Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 01, January -2018

MEASURING SEMANTIC RELATIONSHIP OF MODEL IN INFORMATION GRID

GUDA SRIDHAR

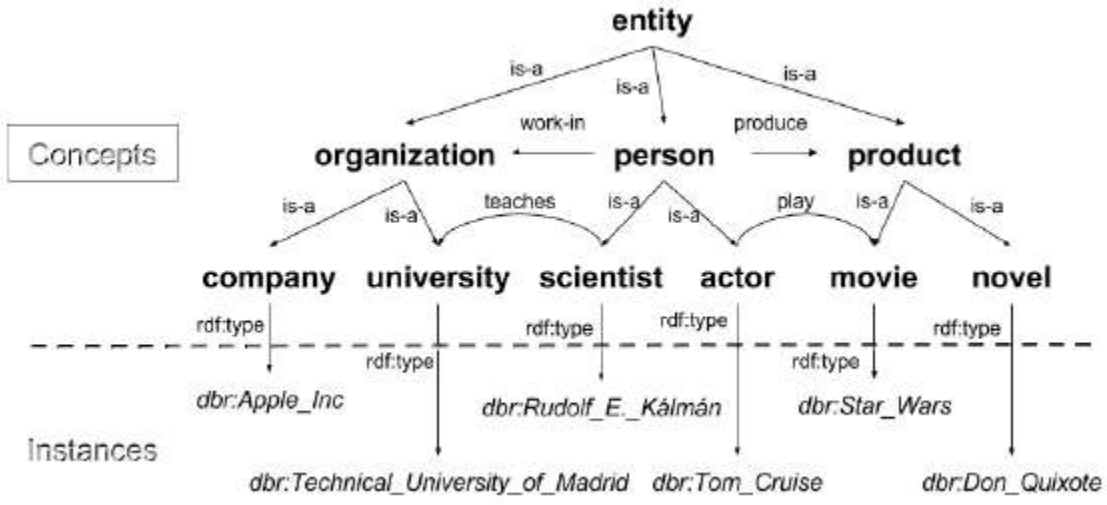
* (Department of CSE, JNTUH College of Engineering Manthani, INDIA)

Abstract: *This paper gives a way for measuring the semantic similarity between standards in Knowledge Graphs (KGs) which includes WordNet and DBpedia. Previous work on semantic similarity methods have focused on either the structure of the semantic community among principles (e.G. Direction period and intensity), or only on the Information Content (IC) of principles. We recommend a semantic similarity technique, namely wpath, to combine these two approaches, the use of IC to weight the shortest path period among ideas. Conventional corpus-primarily based IC is computed from the distributions of ideas over textual corpus, that is required to prepare a site corpus containing annotated standards and has excessive computational fee. As instances are already extracted from textual corpus and annotated through ideas in KGs, graph-based totally IC is proposed to compute IC primarily based on the distributions of concepts over instances. Through experiments accomplished on widely recognized phrase similarity datasets, we display that the wpath semantic similarity method has produced statistically full-size improvement over other semantic similarity strategies. Moreover, in a real category type assessment, the wpath method has shown the first-class performance in terms of accuracy and F rating.*

Index Terms: *Semantic Similarity, Semantic Relatedness, Information Content, Knowledge Graph, WordNet, DBpedia*

I. INTRODUCTION

With the increasing recognition of the connected facts initiative, many public Knowledge Graphs (KGs) have grow to be available, which include Freebase [1], DBpedia [2], YAGO [3], which are novel semantic networks recording tens of millions of ideas, entities and their relationships. Typically, nodes of KGs consist of a fixed of standards C1;C2; : : : ;Cn representing conceptual abstractions of factors, and a hard and fast of instances Ii; I2; : : : ; Im representing real international entities. Following Description Logic terminology [4], information bases comprise two varieties of axioms: a hard and fast of axioms is referred to as a terminology container (TBox) that describes constraints at the structure of the domain, much like the conceptual schema in database placing, and a hard and fast of axioms is known as announcement field (ABox) that announces facts about concrete conditions, like facts in a database placing [4]. Concepts of the KG include axioms describing idea hierarchies and are usually refereed as ontology classes (TBox), while axioms about entity instances are generally referred as ontology instances (ABox). Fig. 1 indicates a tiny instance of a KG using the above notions. Concepts of TBox are constructed hierarchically and classify entity instances into different types (e.G., actor or film) thru a special semantic relation rdf:type1(e.G., dbr:Star Wars is a instance of idea film). Concepts and hierarchical relations (e.G., is-a) compose a idea taxonomy that's a concept tree where nodes denote the standards and edges denote the hierarchical relations. The hierarchical members of the family among standards specify that a concept Ci is a type of concept Cj (e.G., actor is a person).



Apart from hierarchical relationships, concepts can have other semantic relationships among them (e.g., actor plays in a movie). Note that the tiny KG is a simplified example from DBpedia for illustration, and Table 1 shows examples of DBpedia entities and their types which are mapped to the example KG in Fig. 1.
The lexical database WordNet [5] has been conceptualized as a traditional semantic network of the lexicon of English words. WordNet can be regarded as a concept taxonomy where nodes denote WordNet synsets representing a fixed of phrases that percentage one common feel (synonyms), and edges denote hierarchical relations of hypernym and hyponymy (the relation among a sub-concept and a splendid idea) between synsets. Recent efforts have transformed WordNet to be accessed and implemented as concept taxonomy in KGs by changing the conventional illustration of Word- Net into novel related information illustration. For example, KGs such as DBpedia, YAGO and BabelNet [6] have integrated WordNet and used it as part of idea taxonomy to categorize entity times into different sorts. Such integration of traditional lexical assets and novel KGs have provided novel opportunities to facilitate many different Natural Language Processing (NLP) and Information Retrieval (IR) duties [7], inclusive of Word Sense Disambiguation (WSD) [8], [9], Named Entity Disambiguation (NED) [10], [11], query interpretation [12], report modeling [13] and question answering [14] to call a few. Those KG-based totally packages rely on the understanding of standards, instances and their relationships. In this work, we especially exploit the concept stage information, whilst the example level knowledge is used to aid the concept information. More specially, we cognizance at the problem of computing the semantic similarity between concepts in KGs.

II. METHODOLOGY

There is a relatively large number of semantic similarity metrics which were previously proposed in the literatures. Among them, there are mainly two types of approaches in measuring semantic similarity, namely corpus-based approaches and knowledge-based approaches [25]. Corpus based semantic similarity metrics are based on models of distributional similarity learned from large text collections relying on word distributions. Two words will have a high distributional similarity if their surrounding contexts are similar. Only the occurrences of words are counted in corpus without identifying the specific meaning of words and detecting the semantic relations between words. Since corpus based approaches consider all kinds of lexical relations between words, they mainly measure semantic relatedness between words. On the other hand, knowledge-based semantic similarity methods are used to measure the semantic similarity between concepts based on semantic networks of concepts. This section reviews briefly corpus-based approaches (Section 2.1) and knowledge-based semantic similarity metrics that have been observed good performance in NLP or IR applications (Section 2.2).
Corpus-based Approaches: Corpus-based strategies measure the semantic similarity among standards primarily based on the records received from big corpora such as Wikipedia. Following this concept, a few works take advantage of concept associations which include Point sensible Mutual Information [26] or Normalized Google Distance [27], while some other works use distributional semantics techniques to symbolize the concept meanings in excessive-dimensional vectors including Latent Semantic Analysis [28] and Explicit Semantic Analysis [29]. Recent work based on allotted semantics techniques consider advanced computational fashions including Word2Vec [30] and GLOVE [31], representing the words or standards with low-dimensional vectors.

III. TECHNIQUES IMPLEMENTED

The foremost idea of the wpath semantic similarity method is to encode each the structure of the idea taxonomy and the statistical facts of ideas. Furthermore, which will adapt corpus-based IC strategies to structured KGs, graph based IC is proposed to compute IC based at the distribution of standards over times in KGs. Consequently, using the graph-based IC inside the wpath semantic similarity method can represent the specificity and hierarchical structure of the principles in a KG. Section 3.1 gives the wpath semantic similarity technique for measuring semantic similarity between principles in KGs and Section three.2 describes the proposed method to compute graph-based IC of principles primarily based on KGs.
WPath Semantic Similarity Metric: The information-based totally semantic similarity metrics noted in the previous phase are specially advanced to quantify the diploma to which ideas are semantically similar the usage of records drawn from idea taxonomy or IC. Metrics take as input a couple of standards, and go back a numerical price indicating their semantic similarity. Many applications depend on this similarity rating to rank the similarity among one-of-a-kind pairs of principles. Take a fragment of WordNet concept taxonomy in Fig. 2 as example, given the idea pairs of (red meat; lamb) and (beef; octopus), the applications require similarity metrics to present better similarity price to sum (red meat; lamb) than sum (red meat; octopus) due to the fact the concept beef and concept lamb are sorts of meat whilst the concept octopus is a form of seafood. The semantic similarity rankings of a few idea pairs computed from the semantic similarity techniques have been illustrated in Table. 2. It can be seen on this desk how the row of idea pair (beef; lamb) has better similarity rankings than the row of idea pair (beef; octopus).
Graph-Based Information Content: Conventional corpus-primarily based IC calls for to prepare a domain corpus for the idea taxonomy after which to compute IC from the area corpus in offline. The inconvenience lies within the high

computational fee and difficulty of making ready a website corpus. More particularly, to be able to compute corpus-based totally IC, the concepts inside the taxonomy need to be mapped to the phrases in the area corpus. Then the advent of standards is counted and the IC values for concepts are generated. In this way, the extra area corpus training and offline computation may prevent the software of those semantic similarity methods relying on the IC values (e.g., res, lin, jcn, and wpath) to KGs, especially when the area corpus is insufficient or the KG is often up to date. Since KGs already mined structural information from textual corpus, we present a convenient graph-based IC computation approach for computing the IC of principles in a KG based totally on the instance distributions over the concept taxonomy. The graph-based totally IC is proposed to immediately take benefit of KGs while preserving the concept of corpus-based totally IC representing the specificity of ideas. In outcome, the IC-based totally semantic similarity technique such as res, lin, jcn and the proposed wpath can compute the similarity score between ideas immediately counting on the KG. As we noted previously in Section 1, ideas in KGs are typically represented as TBox and arranged into idea taxonomies. Those ideas categorize entity times of ABox into different sorts through the special relation rdf:type. For example, the idea film agencies all movie instances in DBpedia. Moreover, if idea A is a discern concept of concept B and idea C within the taxonomy, then the set of times of A is the union of the times of B and C. In different words, a idea in KG could have multiple entity times indicating the semantic kind of the ones entities, whilst an example can have more than one standards to describe entity classes from fashionable to particular. For example, a DBpedia entity example dbr:Tom Cruise may have numerous concepts describing its kinds from general to specific, Person, Actor, AmericanFilmActo.

IV. CONCLUSION

Measuring semantic similarity of ideas is a vital component in lots of packages which has been supplied within the creation. In this paper, we recommend wpath semantic similarity technique combining direction duration with IC. The primary idea is to apply the path length between ideas to symbolize their distinction, while to apply IC to take into account the commonality among ideas. The experimental results display that the wpath approach has produced statistically full-size development over other semantic similarity strategies. Furthermore, graph-primarily based IC is proposed to compute IC primarily based at the distributions of ideas over times. It has been shown in experimental effects that the graph-based IC is powerful for the rest, lin and wpath strategies and has similar performance because the traditional corpus-based totally IC. Moreover, graph-based totally IC has some of advantages, since it does now not requires a corpus and permits on-line computing primarily based on to be had KGs. Based on the evaluation of a easy factor category class task, the proposed wpath method has also proven the high-quality performance in terms of accuracy and F score.

V. FUTURE SCOPE

In this paper, we evaluated the proposed method in the word similarity dataset and simple classification using the most established evaluation method. More evaluation of semantic similarity methods in other applications considering the taxonomical relation could be useful and can be one of our future works. Furthermore, this paper mainly discussed semantic similarity rather than general semantic relatedness. Therefore, another future work could be in studying the combination of knowledge-based methods with the corpus-based methods for semantic relatedness. Finally, since we combined WordNet and DBpedia together in this paper, we would further explore using the proposed approaches for measuring the entity similarity and relatedness in KG.

VI. REFERENCES

[1] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, “Freebase: a collaboratively created graph database for structuring human knowledge,” in Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008, pp. 1247–1250.

[2] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, “Dbpedia-a crystallization point for the web of 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 ’13. AAAI 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] R. Navigli and S. P. Ponzetto, “Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network,” Artificial Intelligence, vol. 193, pp. 217–250, 2012.

[7] E. Hovy, R. Navigli, and S. P. Ponzetto, "Collaboratively built semi-structured content and artificial intelligence: The story so far," Artificial Intelligence, vol. 194, pp. 2 – 27, 2013, artificial Intelligence, Wikipedia and Semi-Structured Resources.
[8] R. Navigli, "Word sense disambiguation: A survey," ACM Computing Surveys (CSUR), vol. 41, no. 2, p. 10, 2009.
[9] A. Moro, A. Raganato, and R. Navigli, "Entity linking meets word sense disambiguation: a unified approach," Transactions of the Association for Computational Linguistics, vol. 2, pp. 231–244, 2014.
[10] J. Hoffart, S. Seufert, D. B. Nguyen, M. Theobald, and G. Weikum, "Kore: Keyphrase overlap relatedness for entity disambiguation," in Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ser. CIKM '12. New York, NY, USA: ACM, 2012, pp. 545–554.
[11] I. Hulpus, N. Prangnawarat, and C. Hayes, "Path-based semantic relatedness on linked data and its use to word and entity disambiguation," in International Semantic Web Conference, 2015.
[12] J. Pound, I. F. Ilyas, and G. Weddell, "Expressive and flexible access to web-extracted data: A keyword-based structured query language," in Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD '10. New York, NY, USA: ACM, 2010, pp. 423–434.
[13] M. Schuhmacher and S. P. Ponzetto, "Knowledge-based graph document modeling," in Proceedings of the 7th ACM International Conference on Web Search and Data Mining, ser. WSDM '14. New York, NY, USA: ACM, 2014, pp. 543–552.

ABOUT AUTHOR:

Mr. Guda Sridhar is currently working as a Lecturer in Department of Computer Science and Engineering in Jawaharlal Nehru Technological University Hyderabad College of Engineering Manthani. I received M.Tech in Computer Science and Engineering from Jawaharlal Nehru Technological University College of Engineering Jagtial and B.Tech in Information Technology from Kakatiya University,Warangal,Telangana, India.

