International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 1

ISSN: 2454-4248
419 – 423

A Review on Computing Semantic Similarity of Concepts in Knowledge Graphs

Harshal Wanjari, Prof. Nutan Dhande

Department of CSE,

ACE NagthanaWardha MH India

Abstract: Semantic similarity is a metric defined over a set of documents or terms, where the idea of distance between them is based on the likeness of their meaning or semantic content as opposed to similarity which can be estimated regarding their syntactical representation (e.g. their string format). One of the drawbacks of conventional knowledge-based approaches (e.g. path or lch) in addressing such task is that the semantic similarity of any two concepts with the same path length is the same (uniform distance problem).To propose a weighted path length (wpath) method to combine both path length and IC in measuring the semantic similarity between concepts. The IC of two concepts' LCS is used to weight their shortest path length so that those concept pairs having same path length can have different semantic similarity score if they have different LCS.

*****

I. INTRODUCTION

Semantic similarity is a metric defined over a set of documents or terms, where the idea of distance between them is based on the likeness of their meaning or semantic content as opposed to similarity which can be estimated regarding their syntactical representation (e.g. their string format). These are mathematical tools used to estimate the strength of the semantic relationship between units of language, concepts or instances, through a numerical description obtained according to the comparison of information supporting their meaning or describing their nature.The term semantic similarity is often confused with semantic relatedness. Semantic relatedness includes any relation between two terms, while semantic similarity only includes "is a" relations. For example, "car" is similar to "bus",but is also related to "road" and "driving".

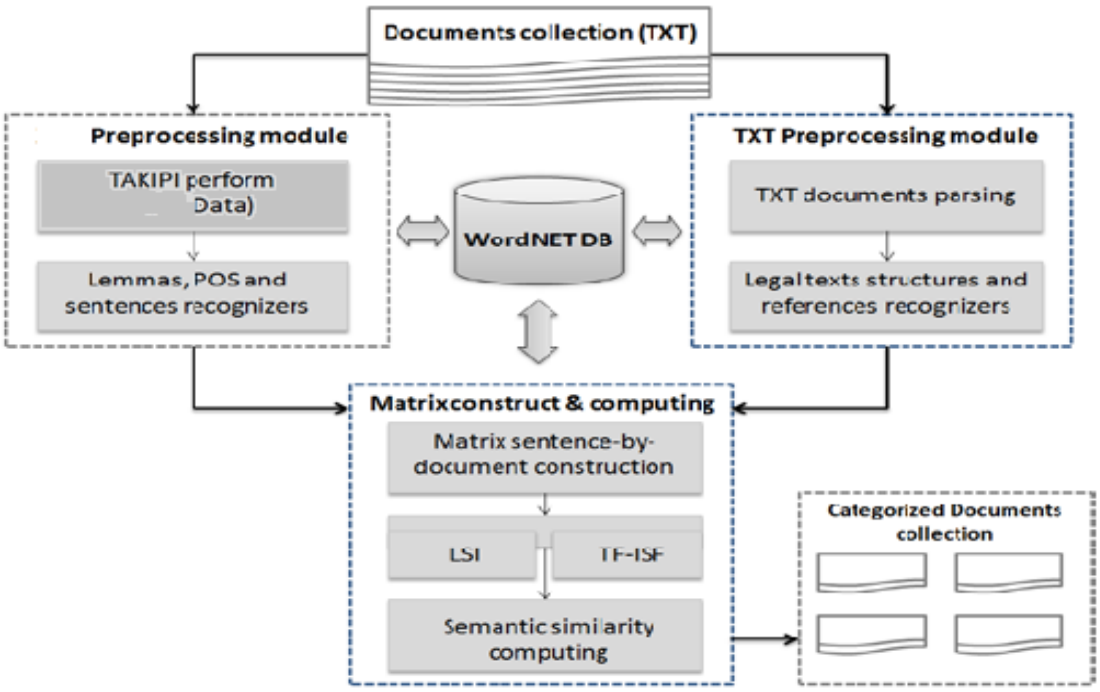

Figure.1.1 Architecture of Computing Semantics Similarity

 ISSN: 2454-4248
419 – 423

To proposed a method for measuring the semantic similarity between concepts in Knowledge Graphs (KGs) such as WordNet and DBpedia. Previous work on semantic similarity methods have focused on either the structure of the semantic network between concepts (e.g. path length and depth), or only on the Information Content (IC) of concepts. We propose a semantic similarity method, namely wpath, to combine these two approaches, using IC to weight the shortest path length between concepts. Conventional corpus-based IC is computed from the distributions of concepts over textual corpus, which is required to prepare a domain corpus containing annotated concepts and has high computational cost. As instances are already extracted from textual corpus and annotated by concepts in KGs, graph-based IC is proposed to compute IC based on the distributions of concepts over instances. Through experiments performed on well known word similarity datasets, we show that the wpath semantic similarity method has produced statistically significant improvement over other semantic similarity methods. Moreover, in a real category classification evaluation, the wpath method has shown the best performance in terms of accuracy and F score.

II. Literature Survey

All researches have aimed to develop and provide the generalized solution to monitor systematic way representing semantics in words using knowledge graph which can improve the efficiency of the database record and reduce the space between the data retrieval system. The major contributions to these topics are summarized below.

1. Information Retrieval by Semantic Similarity [1]

Author: AngelosHliaoutakis,GiannisVarelasDepartment of Electronics and Computer Engineering, Greece

Publication:Volume 4, Issue 4, April 2016 International Journal of Advance Research.

Semantic Similarity relates to computing the similarity between conceptually similar but not necessarily lexically similar terms. Typically, semantic similarity is computed by mapping terms to an ontology and by examining their relationships in that ontology. We investigate approaches to computing the semantic similarity between natural language terms (using WordNet as the underlying reference ontology) and between medical terms (using the MeSH ontology of medical and biomedical terms). The most popular semantic similaritymethods are implemented and evaluated using WordNet and MeSH. Building upon semantic similarity we propose the Semantic Similarity based Retrieval Model (SSRM), a novel information retrieval method capable for discovering similarities between documents containing conceptually similar terms. The most effective semantic similarity method is implemented into SSRM.SSRMhas been applied in retrieval on OHSUMED(a standard TREC collection available on the Web). The experimental results demonstrated promising performance improvements over classic information retrieval methods utilizing plain lexical matching (e.g., Vector Space Model) and also over state-of-the-art semantic similarity retrieval methods utilizing ontologies.mputer Science and Management Studies.

2. Design and Evaluation of Semantic Similarity Measures for Concepts Stemming from the Same or Different Ontologies [2]

Author:Euripides G.M. Petrakis.

Semantic Similarity relates to computing the similarity between concepts (terms) which are not necessarily lexically similar. We investigate approaches to computing semantic similarity by mapping terms to an ontology and by examining their relationships in that ontology. More specifically, to investigate approaches to computing the semantic similarity between natural language terms (using WordNet as the underlying reference ontology) and between medical terms (using the MeSH ontology of medical and biomedical terms). The most popular semantic similarity methods are implemented and evaluated using WordNet and MeSH. The focus of this work is also on cross ontology methods which are capable of computing the semantic similarity between terms stemming from different ontologies (WordNet and MeSH in this work). This is a far more difficult problem (than the single ontology one referred to above) which has not been investigated adequately in the literature. X-Similarity, a novel cross-ontology similarity method is also a contribution of this work. All methods examined in this work are integrated into a semantic similarity system which is accessible on the Web.

3.SURVEY OF SEMANTIC SIMILARITY MEASURES IN PERVASIVE COMPUTING [3]

Author: DjamelGuessoum,MoeizMiraoui,

Publication:INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS VOL. 8, NO. 1, MARCH 2015

Semantic similarity measures usage is prevalent in pervasive computing with the following aims: 1) to compare the components of an application; 2) to recommend and rank services by degree of relevance; 3) to identify services by matching the description of a query with the available services; 5) to compare the current context with already known contexts. The existing works that apply semantic similarity measures to pervasive computing focus on one particular issue. Furthermore, surveys in this domain are limited to the recommendation or discovery of context-aware services. In this article, we therefore present a survey

 ISSN: 2454-4248
419 – 423

of context-aware semantic similarity measures used in various areas of pervasive computing.

4.A Survey on Semantic Similarity Measure [4]

Author: S. Anitha Elavarasi1,Dr. J. AkilandeswariDepartment of Computer Science and Engineering Department of Information Technology Sona College of Technology

Publication: International Journal of Research in Advent Technology, Vol.2, No.3, March 2014.

Measuring semantic similarity between concepts is an important problem in web mining and text mining which needs semantic content matching. Semantic similarity has attracted great concern for a long time in artificial intelligence, psychology and cognitive science. Many measures have been proposed. The paper contains a review of the state of art measures including path based measures information based measures, feature based measures and hybrid measures. The features, performance advantages, disadvantages and relatedissues of different measures are discussed. This paper makes a review of semantic similarity measures with various approaches.

5.Development and application of a metric on semantic nets [5]

Author: R. RadaDept. of Comput. Sci., Liverpool Univ., UK.E. Bicknell

Publication: IEEE Transactions on Systems, Man, and Cybernetics (Volume: 19, Issue: 1, Jan/Feb 1989)

Motivated by the properties of spreading activation and conceptual distance, the authors propose a metric, called distance, on the power set of nodes in a semantic net. Distance is the average minimum path length over all pairwise combinations of nodes between two subsets of nodes.Distance can be successfully used to assess the conceptual distance between sets of concepts when used on a semantic net of hierarchical relations. When other kinds of relationships, like 'cause', are used, distance must be amended but then can again be effective. The judgements of distance significantly correlate with the distance judgements that people make and help to determine whether one semantic net is better or worse than another. The authors focus on the mathematical characteristics of distance that presents novel cases and interpretations. Experiments in which distance is applied to pairs of concepts and to sets of concepts in a hierarchical knowledge base show the power of hierarchical relations in representing information about the conceptual distance between concepts.

6. An approach for measuring semantic similarity between words using multiple information sources.[6]

Author:Y. LiManchester Sch. of Eng., Manchester Univ., UK.D. Mclean

Publication:IEEE Transactions on Knowledge and Data Engineering (Volume: 15, Issue: 4, July-Aug. 2003)

Semantic similarity between words is becoming a generic problem for many applications of computational linguistics and artificial intelligence. This paper explores the determination of semantic similarity by a number of information sources, which consist of structural semantic information from a lexical taxonomy and information content from a corpus. To investigate how information sources could be used effectively, a variety of strategies for using various possible information sources are implemented. A new measure is then proposed which combines information sources nonlinearly. Experimental evaluation against a benchmark set of human similarity ratings demonstrates that the proposed measure significantly outperforms traditional similarity measures**.**

7.Measuring Semantic Similarity Based on Weighting Attributes of Edge Counting.[7]

Author: JuHum Kwon, Chang-Joo Moon,Soo-Hyun Park, Doo-Kwon Baik

Publication: International Conference on AI, Simulation, and Planning in High Autonomy Systems.Semantic similarity measurement can be applied in many different fields and has variety of ways to measure it. As a foundation paper for semantic similarity, we explored the edge counting method for measuring semantic similarity by considering the weighting attributes from where they affect an edge's strength. We considered the attributes of scaling depth effect and semantic relation type extensively. Further, we showed how the existing edge counting method could be improved by considering virtual connection. Finally, we compared the performance of the proposed method with a benchmark set of human judgment of similarity. The results of proposed measure were encouraging compared with other combined approaches.

8. Verbs semantics and lexical selection[8]

Author: Zhibiao Wu National University of Singapore, Republic of Singapore.Martha Palmer University of Pennsylvania, Philadelphia, PA.

Publication:ACL '94 Proceedings of the 32nd annual meeting on Association for Computational Linguistics

This paper will focus on the semantic representation of verbs in computer systems and its impact on lexical selection problems in machine translation (MT). Two groups of English and Chinese verbs are examined to show that lexical selection must be based on interpretation of the sentences as well as selection restrictions placed on the verb arguments. A novel representation scheme is suggested, and is compared to representations with selection restrictions used in transfer-based MT. We see our approach as closely aligned with knowledge-based MT approaches (KBMT),

 ISSN: 2454-4248
419 – 423

and as a separate component that could be incorporated into existing systems. Examples and experimental results will show that, using this scheme, inexact matches can achieve correct lexical selection.

9.Word association norms, mutual information, and lexicography[9]

Author: Kenneth Ward Church Bell Laboratories, Murray Hill, N.J.Patrick Hanks Collins Publishers, Glasgow, Scotland

Publication: Journal Computational Linguistics archive Volume 16 Issue 1, March 1990.

The term word association is used in a very particular sense in the psycholinguistic literature. (Generally speaking, subjects respond quicker than normal to the word nurse if it follows a highly associated word such as doctor.) We will extend the term to provide the basis for a statistical description of a variety of interesting linguistic phenomena, ranging from semantic relations of the doctor/nurse type (content word/content word) to lexico-syntactic co-occurrence constraints between verbs and prepositions (content word/function word). This paper will propose an objective measure based on the information theoretic notion of mutual information, for estimating word association norms from computer readable corpora. (The standard method of obtaining word association norms, testing a few thousand subjects on a few hundred words, is both costly and unreliable.) The proposed measure, the association ratio, estimates word association norms directly from computer readable corpora, making it possible to estimate norms for tens of thousands of words.

10.Corpus-based and knowledge-based measures of text semantic similarity [10]

Author: RadaMihalceaDepartment of Computer Science, University of North Texas,Courtney Corley Department of Computer Science, University of North Texas.CarloStrapparavaIstituto per la RicercaScientifica e Tecnologica, ITC.

Publication:AAAI'06 Proceedings of the 21st national conference on Artificial intelligence - Volume 1

This paper presents a method for measuring the semantic similarity of texts, using corpus-based and knowledge-based measures of similarity. Previous work on this problem has focused mainly on either large documents (e.g. text classification, information retrieval) or individual words (e.g. synonymy tests). Given that a large fraction of the information available today, on the Web and elsewhere,consists of short text snippets (e.g. abstracts of scientific documents, imagine captions, product descriptions),in this paper we focus on measuring the semantic similarity of short texts. Through experiments performed on a paraphrase data set, we show that the semantic similarity method out-performs methods based on simple lexical matching, resulting in up to 13% error rate reduction with respect to the traditional vector-based similarity metric.

III. CONCLUSION

Measuring semantic similarity of concepts is a crucial component in many applications which has been presented in the introduction. In this paper we have reviewed various techniques for computing similarity index. Of all the techniques that we have surveyed we have found a technique for faster evaluation. That is propose wpath semanticsimilarity method combining path length with IC. The basicidea is to use the path length between concepts to representtheir difference, while to use IC to consider the commonality between concepts.

Bibliography

[1]. K. Bollacker,C. Evans,P. Paritosh, T. Sturge,andJ. Taylor,"Free- base:acollaboratively createdgraphdatabaseforstructuring humanknowledge," inProceedings of the2008ACMSIGMOD internationalconference on Managementof data. ACM, 2008, pp.1247–1250.

[2]. C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, andS. Hellmann, "Dbpedia-a crystallization pointfor the webof data," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, no. 3, pp. 154 – 165, 2009, the Web of Data.

[3]. J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum, "Yago2: A spatially and temporally enhanced knowledge base from wikipedia (extended abstract)," in Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, ser. IJCAI '13AAAI Press, 2013, pp. 3161–3165.

[4]. I.Horrocks, "Ontologies and the semantic web," Commun.ACM, vol.51, no.12, pp.58–67, Dec. 2008. [Online]. Available: http://doi.acm.org/10.1145/1409360.1409377

[5]. G. A. Miller, "Wordnet: a lexical database for English," Communications of the ACM, vol. 38, no. 11, pp. 39–41, 1995.

[6]. Z.Wu and M.Palmer,"Verbs semantics and lexical selection," in Proceedings of the 32nd annual meeting on Association for Computational Linguistics, ser. ACL '94.Stroudsburg, PA, USA: Association for Computational Linguistics, 1994, pp. 133–138.

[7]. Y. Li, Z.Bandar, and D.Mclean,"Anapproach formeasuring semantic similarity between wordsusing multiple information sources," Knowledge andDataEngineering,IEEE Transactionson, vol. 15, no. 4, pp. 871–882, 2003.

[8]. J. J. Jiang andD.W.Conrath, "Semantic similarity basedon corpus statistics and lexical taxonomy," Computational Linguistics, vol. cmp-lg/970, no. Rocling X, p. 15, 1997.

[9]. D. Lin, "An information-theoretic definition of similarity," in Pro- ceedings of the Fifteenth International Conference

ISSN: 2454-4248
419 – 423

on Machine Learning, ser.ICML '98.San Francisco,CA, USA:Morgan Kaufmann Publishers Inc., 1998, pp. 296–304.

[10]. M. Pontiki, D.Galanis, H.Papageorgiou, S.Manandhar, and I. Androutsopoulos,"Semeval-2015 task12: Aspectbasedsenti- mentanalysis," in Proceedings of the 9th International Workshop on Semantic Evaluation(SemEval 2015), Association for Computational Linguistics, Denver, Colorado, 2015, pp. 486–495