TEST Engineering & Management

March-April 2020
ISSN: 0193-4120 Page No. 5441 - 5447

A Hybridization Approach based Semantic Approach to the Software Engineering

Debabrata Samanta[1], Mohammad Gouse Galety[2], Shivamurthaiah M[3], Siddalingappa Kariyappala[4]
[1]Department of Computer Science, Christ (Deemed to be University), Bangalore, India
[2]Department of Networking, College of Engineering and Computer Science, Lebanese French University, Erbil, Iraq
[3]Department of Computer Science, Garden City University, Bangalore, India
[4]Department of Computer Science, Bapuji Institute of Hi-Tech Education, S S Layout, Davangere, India

Article Info
Volume 83
Page Number: 5441 - 5447
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 28 March 2020*

Abstract:
It is a major issue to measure semantic accuracy for determining the similarity between words, for developing a web and retrieving information thereby using natural language processing or mining of the text. There have been several proposals and adoption of methods for measuring the similarity by using WordNet and MeSH ontology. A new approach based on hybridization is proposed in this work. Feature based methodology and fuzzy set theory has been used here.

***Keywords**: Semantic Accuracy, Retrieving Information, NPL, MeSH.*

INTRODUCTION

The semantic web has a shared understanding, a defined structural and extension of current World Wide Web. Ontology play‟s an extremely important role for the advance of a semantic web. Ontology is defined by the Author Gurber et al "it is a proper requirement of a communal conceptualization" [1][2]. All-though ontologies be created for the different aspects and domains, they regularly contain overlapping data and the information. This information can be further used to evaluate the similarity among the words. Semantic similarity measures a numerical value that specifies the closeness among the words/concepts [3]. The idea of similarity or likeness is for identifying the concepts with common attributes. Appropriate assessment of similarity improves the understanding of textual data[4][5][6]. More semantically close words are those that share same idea or terms, specified by the synonym in the English language like "pain and hurt", two different words but can be used inter-changeably. In order to measure the accuracy of the similarity among the words/concepts is the open research problem in the area of computational linguistic and natural language processing. In the first description, Car, ship and bus which are different class, having „is-a‟ relation with Automobile, in travel ontology. Similarity can be computed between the {automobile, car},{automobile, bus},{bus, car} and so on.Semantic closeness shows how similar two Terms are, as they share some common attribute of the same meaning [7]. Two terms highly associated with the concept of semantic similarity is semantic distance and semantic relatedness [8]. Semantic distance is computed as edge or link difference between the two terms, how far the two terms; like dist(bus, car)=dist(car, ship) but sim(bus, car) ≠ sim(car, ship). Another term associated with semantic similarity is 'semantic relatedness', which does not only rely on the taxonomic relation "is-a" more relations to be considered like part –of, antonym relationship‟s that are present in the WordNet can be considered [9]. Words {ink, pencil} are less related to each other as compared to

March-April 2020
ISSN: 0193-4120 Page No. 5441 - 5447

{pencil, paper} in terms of semantic relatedness. A new approach has been proposed in this work, for feature-based approaches. In this WordNet is used for taxonomical information, Formal Concept Analysis (FCA) is exploit for extraction of features for given concept or words[10][11][12]. Linguistic value to each feature is assigned to show how much an attribute/feature is related to the corresponding concept/word/class accordance with fuzzy logic and mapped into fuzzy formal concept Analysis. A multi-attribute features table is constructed for the concepts [13]. A fuzzy set similarity measure is used to compute similarity among the words. The overall process yields us a better result in terms of accuracy on benchmark datasets. The same approach can be applied in Global ontologies, cross ontologies and fuzzy ontology.

.

RELATED WORK

As discussed earlier, implementing SPC and SPI has always been a hazard, as measuring is quite impossible in literature. A review of the literature was carried out in an organized manner, focusing on the evaluation of the software's efficiency on the implementation of SPC[14][15]. This study was conducted for identifying and analyzing: (i). positive factors influencing SPC implementation that is correlated to software measurement (ii). negative factors influencing SPC implementation that is correlated to software measurement. The process is comprised of three steps: (i). Development of the research protocol : The possible topic of interest and the context of the study is defined by the researcher along with the objective of the analysis. The next step characterizes the etiquette that would be used for performing the investigate [16][17]18]. The procedure should be comprised of information required to conduct the research: questions, search for the sources, criteria of selection, result storage procedure, analysis... etc. The feasibility of the procedure must be confirmed, i.e…, in case if satisfactory result is obtained, and if the procedure implementation is suitable in conditions of effort and time[19][20][21][22]. The test result should provide a scope of improvement of the procedure, if need arise. In case of a suitable protocol, it should have the approval of an expert, which then can be employed for the research.

ONTOLOGY-BASED SEMANTIC ROLE

The concepts of static software functionality description from ontological perspective have been covered by Roth et al., 2014 [23]. The ontological relations are as follows: Action: An procedure execute by an Actor or any Object is described as an action."HAS ACTOR" and "ACTS ON", respectively indicates the relation of a participant with the action who is a vigorous contributor of an Action and can be the user of a organization or a software structure itself. Any other entity other than the Actor himself is labeled as an "Object". The characteristic of an action and the attribute of an Object refers to Property [24] [25]. We have used the aforesaid material for a semantic role labeling approach for Training and Testing.

Table 1: reckon of annotated occurrence of relations and concepts

Relations	Instances	Concept	Instances
		ACTION	390
HAS_ACTOR	478	ACTOR	289
ACTS_ON	507	OBJECT	578
HAS_PROPERTY	604	PROPERTY	589
Total	1589	Total	1846

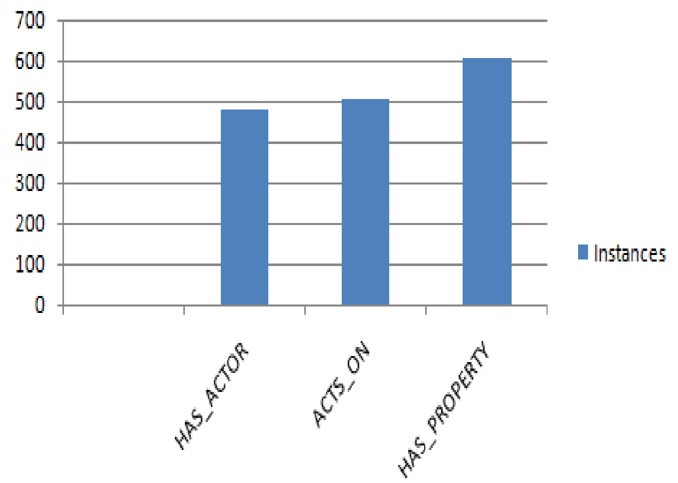

Figure 1: Tabulation form of Relations Vs

TEST Engineering & Management
March-April 2020
ISSN: 0193-4120 Page No. 5441 - 5447

Instances

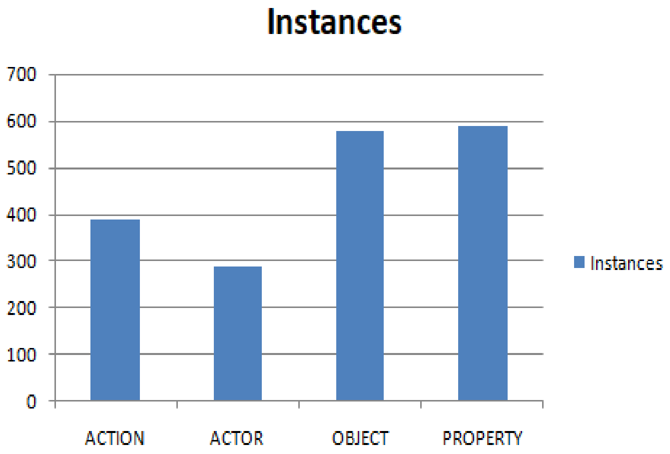

Figure 2: Tabulation form of Concept Vs Instances

As we have used all the annotations to mark for the tokens in text, the annotated data set can be used for role labeling approach in the following sections as well. The Actor and Action should be considered as instances and can differ from the number of associated concepts, provided in Table 1. The main purpose of the work is to express the relation between software requirements and extraction of Ontological concepts in natural language text. In section 3 we developed an interpretation, based on the Ontology and dataset, of statistical models for application in new data. A close resemblance has been observed in our task with reference to the kind of architecture, resembling previous semantic role labeling (SRL) tasks, so that the system may achieve performance (Hajiˇc et al., 2009). The starting point of our technique for labeling roles of the kind used as a standard in semantic analysis independent of the domain, extended to the definition of concepts and relation in the ontology (cf. Section 3), to fulfill the constraint and characteristics of the IT domain. A detailed description of the implementation is provided in the below subsections:

First one is an introduction of the preprocessing pipeline which is applied for expressing a requirement in the form of a sentence in English language, in order to compute a syntactic analysis. Second one is a description of the semantic analysis modules from the ontology implemented.

SYNTACTIC PSYCHOANALYSIS

Four ladders are adapted for extraction of the ontological concepts and relations for semantic role labeling from natural language texts: 1). Identification of the occurrence of the exploit and Object 2). Assign the respective concept type 3). Identifying the instances of the concepts that are similar 4). Relationship classification of the impression occurrence. The re-ranker that is built in is used for implementing to discover the united efficiency of step (3) and (4). Mate tool is then used for implementation of the semantic role labeler. It is further employed for continuous feature and collective label types. The following paragraph describes each component of our implementation:

Step (1) The related instances ACTS ON, HAS ACTOR and HAS PROPERTY influence in identifying the relation between the other concepts of ontology. Hence our first component is to identify these two-ontology concept of Action and Object. We therefore expect correspondence of linguistic units to behave in a similar manner to Prop-Bank/Nombank predictions and the related attributes applied to the step that identifies the predicate that is implemented in the Mate tools. Each and every verb and noun is considered in our implementation and based on lexical semantic and syntactic characteristics, binary is classified. Step (2) This step identified and determines the concept of ontology which is applied to every instance. Step (1) The component forecasts the sactual ontology concept, thereby classifying each noun and verb as an instance of Action and Object (e.g., “upload”!action, “search”!action). This step is a correspondence to the disambiguation of the predicate which is applied in the previous step Prop Bank/NomBank , but the label set is previously decided in the ontology and has no dependency on the “predicate” that is later identified.”. Step (3) The related instances of the components are determined by the components, to detect the words and phrases that are like the already identified

March-April 2020
ISSN: 0193-4120 Page No. 5441 - 5447

instances in Step (1). The main aim of this is to identify the actor of an Action and the objects that are affected and the Property that has relation with any of the above. This step is the same as argument identification in semantic role. Binary decision making is done taking consideration "arguments' as potential input, for determining if any word or phrase points out towards the same ontology concept. Example (a) - "The user" and "photos" both have relation with the Action that is expressed through the word "upload". Example (b) depicts relation of "search", "any user", "by tag" and "the public bookmarks of all RESTMARKS users". In this example, "of all RESTMARKS users", is a depiction of Property which is in relation with the Object manifested with the help of "the public bookmarks".

Table 2: statistical classification based on Linguistic properties

	Action and Object		Related concepts	
	Identification	classification	Identification	classification
AFFECTED WORD FORMS	Yes	Yes	Yes	Yes
AFFECTED WORD LEMMATA	Yes	------	------	------
PARENT PART-OF-SPEECH	Yes	Yes	------	Yes
SET OF DEPENDENT RELATIONS	------	Yes	------	------
SINGLE CHILD WORDS	Yes	------	Yes	------
SINGLE CHILD PAIL-OF-SPEECH	Yes	Yes	------	------
DISTANCE BETWEEN WORDS	------	------	Yes	------

Table 3: Performance of our full model and two simplified baselines

MODEL	PRECISION	RECALL	F1-SCORE
Baseline 1 (WORD-LEVEL PATTERNS)	57.8	45.7	43.86
Baseline 2 (SYNTAX-BASED PATTERNS)	61.7	57.8	62.34
Full SRL model	69.63	67.23	72.98

March-April 2020
ISSN: 0193-4120 Page No. 5441 - 5447

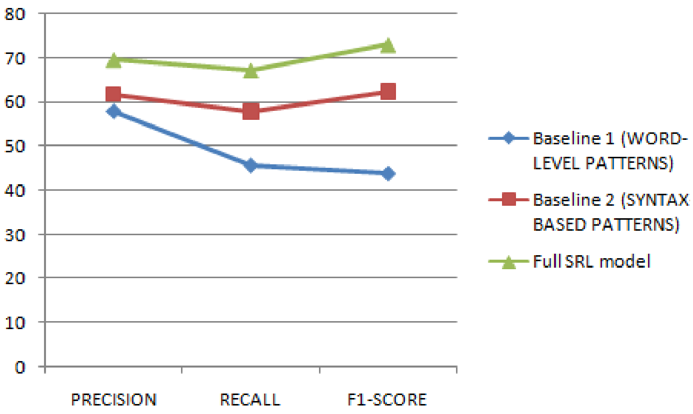

Figure 3: Graphical representation of Performance comparison

EVALUATION

The annotated dataset described in Section 3, is used for the evaluation of the efficiency of the approach of role labeling depicted in Section 4. Label precision refers to fraction of the idea and relation of the instances that are predicted and recalls the fraction of the correctly labelled annotations that are forecasted by the parser.

.

CONCLUSION

Relevant and organized portrayal of the software components is possible with the help of Ontology and semantic annotation of the requirements. For real time assistance, the mapping must be computed from requirements to ontology instances. The approach based on semantic role labelling, induced with ontology-based depiction, helps in achieving accuracy and portrays an outstanding performance when in comparison with two pattern-based baselines. The next work will comprise empirical validation of the proposed approach in downstream applications.

REFERENCES

1. Halevy, A. Y. (2005, October). Why your data won't mix: Semantic heterogeneity. Queue, 3(8), 50–58.
2. Heindl, M., & Bi , S. (2005). A case study on value-based requirements tracing. In Esec/fse-13:Proceedings of the 10th european software engineering conference held jointly with 13th acm sigsoft international symposium on foundations of software engineering (pp. 60–69). New York, NY, USA: ACM.
3. Hohpe, G., & Woolf, B. (2004). Enterprise integration patterns : designing, building, and deploying messaging solutions (I. Pearson Education, Ed.). Addison-Wesly.
4. Hu, G. (2006). Global schema as an inverse view of local schemas for integration. In Sera'06: Proceedings of the fourth international conference on software engineering research, management and applications (pp. 206–212). Washington, DC, USA: IEEE Computer Society.
5. Jagatheesan, A., Weinberg, J., Mathew, R., Ding, A., Vandekieft, E., Moore, D., et al. (2005).
6. Datagridflows: Managing long-run processes on datagrids. In J.-M. Pierson (Ed.), Data management in grids: first vldb workshop (pp. 113–128). Springer-Verlag.
7. Jamadhvaja, M., & Senivongse, T. (2005). An integration of data sources with uml class models based on ontological analysis. In Ihis '05: Proceedings of the first international workshop on interoperability of heterogeneous information systems (pp. 1–8). New York, NY, USA: ACM.
8. Jarke, M. (1998). Requirements tracing. Commun. ACM, 41(12), 32–36.
9. Jieming Wu, X. T. (2010, March). Research of enterprise application integration based-on esb. In Advanced computer control (icacc), 2010 2nd international conference on advanced computer control (pp. 90–93).
10. Weng, L., Agrawal, G., Catalyurek, U., Kurc, T., Narayanan, S., & Saltz, J. (2004). An approach for automatic data virtualization. In Hpdc '04: Proceedings of the 13th ieee international symposium on high performance distributed computing (pp. 24–33). Washington, DC, USA: IEEE Computer Society.

March-April 2020
ISSN: 0193-4120 Page No. 5441 - 5447

11. Xiong, F., Han, X., & Liqun, K. (2009, August). Research and implementation of heterogeneous data integration based on xml. In Electronic measurement and instruments, 2009. Icemi ’09. 9th international conference on (pp. 711–715).
12. Yan Du, L. Z., Wuliang Peng. (2008, December). Enterprise application integration: an overview. In Intelligent information technology application workshops, 2008. iitaw ’08. international symposium on intelligent information technology application workshops (pp. 953–957).
13. S K Podder, Debabrata Samanta, Gurunath R, Impact of Business Analytics for Smart Education System and Management Functions, Proc. of International Conference on Advancements in Computing & Management (ICACM-2019), April 13-14, 2019 by JaganNath University Rajasthan.@ Algorithms for Intelligent Systems (AIS)
14. Pradhan K N, Siddappa M, Kavitha S, Debabrata Samanta , Analysis & Improvement of Wireless Network Security based on Biometrics , Proc. of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM-2019) , February 26-28, 2019 by Amity University Rajasthan.
15. S K Podder, Debabrata Samanta , Impact of Climate Changes on Rural Livelihoods and Re-Orienting the Situation through Human Resources towards a Sustainable Society , Proc. of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM-2019) , February 26-28, 2019 by Amity University Rajasthan.
16. Rajesh Rao K, Debabrata Samanta , M. T. Somashekara, Gurunath R,A Paul, Automatic Segmentation based Graphical Interface for Recognition of Tomography Image Analysis of Damage cell , Proc. of The 3rd International Conference on Inventive Systems and Control (ICISC 2019) , 10-11 January 2019 by JCT College of Engineering and Technology, Coimbatore.
17. Shivamurthaiah Mallaiah, Sitesh Kumar Sinha, Praveen Kumar K, Manasa L R “Session password Authentication using Image Based System”, December, 2017,India, Volume 8, Issue 9, 2017, pp.34-36.
18. Shivamurthaiah M, Manjula S, “A Quadra Techniques for Face Recognition through OZMN”, An International Conference on Recent Trends in IT Innovations-Tec’afe 2017, Volume 5, Issue 2, April 2017.pp154-158.
19. Praveen Kumar K Sitesh Kumar Sinha Shivamurthaiah M, “Online Voting System using Aadhaar Number with OTP”, IJRECE, Volume 7, Issue 1, 2019,India
20. Shivamurthaiah M, Manjula S, Praveen Kumar K,“A Restoration Strategy to Extract a Definite Image from Noisy Image in Medical Image Processing”, An International Conference on Recent Trends in IT Innovations-Tec’afe 2017, Volume 5, Issue 2, April 2017.pp159-165.
21. Zhang Yi, Z. J. (2009, December). Research and implementation of eai based on soa. In Computational intelligence and software engineering, 2009. cise 2009. international conference on computing.
22. Dhanush V, Mahendra A R, Kumudavalli MV,Samanta D , Application of deep learning technique for automatic data exchange with Air-Gapped Systems and its Security Concerns, Proc. of IEEE International Conference on Computing Methodologies and Communication , 18-19, July 2017 , Erode.
23. R Kumar, Rishabh K,Samanta D , M Paul, CM Vijaya Kumar, A Combining approach using DFT and FIR filter to enhance Impulse

response, Proc. of IEEE International Conference on Computing Methodologies and Communication , 18-19, July 2017 , Erode.

24. G Ghosh, Samanta D, M Paul, N Kumar Janghel, Hiding Based Message Communication techniques depends on Divide and Conquer Approach, Proc. of IEEE International Conference on Computing Methodologies and Communication, 18-19, July 2017, Erode.
25. R. K. Singh, T. Begum, L. Borah, Samanta D, Text Encryption: Character Jumbling, Proc. of IEEE International Conference on Inventive Systems and Control @IEEE, 19-20 January 2017, Coimbatore.
26. G. Md. Gouse, Dr.B. Kavitha, Web Mining: A Review of Present Research Techniques, and Its Software,Journal of Computational Information Systems Volume 5,Issue 1,Pages 75-83.