to first and second language classroom research, classroom research has played only a minor role in interlanguage pragmatics thus far[3]. Pragmatics, as Kasper and Blum-Kulka put it, is the study of people's comprehension and production of linguistic action in context. Interlanguage pragmatics, thus, refers to the study of nonnative speakers' use and acquisition of linguistic action patterns in a second language [4]. It seems that Iranian EFL students are paralyzed

in pragmatic areas. Although they are proficient and competent in different aspects of language, they have a lot of problems facing the pragmatic issues.

The main objectives of this study is discovering pragmatic failure, and finding out the relationship between pragmatics and second language learning. It is striving to respond to the question that if those learners who are competent in language proficiency have this ability to deal with pragmatic issues, or being linguistically competent guaranties coping with pragmatic problems. So to respond these issues, the following research questions are posed:

- How is the performance of Iranian EFL learners in pragmatic situations?
- Is there any sex difference regarding the use of pragmatic features of English?
- What is the relationship between proficiency and knowledge of pragmatic features of English?
- Can knowledge of pragmatic features be predicted by a language proficiency test?.

INTRODUCTION TO PRAGMATICS

Learning language is not a linear process in which the individual learn the materials one after another but it is a spiral one in which the learners try to acquire the materials in several grammatical areas simultaneously. Learning to use language accurately (grammatical accuracy) as well as appropriately (pragmatic appropriacy) is a must. Pragmatic appropriacy and linguistic accuracy are two wings of language learning in which absence of one make the flying impossible. It presumably possible to produce pure correct language in terms of grammatical accuracy, but it makes no sense without considering pragmatic appropriacy and it makes communicative barriers in the process of language learning.

As Rintell and Mitchell point out, resorting to only literal meaning of the words can lead to misunderstanding or create offence when the person does not know the rules to interpret them[5]. According to Fraser, who insists a theory of linguistic communication, "any effects beyond the successful recognition of the speaker's intentions, such as convincing, annoying, or confusing the hearer, are not part of communication but the result of communication or perhaps the result of failure to communicate" [6]. What Fraser describes means "pragmatic failure to the inability to understand what is meant by what is said" presented by[7,6] . Pragmatic failure is a sub branch of cross-cultural pragmatics which has grown tremendously in last twenty years.

Pragmatics

The general area in language learning from the view point of its use is labeled as pragmatics. The main focus of pragmatics is on language users- the choice they make, kinds of words regarding formality or informality, the effect the language has on the other participants in the process of communication and the like.

Thomas believes that pragmatics is meaning in interaction [8]. He holds that pragmatics is concerned with the negotiation of meaning between the participants, the context in which the utterance is taking place, whether it is physical, social, linguistic or potential meaning of utterance.

LoCastero views language as an attempt to create meaning in a joint action by the speaker and the hearer that include both linguistic and non-linguistic signals in the sociocultural context[9]. Pragmatics is also the study of invisible meaning (2)– it regards how a lot of unproduced language is communicated between the speaker and the hearer. Speakers have a presupposition in their mind to assume that there is a great deal of information shared between the participants.

Pragmatic failure

It is worth to say that a great deal of misunderstanding is not because we do not hear the speaker or his words are not grammatical, but it is the inability to understand 'what is meant by what is said', Thomas exploited the term 'pragmatic failure'[7]. It is pragmatic failure which leads to cross-cultural communication failure, thus it seems essential to investigate the causes of pragmatic failure and seek for the ways to avoid them by choosing unwise linguistic forms for not to be offended or make a hindrance in the process of communication
Sharifian offers the following example from an Iranian student:

An Iranian student at Shiraz University receives the recommendation letter from her American lecturer that she has asked him to write for her and then turns to him and says, "I'm ashamed." Bewildered by the student's response, the lecturer asks, "What have you done?!!!" [10]

This is a case in intercultural miscommunication because of inappropriate use of the illocutionary force for appreciating. This expression is used in a situation in which an offence has taken place rather that expressing appreciation.

Pragmalinguistic and sociopragmatic failure

Leech, quoting Thomas, classifies general pragmatics into pragmalinguistics and sociopragmatics[11]. He describes “pragmalinguistics is related to grammar” while “sociopragmatics is related to sociology." In other words, “pragmalinguistic studies are language-specific” while “sociopragmatic studies are culture-specific”.
Pragmalinguistics concerns with the conveying communicative acts and interpersonal meaning. It includes some resources such as directness and indirectness, routines, and a large range of linguistic forms, which can intensify or soften communicative act (1[12]

Sociopragmatics, on the other hand, refers to the way in which language is used in a social situation, in other words, it concerns with the function of a language in the situation, and thus, learning of the language is interwoven with the learning of the culture in which language is taking place. The relationship that holds between both, language and culture is depicted, as shown in **Figure- 1** below:

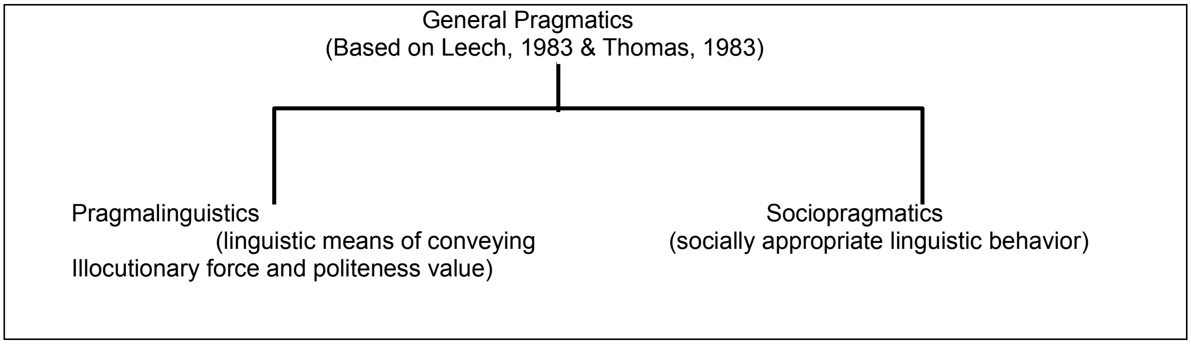

General Pragmatics
(Based on Leech, 1983 & Thomas, 1983)
Pragmalinguistics
(linguistic means of conveying
Illocutionary force and politeness value)
Sociopragmatics
(socially appropriate linguistic behavior)

Fig:1.the pragmatic continuum: language- culture

……………………………………………………………………………………………………..

As illustrated in **Figure- 1** above, Leech regards general pragmatics as “the study of linguistic communication in terms of conversational principles”, whereas pragmalinguistics and sociopragmatics belong to “more specific local conditions of language use” [11].

Speech act

While talking, people do not only produce meaningful utterances, they also perform actions via those utterances, which are known as speech acts – apologizing, promising, complaining, complimenting, inviting, etc. This can be done directly or indirectly.

The term speech acts was proposed for the first time by Austin [13]. He indicates that “the uttering of the sentence is, or is a part of, the doing of an action” [13], and he provides the concept of speech as an act. Speech acts, according to Searl, are “the basic or minimal units of linguistic communication” [14] Schmidt and Richards state that speech acts are “all the acts we perform through speaking and all the things we do when we speak” [15]. Speech acts are the actions we do through speaking, and they can cause a change as we speak or have an effect on the person we are talking to. For example, in saying I apologize, one is not only stating something, but is also performing an act of apology.

MATERIALS AND METHODS

Totally 60 EFL learners participated in this study. They constituted 30 fourth-year undergraduate, and 30 first-year graduate students from Shiraz Azad University, majoring in English teaching. They were studying English as a foreign language for four and five years, respectively. All learners were native speakers of Persian and their age ranged from 22 to 32. They were considered to be at a high level of language proficiency. None of them had visited any English-speaking country. The instruments utilized in this study consisted of two parts; a language a pragmatic awareness quiz and proficiency test. In order to examine participants' pragmatic knowledge, a pragmatic awareness quiz was employed. It constituted 40 items and all the items were in multiple choice format. It was provided by Professor Yarmohammadi and its content validity was approved by him. For the sake of reliability, it was calculated and 0.7 was obtained. The purpose of the pragmatic awareness quiz is to elicit participants' pragmatic awareness and the level of their pragmatic knowledge and to know how much they have acquired pragmatics during their academic studies.

To assess participants' language proficiency, Oxford Quick Placement Test (1999) was employed. It was a 60-item multiple-choice test consisting of three sections: Reading with fill in the blanks (25 items), Grammar (10 items) and Vocabulary (20 items) and language use (5 items). The reading Section assesses the students' ability to fill in the blanks with appropriate words. The grammar section assesses students' grammatical knowledge and asks them to choose the item which is grammatically correct. The purpose of the vocabulary section is to assess students' range of vocabulary understanding. Since this is a standard test and is used in language institutes to assess students' level of language proficiency, and its validity and reliability have already been established by Oxford University, we preferred to utilize it.

The data were collected over two weeks during the fall semester 2008- 2009. The data collection was carried out in paper and pencil. The nature of the test was explained to the participants of the study on the exam session. Moreover, the participants were assured on the confidentiality of the results and the advantage of their contributions to the study. On the exam session, half of the participants were given pragmatic quizzes and at the same time, the other half were given the language proficiency tests. After they completed their task, those who had pragmatic quizzes, were given language proficiency tests and those who had language proficiency tests were given pragmatic quizzes to have counter balance. They were given 40 minutes on the pragmatic quiz and 40 minutes on the language proficiency test. The exam sessions were observed and the participants' possible questions were answered.

The data were analyzed through the following statistical procedures through SPSS software (version 13) to answer the research questions. To find out the level of Iranian EFL students' pragmatic knowledge, descriptive analysis was used. Correlation test and One Way ANOVA were run to see the relationship between participants' language proficiency and pragmatic knowledge. The test which was employed to compare the differences between the males and the females concerning use of pragmatic features of English was Independent Samples T.test. To find out whether pragmatic features of English were predictable by language proficiency test, regression test was employed. All these constituted the quantitative part of the analysis.

RESULTS

Results of the analysis of the data along with their interpretation and discussion are presented in this chapter, which is divided into two main parts. The first part deals with the quantitative analysis of the data, the tables and their interpretations, and in the second part discussions of the data analyses are presented.

.

Participants' performance in pragmatic issues

The first research question refers to the performance of Iranian EFL learners in pragmatic issues in which descriptive analysis was used. The participants' scores on pragmatics are shown in **Table- 1**, including the mean, standard deviation, minimum, and maximum of the participants.

Table: 1 .Descriptive statistic results of the participants' pragmatic performance

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
pragmatic test 40	60	9	26	18.35	3.839
Valid N (listwise)	60				

Table- 1 indicates that the number of the participants is 60, the minimum achieved score is 9, the maximum score is 26, the mean obtained is 18.35 and the standard deviation is 3.839. Since pragmatic quiz consists of 60 items, the achieved mean of the participants is less than the total half.

Relationship between language proficiency and pragmatics

The second research question is concerned with the relationship between language proficiency and knowledge of pragmatic features of English. Table 2 reveals the relationship between language proficiency and pragmatics. It can be observed that the Pearson correlation between pragmatic quiz and language proficiency is 0.506.

Table: 2 .Correlation between linguistic knowledge and pragmatics

Correlations			
		Pragmatic test (40)	placement test (60)
Pragmatic test (40)	Pearson Correlation	1	.506**
	Sig. (2-tailed)		.000
	N	60	60
placement test (60)	Pearson Correlation	.506**	1
	Sig. (2-tailed)	.000	
	N	60	60
**. Correlation is significant at the 0.01 level (2-tailed).			

Table -3 shows the results achieved by On Way ANOVA test. It reveals that the level of significant between language proficiency and pragmatic quiz is 0.000 (between groups).

Table: 3. Statistic result of relationship between pragmatics and language proficiency

ANOVA					
pragmatic test 40					
	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	257.216	2	128.608	11.970	**.000**
Within Groups	612.434	57	10.744		
Total	869.650	59			

Table- 4 indicates the multiple comparisons of the participants in pragmatics. We divided them into three groups as High, Mid, and Low, the level of significance between High and Low group is 0.000. It is 0.432 between High and Mid, and finally the level of significance between Low and Mid is 0.002.

Table: 4 . Multiple comparisons of three different groups of pragmatics

	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	High	Mid	1.335	1.022	.432	-1.23	3.90
		Low	5.235*	1.124	.000	2.41	8.06
	Mid	High	-1.335	1.022	.432	-3.90	1.23
		Low	3.900*	1.022	.002	1.33	6.47
	Low	High	-5.235*	1.124	.000	-8.06	-2.41
		Mid	-3.900*	1.022	.002	-6.47	-1.33
*. The mean difference is significant at the 0.05 level.							

Sex difference regarding the use of pragmatic features of English

The third research question concerns the sex difference with regard to the use of pragmatic features of English in which the results are shown in the **Table- 5**.

Table- 5 pertains to the sex difference regarding the use of pragmatic features of English. It is observed that numbers of male and female participants are 27 and 33, respectively. Their mean scores achieved are 18 for the males and 18.64 for the females.

Table :5 .Sex difference regarding the pragmatic features of English

Group Statistics					
	Sex	N	Mean	Std. Deviation	Std. Error Mean
Pragmatic test 40	Male	27	18.00	4.010	.772
	Female	33	18.64	3.732	.650

Independent Sample test was used to see if the males and the females differ concerning the use of pragmatic features of English. As can be seen from the table 6, the levels of significance between the males and the females are 0.528 and 0.531.

Table :6 .Independent Sample Test for sex difference regarding use of pragmatics

Independent Samples Test								
		t-test for Equality of Means						
		T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Pragmatic test 40	Equal variance assumed	-.636	58	.528	-.636	1.001	-2.641	1.36
	Equal variances not assumed	-.631	53.911	.531	-.636	1.009	-2.659	1.38

Prediction of English pragmatic features by language proficiency test

Regression test was conducted to see if it is possible to predict pragmatic features with a language proficiency test. **Table- 7** reveals the descriptive analysis of the regression test. As it is demonstrated, the R is 0.506 and R square or correlation coefficient is 0.256.

Table: 7. Correlation between language proficiency and pragmatics

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.506a	.256	.243	3.340
a. Predictors: (Constant), placement test 60				

The main statistical data concerning the subjects' results in the language proficiency test and the pragmatic quiz are shown in **Table- 8**. It reveals the achieved results by ANOVA test. As it is observed, the significance is 0.000.

Table: 8 . Predictions of pragmatic features by linguistic knowledge test

ANOVA b						
Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	222.519	1	222.519	19.944	.000a
	Residual	647.131	58	11.157		
	Total	869.650	59			

Pragmatic and grammatical competences show a regular imbalance in the sense that grammatical competence exceeds pragmatic competence [16].

In order to answer question one in the research question, which is how the performance of Iranian EFL students is in pragmatic situations, the descriptive statistics was used. As can be seen from the **Table- 1**, the pragmatic knowledge test yielded lower scores than it was expected. The maximum obtained score was 26 and the minimum achieved score was 9 out of 40. And considering mean score which is 18.36, it is drastically low, and since the participants' mean score is less than the total half, it indicates that Iranian EFL students are poor in the pragmatic awareness.

The second hypothesis question suggested what relationship there is between language proficiency and knowledge of pragmatic features of English. Regarding the results illustrated in the **Table- 2**, we can see that the correlation between proficiency test and pragmatic knowledge test is 0.506. And **Table- 3** shows the level of significance which is 0.000. It means that the correlation between these two kinds of tests is moderate, but it is significant. From **Table- 2**, these conclusions can be drawn as follows:

(a) The linguistic ability of the participants is not tied inextricably with their pragmatic awareness.
(b) There is a relationship between language proficiency test and pragmatic knowledge quiz (sig: 0.000) but it is not so strong. We can say that the relationship between these two tests is moderate (correlation: 0.506).
The level of significance between Low Level and Mid level is 0.002. It means that as the subjects go from low level to high level in language proficiency their pragmatic knowledge increases and in the same line, as the subjects go from low level to high level their pragmatic knowledge increases as well (sig, 0.000). But the difference between mid level and high level is not significant. This indicates that there is no significant difference between these two groups namely as their levels of language proficiency increases, it has no effect on their pragmatic knowledge.

The third hypothesis question is striving to answer the question if there is a sex difference regarding the use of pragmatic features of English. To answer this question Independent Sample t-test was employed. **Table- 5** demonstrates that the mean score of the males is 18 and the mean score of the females is 18.6 out of 40. **Table- 6** reveals the level of significance which is 0.5. This means that there is no significant relationship between the two sexes regarding the use of pragmatic features of English. The Two groups do not differ significantly in term of their English pragmatic knowledge and both sexes did the same job in pragmatic awareness.

The fourth question is dealing with the question if pragmatic knowledge is predictable by a language proficiency test or not. The results reported in **Table- 7** and **-8** reveal that the R Square which is correlation coefficient between language proficiency test and pragmatic knowledge quiz is 0.25 and the level of significance is 0.000, so it is predictable. This indicates that there is a significant relationship between these two tests, but the correlation coefficient is very low. This shows that those students who are in higher level of language proficiency do a better job in pragmatic contexts. Namely, the higher the level of subjects' language proficiency is, the higher their pragmatic knowledge would be.

CONCLUSION

Despite the importance of teaching English in Iran, its functional goals have not been achieved yet. The assumption is that there are some problems in teaching English regarding the methods employed as well as textbooks which are currently used in the country. Consequently, such problems hinder the learners to acquire, or if acquire, develop communicative abilities. The present study was an attempt to prove the existence of such problems regarding the teaching methods in Iranian colleges.

Besides, this study explored to find out if pragmatics and grammatical knowledge develop simultaneously. This study suggests that linguistic competence is necessary for pragmatic competence, but not sufficient. In other words, linguistic competence is a necessary prerequisite to pragmatic competence, but it does not guarantee pragmatic competence.

There has been a large number body of studies regarding pragmatics in the field of applied linguistics during the last four or five decadesm [17-27].

In the last decade works of some scholars such as Kramsch indicate the close relationship between L2 teaching and pragmatics [28]. Some studies suggest that participants, who had higher language performance, completed the pragmatic tests much better comparing low language proficiency participants in ESL contexts [29-31]. On the other hand, some studies indicated that even learners who performed well in language proficiency, may exhibit a wide range of pragmatic competence when compared with native speakers in conversations and elicited conditions [32-34]. In addition Bardovi-Harlig & Dörnyei believe that a good level of language proficiency may not guarantee a good level of pragmatic competence for "The disparity between learner's and NS's pragmatic competence may be attributed to two key factors related to input and the salience of relevant linguistic features in the input from the point of view of the learner" [16].

In a study conducted by He and Yan on Chinese EFL learners, they found that the level of pragmatic competence was not proportional to their grammatical competence [35]. In another study carried out by Hong et al in which they compared freshmen and seniors in terms of their pragmatic competence, they understood senior who were linguistically competent, did not perform significantly better than the freshmen in pragmatic development [36]. Bardovi-Harlig et al suggest two reasons that a grammatical competence does not guarantee pragmatic competence: "The disparity between non-native and native learner's pragmatic competence may be attributed to two key factors related to input and the salience of relevant linguistic features in the input from the point of view of the learners " [16].

In this study, we found out that Iranian EFL students are poor in pragmatic situations. Although they have gained mastery in language proficiency, they encounter a great deal of difficulties when they are put in a pragmatic situation. We also investigated the relationship between language proficiency and pragmatics and found out that they are correlated with each other and that there is a significant relationship between language proficiency and pragmatic awareness, but this correlation is weak. We also found that there is no difference between male and female learners in pragmatic field, and eventually we came to this conclusion that pragmatic feature of English is predictable, namely, those students who are in a high level of language proficiency, do better in pragmatic situations.

This study shows that the learners of English have a good command of linguistic forms but it doesn't mean they can communicate with native speakers appropriately. An EFL learner may speak English frequently, and may not make any errors in vocabulary or grammar, but that doesn't necessarily mean that he can communicate with native speakers without pragmatic failures. Developing pragmatic competence is the most important task in cross-cultural communication.

CONFLICT OF INTEREST

The authors declare no competing interests

ACKNOWLEDGEMENT

We really thank Dr. Yarmohammadi, Dr. Yamini, and Dr Bagheri who provided insight and expertise that greatly assisted the research.We thank students of Azad University of Shiraz for assistance in completinf this paper.

FINANCIAL DISCLOSURE

None.

REFERENCES

[1] Bachman LF. [1990] Fundamental considerations in language testing: Oxford University Press;.

[2] Yule G.[1996] Pragmatics: Oxford University Press..

[3] Kasper G, Schmidt R.[1996] Developmental issues in interlanguage pragmatics. Studies in second language acquisition,18(02):149-169.

[4] Kasper G, Blum-Kulka S. [1989]Interlanguage pragmatics: An introduction. *Interlanguage pragmatics* 1993;3:15.

[5] Rintell E, Mitchell CJ. Studying requests and apologies: An inquiry into method. Cross-cultural pragmatics: Requests and apologies.:248-272.

[6] FRASER B.[1983] The domain of Pragmatics)). In J. RICHARDS & R. SCHMIDT (Eds): Language and Communication. New York: Longman.

[7] Thomas J.[1983]Cross-cultural pragmatic failure.

[8] Thomas JA. [2014]Meaning in interaction: An introduction to pragmatics: Routledge.

[9] LoCastro V. [2003]An introduction to pragmatics: Social action for language teachers: University of Michigan Press Ann Arbor, MI;.

[10] Sharifian F. [2004]Cultural schemas and intercultural communication: A study of Persian.

[11] Leech GN.[2016] Principles of pragmatics: Routledge.

[12] Kasper G, Rose KR. [2001]Pragmatics in language teaching: Ernst Klett Sprachen;.

[13] Austin JL. [1962]How to do things with words Cambridge. Mass: Harvard.

[14] Searle JR. [1969]Speech acts: An essay in the philosophy of language: Cambridge university press;.

[15] Schmidt RW, Richards JC. [1980]Speech acts and second language learning. Applied linguistics.1(2):129-57.

[16] Bardovi-Harlig K, Dörnyei Z. [1998]Do language learners recognize pragmatic violations? Pragmatic versus

grammatical awareness in instructed L2 learning. Tesol Quarterly 32(2):233-259.

[17] Henrichsen LE.[1998] Understanding culture and helping students understand culture. Web document.

[18] Harrison B. [1990]Culture and the language classroom: Macmillan Modern English.

[19] Halliday MAK. [1978]Language as social semiotic: London Arnold.

[20] AbiSamra N.[2001] Strategies and techniques for teaching culture. American University of Beirut Retrieved on. 11(10)

[21] Kramsch C. [1998]Language and culture: Oxford University Press.

[22] Moran PR, Lu Z.[2001] Teaching culture: Perspectives in practice: Heinle & Heinle Boston, MA

[23] Nemetz-Robinson G.[1985] Cross-cultural understanding. Processes and approaches for foreign language ESL and bilingual educators. London, Prentice Hall.

[24] Stern H.[1992] Issues and options in language teaching (edited posthumously by Patrick Allen & Birgit Harley). Oxford: Oxford University Press.

[25] Spencer-Oatey H. [2004]Culturally speaking: Managing rapport through talk across cultures: A&C Black;.

[26] Adaskou K, Britten D, Fahsi B.[1990] Design decisions on the cultural content of a secondary English course for Mororcco. *ELT journal* 44(1):3-10.

[27] Cortazzi M, Jin L. [1999]1 1 Cultural mirrors. Culture in second language teaching and learning.196.

[28] Kramsch C. [1988]The cultural discourse of foreign language textbooks. *Towards a new integration of language and culture,* 63-68.

[29] Hill T. [1997]The Development of Pragmatic Competence in an EFL Context. Temple University Japan: Ph. D. Dissertation, Tokyo.

[30] ROEVER C. [2005]Language Testing and Evaluation: Testing ESL Pragmatics: Development and Validation of a Web-Based Assessment Battery.

[31] Yamashita SO.[1996] Six measures of JSL pragmatics: Univ of Hawaii Pr;.

[32] Bardovi-Harlig K, Hartford BS. [1991] Saying "no" in English: Native and nonnative rejections. *Pragmatics and language learning* 2:41-57.

[33] Omar AS. [1992] Conversational Openings in Kiswahili: The Pragmatic Performance of Native and Non-Native Speakers. *Pragmatics and language learning 3*:20-32.

[34] Takahashi T, Beebe LM. [1987]The development of pragmatic competence by Japanese learners of English. *JALT journal* 8(2):131-155.

[35] He Z, Yan Z. [1986]The Pragmatic failure of Chinese students in communication in English: An investigation of Chinese-English pragmatic differences. *Foreign Languages Teaching and Research* 3:52-57.

[36] Hong Y-Y, Chiu C-Y, Kung TM.[1997] Bringing culture out in front: Effects of cultural meaning system activation on social cognition. *Progress in Asian social psychology*1:135-146.

ABOUT AUTHORS

Professor Morteza Gholami *is a Ph.d candidate and a lecturer at Jahrom University of Medical Sciences for 5 years. He has been Teaching ESP to the students of medicine, nursing , anesthesia, etc. Currently he is also teaching at Jahrom University.*

Narges Nikookar *has an MSc. in speech and language pathology. He is working in a private clinic.*

Naser Salehi *is a lecturer at Islamic Azad University of Darab Branch. He is currently teaching English to students at Iran Language Institute (ILI). His areas of interest are mainly vocabulary acquisition and reading comprehension.*

Mehdi Abbasi *is an instructor az Azad University of Shiraz*

THE IIOAB JOURNAL ISSN: 0976-3104

ARTICLE OPEN ACCESS

ESTIMATION OF SEMANTIC SIMILARITY BETWEEN CONCEPTS AND FUZZY RULES OPTIMIZATION WITH MODIFIED GENETIC ALGORITHM (MGA)

B. Shobhana[1*]and **R. Radhakrishnan**[2]

[1]*Anna University, Chennai, Tamil Nadu, INDIA*
[2]*Sasurie College of Engineering, Vijayamangalam, Tirupur, Tamil Nadu, INDIA*

ABSTRACT

Aims: *Semantic similarity estimation is an important component of analyzing natural language resources like clinical records. In the past, several approaches for assessing word similarity by exploiting different knowledge sources have been proposed. Some of these measures have been adapted to the biomedical field by incorporating domain information extracted from clinical data or from medical ontologies.* ***Materials and methods:*** *In the recent work semantic similarity results from medical ontologies is evaluated based on some rules and assumptions. However these rules generation is completely performed based on the concepts only not by consideration of semantic similarity measures on the concepts. Here a new similarity measure is proposed and it is combining both super concepts of the assessed concepts and their common specificity feature. The similarity measure is performed based on Information Content (IC) and context vector. Then Fuzzy Rule base Modified Genetic Algorithm (DLCS-FRMGA) approach is introduced to rule optimization phase.* ***Results:*** *Using the MGA, the problem of finding an optimal rule base can be reduced to improve their applicability and accuracy. Thus the deterioration problem never happens since the best solution from the current generation will be superior to or at least the same with the past.* ***Conclusion:*** *Using MeSH and SNOMED CT as the input ontology, the accuracy of DLCS-FRMGA proposed method is evaluated according to a standard benchmark datasets of manually ranked medical terms.*

Received on: 1st - April -2016
Revised on: 7t-h - May-2016
Accepted on: 22nd – May-2016
Published on: 5th– August-2016

KEY WORDS

Semantic similarity, Super concepts, Evaluated concepts, Least Common Subsumer (LCS), Information Content (IC), Modified Genetic Algorithm (MGA), Fuzzy Rulebase Modified Genetic Algorithm (DLCS-FRMGA)

***Corresponding author: Email:** sobhanab.scholar@gmail.com; **Tel:** +91-9443901212

INTRODUCTION

An information retrieval process begins when a user enters a query into the system. User queries are matched against the database information. However, as opposed to classical SQL queries of a database, in information retrieval the results returned may or may not match the query, so results are typically ranked. This ranking of results is a key difference of information retrieval searching compared to database searching [1]. Depending on the application the data objects may be, for example, text documents, images [2], audio [3], mind maps [4] or videos. Often the documents themselves are not kept or stored directly in the IR system, but are instead represented in the system by document surrogates or metadata [5].Semantic search seeks to improve search accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable data space, whether on the Web or within a closed system, to generate more relevant results [6].

Ontology is “a formal explicit specification of a shared conceptualization”. Ontology provides a common understanding of a term and also its relationship with other terms. Thus a hierarchy can be formed with the related terms. Ontology compartmentalizes the variables needed for some set of computations and establishes the relationships between them [7] [8]. The fields of artificial intelligence, the Semantic Web, systems engineering, software engineering, biomedical informatics, library science, enterprise bookmarking, and information architecture all create ontologies to limit complexity and to organize information.

Semantic Information Retrieval has become the core part of any search engine. Many papers deal with SWS that uses the OWL language for constructing ontology. DySE System (Dynamic Semantic Engine) [9] implements a context-driven approach in which the keywords are processed in the context of the information in which they are retrieved, in order to solve semantic ambiguity and to give a more accurate retrieval based on user interests. Ontology Construction in Education Domain [10] deals with the construction of Ontology for specific University constructing instances specifically. Here the usage of Protégé tool for constructing the ontology is illustrated. Query sentences as semantic networks [11] paper describes procedure for representing the queries in natural language as semantic networks. Here a syntactic analysis of the query is done by parsing the query using Stanford parser to tag each and every word with their corresponding parts of speech.

Semantic Information Retrieval System [12] is mainly concerned with retrieving information from a sports ontology using the SPARQL query language. Here specific information is retrieved from the ontology. The sports related information is queried from the ontology and it is done using SPARQL language. A Relation-Based Page Rank Algorithm for Semantic Web Search Engines [13] proves that relations among concepts embedded into semantic annotations can be effectively exploited to define a ranking strategy for Semantic Web search engines. This sort of ranking behaves at an inner level and can be used in conjunction with other established ranking strategies to further improve the accuracy of query results. The overview of the existing systems gives multitude approaches for semantic information extraction. Though these above systems perform a semantic analysis, it has been implemented in a more generic way. Hence in order to further enrich this process to retrieve more promising results a system has been proposed for queries relating to university domain (SIEU).

METHODS

In this work, firstly, we review and investigate different measures for semantic similarity computation. Then, propose a new measure considering the multiple inheritances in ontologies and the common specificity feature of the evaluated concepts in order to obtain a more accurate similarity between concepts. For measured semantic similarity score values, rules are formed based on fuzzy function. Created fuzzy rules then optimized by *Modified Genetic Algorithm (MGA)*. Finally, evaluate the proposed DLCS-FRMGA approach using two datasets of biomedical term pairs scored for similarity by human experts and exploiting SNOMED CT as the input ontology. Compare the correlation obtained by our measure with human scores against other measures. The experimental evaluations confirm the efficiency of the proposed measure.

Similarity Measurement

Here concentrated on the Path [14], *Leacock &Chodorow (LCH)* [15], and Wu & Palmer [16] path finding measures that are dependent on the shortest path separating concepts. Consider $p = \mathrm{path}\,(C_1, C_2)$, the quantity of nodes in the shortest path splitting two concepts, C_1 and C_2. The shortest path among two concepts navigates their Least Common Subsumer $(\mathrm{lcs}(C_1, C_2))$, i.e. their nearby common parent. The depth $(\mathrm{depth}(c))$ of a concept is described as the amount of nodes in the pathway to the root of the taxonomy; and d indicates the maximum depth of taxonomy. Path describes the similarity among two concepts basically as the converse of the extent of the path separating them.

One major complication with this description is that the similarity of a concept with itself is below 1 $(\mathrm{if}\, C_1 = C_2, \mathrm{then}\, \mathrm{path}(C_1, \mathrm{lcs}(C_1,C_2)) + \mathrm{path}(c2, \mathrm{lcs}(C_1,C_2)) = 2)$. As an alternative, here adopted the definition of Wu & Palmer employed in the Natural Language Toolkit :

One major complication with this description is that the similarity of a concept with itself is below 1 $(\mathrm{if}\, C_1 = C_2, \mathrm{then}\, \mathrm{path}(C_1, \mathrm{lcs}(C_1,C_2)) + \mathrm{path}(c2, \mathrm{lcs}(C_1,C_2)) = 2)$. As an alternative, here adopted the definition of Wu & Palmer employed in the Natural Language Toolkit :

$$\mathrm{sim}_{wp}(C_1, C_2) = \frac{2 \times \mathrm{depth}(\mathrm{lcs}(C_1, C_2))}{p - 1 + 2 \times \mathrm{depth}(\mathrm{lcs}(C_1,C_2))} \qquad (1)$$

Based on this, $\mathrm{if}\, C_1 = C_2, \mathrm{then}\, p - 1 = 0$, and the similarity measure evaluates to 1.

Measures based on Information Content

Resnik [17] formulated to balance the taxonomical arrangement of ontology with the information distribution of concepts assessed in input corpora. Here also utilized the notion of IC, by means of associating appearance probabilities to every concept in the taxonomy, calculated from their occurrences in a particular corpus. IC of a term a is calculated in accordance with the negative log of its probability of occurrence, $p(a)$ (5). In this way, uncommon words are considered more useful than common ones.

$$\mathrm{IC}(a) = -\log P(a) \qquad (2)$$

This procedure is typically done manually in order to guarantee the appropriateness of the tagging, applicability and hampering the scalability of this scheme with huge corpora. Furthermore, when either the taxonomy or the corpus transformations, re-computations are required to be recursively implemented for the affected concepts. As a result, it is essential to carry out a manual and time consuming investigation of corpora and resultant probabilities would depend on the size and temperament of input corpora. By taking the drawbacks of IC-based schemes because of their dependency on corpora, a few authors attempted to intrinsically derive IC values from ontology. These works depends on the supposition that the taxonomic arrangement of ontologies like WordNet is put in order in a significant manner, in accordance with the rule of cognitive saliency [18]. This confirms that humans specialized concepts when they require distinguishing them from previously existing ones. As a result, concepts with several hyponyms (i.e., specializations) are extremely common and offer fewer details than the concepts in the place of leaves of the hierarchy. Seco et al., [19] and Pirró and Seco [20] based IC computations on the amount of hyponyms. In view of the fact that, $\mathrm{hypo}(a)$ the amount of hyponyms of the concept a and $\mathrm{max\_nodes}$ the amount of hyponyms of the root node, they calculate IC of a concept in the following manner (10):

$$IC_{seco}(a) = 1 - \frac{\log(hupo(a)+1)}{\log(max\_nodes)} \quad (3)$$

The denominator guarantees that IC values are normalized in the limit [0...1]. This scheme only takes hyponyms of a particular concept in the taxonomy; as a result, concepts with the similar number of hyponyms however different degrees of generality come out to be equally comparable. With the aim of dealing with this complication effectively, and in the similar manner as for edge-counting measures, Zhou et al., [21] formulated to balance hyponym-based IC computation with the associated deepness of each concept in the taxonomy. The IC of a concept is found as given below:

$$IC_{zhou}(a) = k\left(1 - \frac{\log(hupo(a)+1)}{\log(max\_nodes)}\right) + (1-k)\left(\frac{\log(depth(a))}{\log(max\_depth)}\right) \quad (4)$$

Besides $hypo$and max_nodes, which has the similar meaning as eq. 20, $depth(a)$ corresponds to the deepness of the concept a in the taxonomy and max_depth indicates the maximum deepness of the taxonomy. The factor k fine-tunes the weight of the two features engaged in the IC assessment. Here used $k = 0.5$.

Context Vector Measures of Semantic Relatedness

Furthermore, the WordNet glosses can be considered as a corpus of contexts comprising of around 1.4 million words. Subsequently, the gloss vector measure got the maximum correlation concerning human judgment by means of different benchmarks.

Gloss vectors for all concepts in WordNet can be computed in this manner. The relatedness of two concepts is then determined as the cosine of the normalized gloss vectors corresponding to the two concepts:

$$related_{vector}(c_1, c_2) = \cos(angle(\overline{v_1}, \overline{v_2})) \quad (5)$$

Where c_1 and c_2 are the two given concepts, $\overline{v_1}$ and $\overline{v_2}$ are the gloss vectors corresponding to the concepts and $angle$ returns the angle between vectors. Using vector products, the above relatedness formula can be rewritten as:

$$related_{vector}(c_1, c_2) = \frac{\overline{v_1} \cdot \overline{v_2}}{|v_1||v_2|} \quad (6)$$

The measure of semantic relatedness based on WordNet and MeSH glosses, which is enhanced with information from a large corpus of text.

Proposed Similarity Measurements

Here, the most common similarity among two concepts (or two text nodes) C_1 and C_2 indicates a weighted sum of the similarities of the two features among them, i.e.:

$$Sim(C_1, C_2) = w_1 \bullet (Sim\ C(c_1, c_2) + w2 \bullet SimP(c_1, c_2)) \quad (7)$$

Data Content Similarity (SimC)

It is the Cosine similarity among the term frequency vectors of $C_1$and C_2:

$$simC(c_1, c_2) = \frac{V_{C_1} \cdot V_{C_2}}{|V_{C_1}| * |V_{C_2}|} \quad (8)$$

Where V_d indicates the frequency vector of the terms within the concept C, $||V_d||$ indicates the length of V_d, and the numerator is the inner product of two vectors.

Presentation Style Similarity (SimP)

It indicates the average of the style Feature Scores (FS) over the entire six presentation style Features (F) among $C_1$and C_2:

$$simP(c_1, c_2) = \sum_{i=1}^{6} {FS_i}/_6 \quad (9)$$

WhereFS_i is the score of the ith style feature and it is defined by $FS_i = 1\, F_i^{C_1} = F_i^{C_2}$ and $FS_i = 0$otherwise, and is the ithstyle feature of data unit d.

Common Specificity Feature

Here, a new modified measure for the purpose of semantic similarity by means of combining both the super concepts of the assessed concepts and common specificity feature which can confine further semantic indication. This measure can accomplish better performance than other measures, since it is completely based on structure and continue their simplicity. Consider c_i stand for ith concept of ontology. Subsequently, $N(c_i)$ is defined as the collection of the entire super concepts of c_i including c_i itself. As a result, the number of non-common super concepts of concepts can be determined as follows:

$$Noncomsub(c_1, c_2) = |N(c_1) \cup N(c_2)| - |N(c_1) \cap N(c_2)| \quad (10)$$

At this point, the $NonComSub$ value can be a sign of the path length of the two concepts. LCS node of two concepts C_1 and C_2 determines the common specificity of C_1 andC_2 in the cluster. So the specificity of two concepts is computed through the process of finding the deepness of their LCS node and subsequently scaling this deepness as follows:

$$ComSpec(c_1,c_2) = D - depth(LCS(c_1,c_2)) \quad (11)$$

Where$D$ indicates the deepness of the ontology and the $ComSpec$ feature decides the common specificity of two assessed concepts. The lesser the $ComSpec$ value of two concepts, the additional details they share, and as a result the more comparable they are. As a result, the semantic distance among concepts c_1 and c_2 is given as follows:

$$SemD(c1,c2) = \log(NonComSub \times ComSpec + 1) \quad (12)$$

It is to be mentioned that whichever concept can be evaluated with itself.

Footrule Similarity

In this proposed similarity is based on the footrule similarity. The expansion of the Spearman footrule distance by means of including the result of a similarity function described on concepts in U. This measure is described in concepts of a generic similarity function. In actual fact, the option of a specific similarity measure for domain is of huge significance, as well as the progress of similarity measures is an active area in the data mining and pattern recognition communities. At this point, take the similarity measure as specified, and presume that the similarity scores returned are significant inside the particular domain. The footrule similarity distance among two (maybe partial) ranked lists $\sigma$and r, provided a similarity function $s(.,.)$, is given as:

$$F_{sim}(\sigma,r,s(c_1,c_2)) = \sum_{C_1\in\sigma_{S|r}} \sum_{C_2\in r_{S|\sigma}} sim(C_1,C_2)|\sigma_{S|r}(C_1) - r_{S|\sigma}(C_2)| \quad (13)$$

Specifically, the footrule similarity distance is computed on similarity projections of $\sigma$and r. The difference in ranks for items in these resultant lists is weighted through the power of the similarity.
This measure can be applied to the entire concepts. Subsequently, the most common similarity between these semantic similarity and footrule similarity is a weighted sum of the similarities:

$$Sim(C_1,C_2) = w_1 * (SemD(c1,c2) + w2 * F_{sim}(\sigma,r,s(c_1,c_2))) \quad (14)$$

Fuzzy Rules for Semantic Similarity Score Evaluation

Let us introduce some precise definitions of what is meant by the rule base solution representation. First of all here given L linguistic variables that is L=3 ontologies, MeSH , SNOMED and MeSH and SNOMED {A^1, ..., A^L} and their semantic score value for concept 1 to concept n. Each linguistic variable A$_i$ has M$_i$ linguistic descriptions $\{A^i_1,\dots,A^i_{M_i}\}$ that arerepresented by triangular membership functions $\mu^i_j, j = 1,\dots,M_i$. A fuzzy rulehas the form [22]: ifA^{i_1} is $A^{i_1}_{j_1}$ and A^{i_2} is $A^{i_2}_{j_2}$ and A^{i_k} is $A^{i_k}_{j_k}$ then concept C Where$i_1,\dots i_k \in \{1,\dots L\}, j_k \in \{1,\dots M_{i_k}\}$and $\sigma \in [0,1]$. A rule base is a set of several rules. Let us assume that we are given a rulebase consisting of n rules: $ifA^{i_1^1}$ is $A^{i_1^1}_{j_1}$ and $A^{i_2^1}$ is $A^{i_2^1}_{j_2}$ and $A^{i_3^1}$ is $A^{i_3^1}_{j_3}$ then concept c_1
$ifA^{i_1^2}$ is $A^{i_1^2}_{j_1}$ and $A^{i_2^2}$ is $A^{i_2^2}_{j_2}$ and $A^{i_3^2}$ is $A^{i_3^2}_{j_3}$ then concept c_2
$ifA^{i_1^n}$ is $A^{i_1^n}_{j_1}$ and $A^{i_2^n}$ is $A^{i_2^n}_{j_2}$ and $A^{i_3^n}$ is $A^{i_3^n}_{j_3}$ then concept c_n
where $i_l^m \in \{1,\dots,L\}$and $j_l^m \in \{1,\dots,M_{i_l^m}\}$. Given a semantic vector from the multiple ontologies $sev \in \mathbb{R}^L$of observed semantic values, whose components are values for the linguistic variables A^1 , ..., A^L, can evaluate the rule base as follows [23]: the function ρ describes the way the rule base interprets semantic values observations $sev$to produce a single output value that is selected semantic vector belongs to concept one or not . However the above mentioned rules only satisfied to semantic measurement of the single attribute only so this can be changed as follows,

$ifA^{i_1^1}$ is $A^{i_1^1}_{j_1}$ and $B^{i_2^1}$ is $A^{i_2^1}_{j_2}$ and $B^{i_3^1}$ is $A^{i_3^1}_{j_3}$ then concept$c_1\&c_2$
$ifA^{i_1^2}$ is $A^{i_1^2}_{j_1}$ and $B^{i_2^2}$ is $B^{i_2^2}_{j_2}$ and $B^{i_3^2}$ is $B^{i_3^2}_{j_3}$ then concept $c_2\&c_3orc_2\&c_1$
$ifA^{i_1^n}$ is $A^{i_1^n}_{j_1}$ and $B^{i_2^n}$ is $B^{i_2^n}_{j_2}$ and $A^{i_3^n}$ is $A^{i_3^n}_{j_3}$ then concept $c_n\&c_{n-1}$ or $c_n * c_{n-1}$

This value has an application specific meaning and can be taken to be a real number. More precisely, $\rho: \mathbb{R}^L \rightarrow \mathbb{R}$ is defined as follows:

$$sev = \begin{pmatrix} sev^1 \\ sev^2 \\ \vdots \\ sev^L \end{pmatrix} \rightarrow \frac{\sum_{m=1}^{n} we(c^m) \prod_{l=1}^{k_m} \mu^{i_l^m}_{j_l^m}(x^{i_l^m})}{\sum_{m=1}^{n} c^m} \quad (15)$$

Evaluation Function

Consider an evaluation function (to minimize) that measures the minimum difference values between ontologies semantic values and as well as their concepts when training a rule base to fit a given biomedical dataset samples. This observed semantic values consists of a set $\{sev_i, y_i\}_{i=1,\dots,N}$, where each$sev_i = \begin{pmatrix} sev_i^1 \\ sev_i^2 \\ \vdots \\ sev_i^L \end{pmatrix}$

is a vector that has as many components as there are linguistic variables, i.e. $sev_i \in \mathbb{R}^L \forall i = 1,\ldots,N,$ and each y_i is a real number, i.e. $y_i \in \mathbb{R} \forall i = 1,\ldots,N$. Then the evaluation function has the form.

$$diff = \sum_{i=1}^{N} (\rho(sev_i) - y_i)^2 = \sum_{i=1}^{N} \left(\frac{\sum_{j=1}^{n} a_{ij} we(c^j)}{\sum_{j=1}^{n} we(c^j)} - y_i \right)^2 \quad (16)$$

Where $a_{zm} = \prod_{i=1}^{k_m} \mu_{j_i^m}^{i}\left(x_s^{i}\right)$ (17)

The major aim of this work is to optimize the rules base in such a way that the evaluation function $diff$ becomes minimal. This involves two separate problems. Firstly, the form of the membership functions μ_j^i may be varied to obtain a better result. Secondly, the rule base may be varied by choosing different rules or by varying the weights $e(c_i)$. In this paper we consider both problems as important however for rule optimization varying the weights $we(c_i)$ is determined via the use of the Modified Genetic Algorithm (MGA). In the MGA the difference between the two ontologies semantic meaning is less means then weight becomes very high or it becomes less, if the difference is less than those concepts are similar and those rules are considered as major important rule in fuzzy theory. Concentrate on the second problem, taking the form of the membership functions to be fixed. For example, we can standardize the number of membership functions for each linguistic variable A^i to be $M_i = 2n_i - 1$ and define

$$\mu_j^i = \begin{cases} 0 : sev \leq \dfrac{j-1}{2n_i} \\ 2n_i sev + 1 - j : sev \in \left[\dfrac{j-1}{2n_i}, \dfrac{j}{2n_i}\right] \\ -2n_i sev + 1 - j : sev \in \left[\dfrac{j}{2n_i}, \dfrac{j+1}{2n_i}\right] \\ 2n_i sev + 1 - j : sev \leq \dfrac{j-1}{2n_i} \end{cases} \quad (18)$$

for$j = 1,..2n_i - 1 = M_i$

Search Space

The search space is the set of all potential rule base solutions. Let us first of all compute the maximum number of rules n_{max} then each rule can be written in the form

If $A^1 is A_{j1}^1 and A^2 is A_{j2}^1$ …. And $A^L is A_{jL}^L$ then concept

where in this case $j_i \in \{0, 1, \ldots, M_i\}$ and$j_i = 0$ implies that the linguistic variable A_i does not appear in the rule. Then we have

$$n_{max} = (M_1 + 1) \times (M_2 + 1) \ldots \times (M_L + 1) - 1 \quad (23)$$

Note that we have subtracted 1 to exclude the empty rule. If include the possible choices of weights $we(c_i)$ with discretization $we(c_i) \in \{0, \frac{1}{d}, \ldots, 1\}$, then we have a system of$(d+1)^{n_{max}}$.

Modified Genetic Algorithm (MGA)

MGA possesses a structure similar to GA. However, the MGA [24] has been distinguished from the GA in that the reproduction of the optimized rules is processed after the completion of both the crossover and mutation. Thus the deterioration problem never happens since the best solution from the current generation will be superior to or at least the same with the past. However, in the recent work, rules typically consist of diverse sets of classes and properties. In addition, this rule base analysis does not execute the rules. So the measurement of semantic similarity between concepts requires much time to complete the task, to solve this problem grouping is introduced in the next approach.

In the beginning the MGA creates an initial population based on the semantic score values. In the next step the algorithm evaluate the objective values (difference value of the semantic score between two ontologies in the multiple ontology stage) of the current population. After that weights are reproduced. During the reproduction, crossover first occurs. New Weights from rules combine to form a whole new rule. The newly created weight values of the concepts then mutates. Mutation means that the elements of chromosome are a bit changed. These changes are mainly caused by errors in copying weights from fuzzy rule base. Then MGA ranked weights represented by their associated cost, to be "minimized", and returns the corresponding individual fitness. Next the most fitted weights from fuzzy rules are selected. Here the objective values of the fuzzy rules in the offspring are evaluated and re-insertion of fuzzy rules in rule phase replacing parents is done. The MGA is terminated when some criteria are satisfied, e.g. a certain number of generations, a mean deviation in the population, or when a particular point in the search space is encountered.

RESUTLS

There are no standard human rating datasets for semantic similarity in biomedical domain. Carried out an experiment similar to the one proposed. Two ontologies are used as background knowledge: WordNet and MeSH. WordNet is a lexical database that describes and structures more than 100,000 general concepts, which are semantically structured in an ontological way. The *Medical Subject Headings (MeSH)* contains a hierarchy of medical and biological terms defined by the U.S National Library of Medicine. Use the benchmarks of Hliaoutakis et al.[25] and Pedersen et al.[26] to evaluate our methods. For the first one, we have taken the ratings given by the 3 physicians, 9 medical coders and the average of both. In this manner, a suitable LCS between the two ontologies must be discovered to enable the similarity assessment. The performance of proposed Fuzzy Rulebase Modified Genetic Algorithm (DLCS-FRMGA) based similarity measure evaluation is compared to existing higher similarity method of Al Fengq in Yang et al and Rule Based Semantic Score Evaluation (RSSE).Precision measure is calculated based on the formula

$$\text{Precision} = \frac{T_p}{(T_p + F_p)} \qquad (24)$$

Recall is calculated based on the formula

$$\text{Recall} = \frac{T_p}{(T_p + F_n)} \qquad (25)$$

Accuracy is calculated based on the formula

$$\text{Accuracy} = \frac{T_p}{(T_p + F_p + F_n)} \qquad (26)$$

Where T_p – True Positive (Correct result), T_N – True Negative (Correct absence of result), F_p– False Positive (Unexpected Result) , F_n– False Negative (Missing result).
F-Measure is calculated based on the formula

$$F=2.\frac{\text{precision.recall}}{\text{prrecision+recall}} \qquad (27)$$

The simulation results for the evaluation of the proposed approach against various performance measures like Precision, Recall, Accuracy and F-Measure.

***Accuracy*:** The proposed similarity measure evaluation using Wordnet Fuzzy Rulebase Modified Genetic Algorithm (DLCS-FRMGA) produced better accuracy rate shown in **[Figure -1]** in which When the number of concepts increases the accuracy of the result is increases.

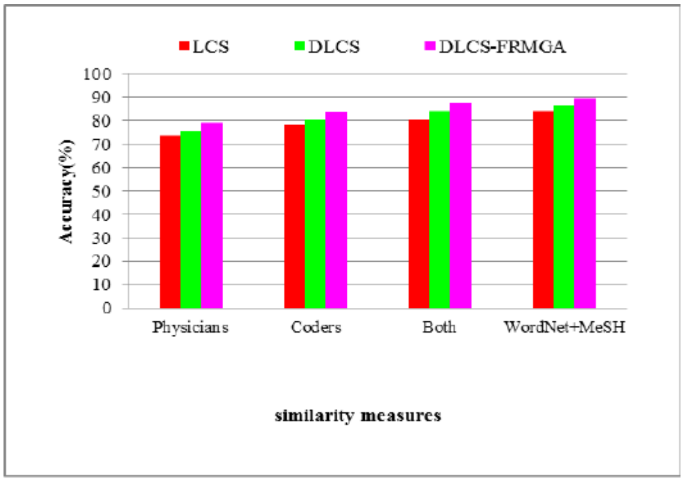


Fig: 1. Accuracy comparison of similarity measures

Figure 1 shows the accuracy comparison results where proposed DLCS-FRMGA produces 89.28% accuracy value to WordNet+MeSH which is better When compared to existing semantic similarity measurements with increased percentage of 2-6% for accuracy parameter.

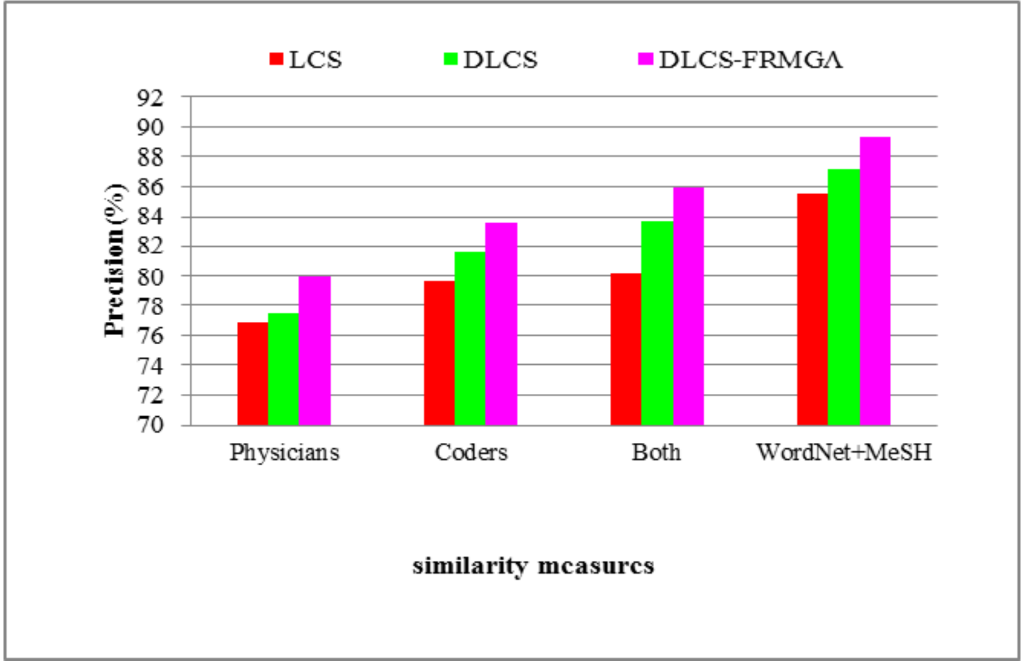


Fig: 2. precision comparison of similarity measures

[Figure- 2] shows the precision comparison results where proposed DLCS-FRMGA produces 89.28% precision value to WordNet+MeSH which performs better with increased percentage of 2-4%.

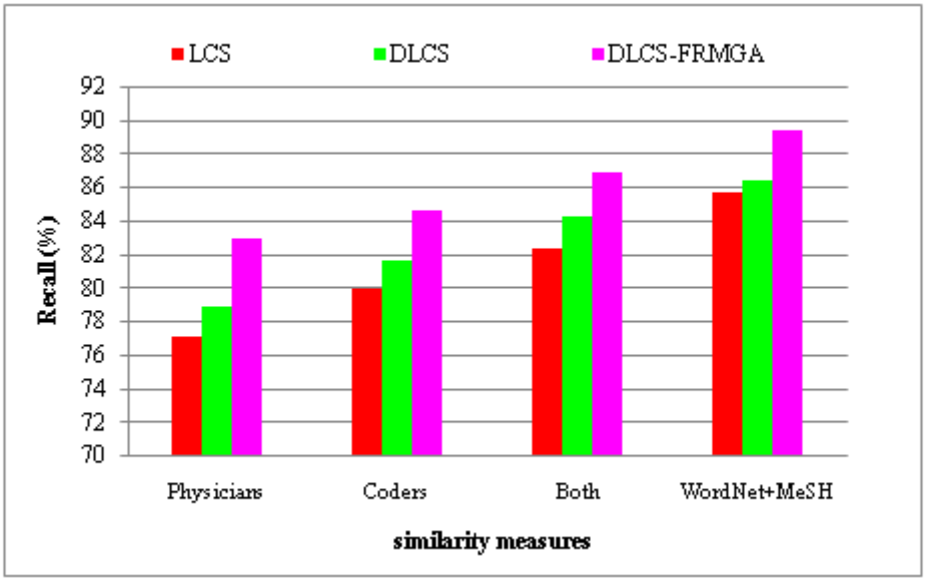


Fig: 3. Recall comparison of similarity measures

[Figure -3] shows the recall comparison results where the proposed DLCS-FRMGA produces 89.36% recall value which performs better with increased percentage of 2-4% for recall parameter.

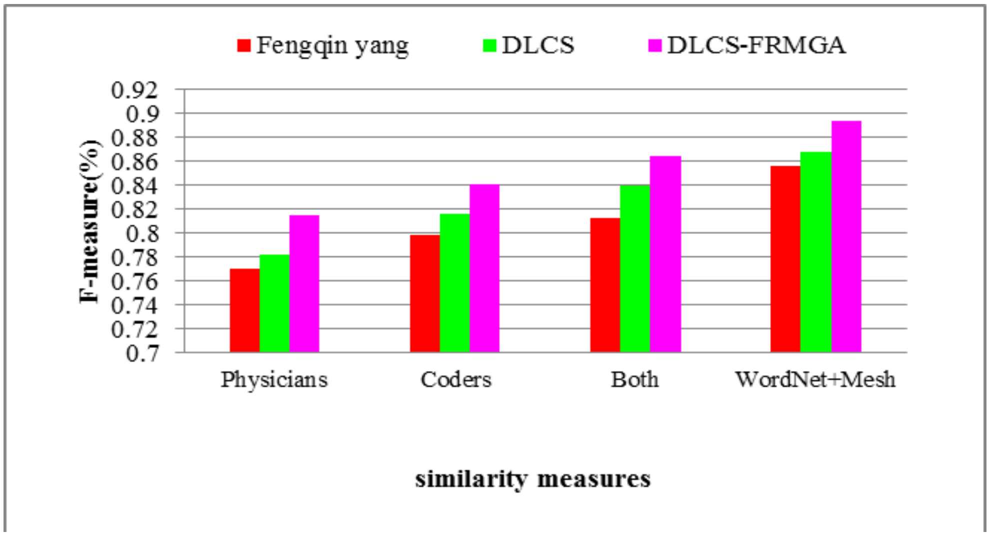


Fig: 4. F-measure comparison of similarity measures

The proposed similarity measure using Fuzzy Rule base Modified Genetic Algorithm (DLCS-FRMGA) produced high F1-measure shown in **[Figure -4]** which is higher than the existing similarity methods.

CONCLUSION AND FUTURE WORK

In the last few years, the amount of clinical data that is electronically available has increased rapidly. Digitized patient health records and the vast amount of medical and scientific documents in digital libraries have become valuable resources for clinical and translational research. With the intention of sorting them tranquil to understand and progress their applicability and precision, anticipate a Fuzzy Rulebase Modified Genetic Algorithm (DLCS-FRMGA) method deal with in fuzzy theory that permits the procedures deliberate to be consistently redefined. The common specificity feature considers the depth of the Least Common Subsumer (LCS) of two concepts and the depth of the ontology to obtain more semantic evidence. In addition the proposed DLCS-FRMGA method, some other features such as Data content and presentation features also extracted from biomedical text domain. For extracted features then data content similarity and Presentation style similarity is also measured. The weight of similarity between two concepts is calculated by using aggregation criteria. The similarity results of these are evaluated based on some rules and assumptions. For this purpose a DLCS-FRMGA approach is introduced to rule optimization phase. However rules in the rule generation phase is optimized using Modified Genetic Algorithm (MGA). Also anticipate calculating DLCS-FRMGA method in a accessible and effectual way from the taxonomical knowledge demonstrated in biomedical ontologies. As a outcome, new-fangled DLCS-FRMGA approach increases the semantic similarity measures expressed in terms of concept Information Content are presented. These measures are evaluated and compared to related works using a benchmark of medical terms and a standard biomedical ontology. The correlation amongst the outcomes of the estimated events and the human authorities' assessments demonstrates that DLCS-FRMGA method outstrips further most of the a fore mentioned events evading, roughly some of their confines.

CONFLICT OF INTEREST

The author declares no competing interest in relation to the work.

ACKNOWLEDGEMENT

Author is thankful to Dr. Radhakrishnan, Principal, Sasurie College of Engineering, Tamilnadu, India for his technical guidance in bringing out this paper efficiently. The author is also thankful to Dr.Makkal.G.Rajan, Chairman, Sri **Rajiv Gandhi polytechnic college**, Tamilnadu for providing moral support.

FINANCIAL DISCLOSURE

Nil

REFERENCES

[1] Jansen BJ, Soo YR, [2010]. The seventeen theoretical constructs of information searching and information retrieval. *Journal of the American Society for Information Science and Technology* 61(8): 1517-1534.

[2] Goodrum, Abby A, [2000]. Image information retrieval: An overview of current research. *Informing Science* 3(2): 63-66.

[3] Foote, Jonathan,[1999]. An overview of audio information retrieval. *Multimedia systems* 7(1): 2-10.

[4] Beel J, Gipp B, [2010]. Link analysis in mind maps: a new approach to determining document relatedness. *4th International Conference on Uniquitous Information Management and Communication*, pp.38.

[5] Baeza-Yates, Ricardo, William Bruce Frakes, eds. [1992]. Information Retrieval: Data Structures & Algorithms. *Prentice Hall*.

[6] Latzina, Markus, Anoshirwan Soltani, [2011]. Mixed initiative semantic search. *U.S. Patent* 7,987,176, No. 26.

[7] Gruber, Thomas R, [1993]. A translation approach to portable ontology specifications. *Knowledge acquisition* 5(2): 199-220.

[8] Fensel, Dieter, Ontologies, [2001]. In Ontologies, *Springer Berlin Heidelberg*. pp. 11-18.

[9] Antonio M, Rinaldi,[2009]. An Ontology-Driven Approach for Semantic Information Retrieval on the Web. *ACM Transactions on Internet Technologies* 9(3):10:1-10:24.

[10] Malik SK., Prakash N, Rizvi S.A.M. [2010]. Developing a University Ontology in Education Domain using Protégé for Semantic Web. *International Journal of Engineering Science and Technology* 2(9):4673-4681.

[11] Joel Booth, Barbara Di Eugenio, Isabel F, Cruz, Ouri Wolfson, [2009]. Query Sentences as Semantic (Sub) Networks, *IEEE International Conference on Semantic Computing*, pp.89-92.

[12] Jun Zhai, Kaitao Zhou, [2010]. Semantic Retrieval for Sports Information Based on Ontology and SPARQL, *International Conference of Information Science and Management Engineering (ICME)*. 19(2): 315-323.

[13] Fabrizio Lamberti, Andrea Sanna, Claudio Demartini, [2009]. A Relation-Based Page Rank Algorithm for Semantic Web Search Engines, *IEEE Transactions on Knowledge and Data Engineering* 21(1):123-136.

[14] Pedersen T, Pakhomov SVS, Patwardhan S, Chute CG. [2007]. Measures of semantic similarity and relatedness in the biomedical domain. *J Biomed Inform* 40:288–299.

[15] Leacock C, Chodorow M, [1998]. Combining local context with WordNet similarity for word sense identification, WordNet: A Lexical Reference System and its Application.

[16] Wu Z, Palmer M, [1994]. Verbs semantics and lexical selection, Association for Computational Linguistics, *Proceedings of the 32nd annual meeting on Association for Computational Linguistics, Las Cruces, New Mexico*, pp.133–138.

[17] Resnik P, [1995]. Using Information Content to Evalutate Semantic Similarity in a Taxonomy, In C.S. Mellish (Ed.), 14th International Joint Conference on Artificial Intelligence, *IJCAI Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc.*, 1: 448-453.

[18] Blank A, [2003]. Words and Concepts in Time: Towards Diachronic Cognitive Onomasiology, In R.Eckardt, K. von Heusinger& C. Schwarze (Eds.), Words and Concepts in Time: towards Diachronic Cognitive Onomasiology Berlin, Germany: Mouton de Gruyter, pp.37-66.

[19] Seco N, Veale T, Hayes J, [2004]. An Intrinsic Information Content Metric for Semantic Similarity in WordNet, In R. López de Mántaras& L. Saitta (Eds.), 16th Eureopean Conference on Artificial Intelligence, *ECAI 2004, including Prestigious Applicants of Intelligent Systems, PAIS Valencia, Spain: IOS Press*, pp.1089-1090.

[20] Pirró G, Seco N, [2008]. Design, Implementation and Evaluation of a New Semantic Similarity Metric Combining Features and Intrinsic Information Content, In R. Meersman& Z. Tari (Eds.), *OTM 2008 Confederated International Conferences CoopIS, DOA, GADA, IS, and ODBASE Monterrey, Mexico: Springer Berlin/Heidelberg*, 5332:1271-1288.

[21] Zhou Z, Wang Y, Gu J, [2008]. A New Model of Information Content for Semantic Similarity in WordNet, In S.S. Yau, C. Lee & Y.-C. Chung (Eds.), *Second International Conference on Future Generation Communication and Networking Symposia, FGCNS, Sanya, Hainan Island, China*, pp.85-89.

[22] JohnsonR, MelichM, MichalewiczZ, SchmidtM, [2005]. Coevolutionary optimization of fuzzy logic intelligence for strategic decision support. *IEEE Transactions on Evolutionary Computation* 9(6) :682–694.

[23] Esmin A, Lambert-TorresG,[2007]. Evolutionary computation based fuzzy membership functions optimization. *IEEE International Conference on Systems, Man and Cybernetics ISIC*, pp. 823–828.

[24] Roeva O,[2006]. A Modified Genetic Algorithm for a Parameter Identification of Fermentation Processes, Biotechnology & Biotechnological Equipment, 20, 1: 202-209.

[25] Hliaoutakis A, Varelas G, Voutsakis E, Petrakis E.G.M, Milios E.E, [2006] Information Retrieval by Semantic Similarity, International Journal on Semantic Web and Information Systems, 2: 55-73.

[26] Petrakis E.G.M, Varelas G, Hliaoutakis A, Raftopoulou P, [2006]. X-Similarity:Computing Semantic Similarity between Concepts from Different Ontologies. *Journal of Digital Information Management*, 4:233-237.