

Semantic Relatedness Measures in E-learning: A study

R.Sunitha

Department of Computer Science
Pondicherry University

G.Aghila

Department of Computer Science
Pondicherry University

ABSTRACT

In this paper, a detailed study of works that have been carried out in finding the semantic relatedness or relatedness (in short) of Learning Objects (LO) in the context of E-learning has been presented. Learning Objects are small instructional chunks of learning elements which can be archived, extracted and shared in the learning process. Semantic relatedness in general specifies the degree of relatedness between two concepts in a taxonomy computed using different types of relations defined between the concepts. Semantic relatedness measures have been used in applications like Word sense Disambiguation, Information Retrieval, Natural Language Processing, Query Expansion etc. In the context of E-learning, there are several scenarios like learning object sequencing, query answering, scaffolding, clustering etc. where the computation of semantic relatedness between LO has promising scope. But only few works have been carried out in the quantification of semantic relatedness between LO. The objective of this paper is to present the existing semantic relatedness measures in general and with respect to learning object in specific and to analyze the adeptness of the measures.

Keywords

E-learning, Learning Objects, Semantic Relatedness Measures

1. INTRODUCTION

The proliferous growth of Internet has facilitated the proportional growth of applications like e-learning. E-learning refers to the use of digital medium for the purpose of learning. E-learning has spawned from the asynchronous self-paced learning to synchronous collaborative experiences with web-based support. Thus E-learning can be defined as the use of Internet technologies to deliver a broad array of learning solutions with the objective of enhancing knowledge and performance [20]. The Internet has facilitated e-learning with a massive amount of rich learning resources which can be termed as information objects [21]. But these crude learning resources in their native form cannot usher efficient e-learning solutions. Hence this mass of content used in the context of learning should be represented in a completely new conceptual model [5] and thus the concept of Learning Objects comes into the picture of e-learning. Learning Objects (LO) are small instructional chunks of learning elements which can be archived, extracted and shared. Metadata standards facilitates the shareability and generativity of LO. LO repositories enable the reusability, shareability and scalability of LO. Learning Object repositories are electronic spaces that are database driven for storing the LOs. Meta-data or tags are the data about the different attributes that describe the learning objects. One of the attributes of LO specifies the type of associativity or relation a LO has with other LOs. By exploiting this knowledge of the relationship between the LOs, better solutions can be given to the learner. Semantic relatedness refers the degree to which two concepts are related. Though semantic relatedness sounds similar to

semantic similarity, it is not so. Semantic relatedness stands in general for the relatedness of concepts, unlike similarity that scores the likeness of the concepts. Semantic similarity can be said as a special case of Semantic relatedness. Semantic similarity and relatedness measures have been used in applications like Word sense Disambiguation (WSD), Information Retrieval (IR), Natural Language Processing (NLP), Query Expansion etc.[1][14] But only a few works have been carried out in quantifying the relatedness between LO in the context of e-learning. As quantification of LO relatedness has promising scope in delivering efficient e-learning solutions, an attempt has been made to study the status of Semantic relatedness measures in e-learning. In this paper, the significant works that have been carried out in the computation of Semantic relatedness of words and concepts in general has been presented. Works that have been carried out with respect to relatedness of LO has also been discussed. The rest of the paper is organised as follows. Section two introduces the LO and its associated concepts like metadata standards and Learning Object Repositories (LOR). Section three discusses about the significant contributions in computing Semantic relatedness of concepts. Section four presents the study of work with respect to relatedness of LO. The inferences derived from the study have been tabulated in Section five. The conclusion is presented in Section Six.

2. LEARNING OBJECTS

2.1 Definition

Learning Objects (LO) are small instructional chunks of learning elements which can be archived, extracted and shared in the context of E-learning. Learning objects are grounded by the object-oriented paradigm which facilitates archiving and sharing thereby eliminating the need for recreation, facilitating collaboration and serving as a rich base for e-learning. The LO may be of different forms like text, video, audio, graphics and may serve as tutorials, scenarios, simulations, lesson modules, case studies and assessment.

The conceptualization of Learning Object was done by Gerard in 1967, and the term was coined by Wayne Hodgins in 1994. The term 'Learning Object' cannot be attached to a single definition. The study of literature reveals different definitions of learning objects based on their granularity, usage and attributes.

The most accepted definition is that of Wiley stated as "any digital resource that can be reused to support learning"[22]. IEEE defines a learning object as "any entity, digital or non-digital, that may be used for learning, education or training"[10].

2.2 LO Meta-data and LOR

The inclusion of LO in E-learning solutions becomes effective only if they are accessible, reusable, general, and scalable. To enable these qualities of LO, meta-data standards and learning object repositories are evolving. Meta-data or tags are the data

about the different attributes that describe the learning objects. Different Meta-data standards have involved. The Dublin Core Metadata Initiative, The Learning Technology Standards Committee (LTSC) of the Institute of Electrical and Electronics Engineers (IEEE), Advanced Distributed Learning (ADL) initiative of Department of Defence are the prominent Meta-data standards. In order to enable LO extractable and sharable it has to be stored. Thus Learning Object repositories are database-driven electronic spaces for storing the LOs. LOR are equipped with well-defined user interfaces to facilitate interactivity. Some of the important LORs are Multimedia Educational Resource for Learning and Online Teaching (MERLOT), Wisconsin Online Resource Center Wisc-Online Learning Object Project, European Knowledge Pool System (ARIADNE) and MACE.

2.3 LO and Semantic Relations

Learning objects have been designed as standalone educational resources meant to satisfy a specific objective. But by relating different Learning objects may it be cross-media object, the learner can be provided with the best learning experience. Learning objects may differ in complexity, granularity, type of media etc. and may possess semantic or didactic dependency with one another. This dependency is tagged as relation in the metadata describing the LOs. These relations may be terminological, domain specific or domain independent semantic relations. The LOs and their relations with other LOs can be represented in a taxonomy. The concept, relation ideology can be mapped to ontologies.

3. CLASSIFICATION OF RELATIONS

Relations in general defined between concepts can be classified as illustrated in figure 1. The important classification is terminological and semantic relations. Semantic relations are further classified as domain specific and domain independent relations.

Terminological relations are defined between terminological units or concepts of a domain. Generic-specific, part-whole, entities-function and cause-effect are the important terminological relations [3].

Semantic relations are relations between concepts or meanings. The important semantic relations are class inclusion, attribution, antonyms, synonyms[17].

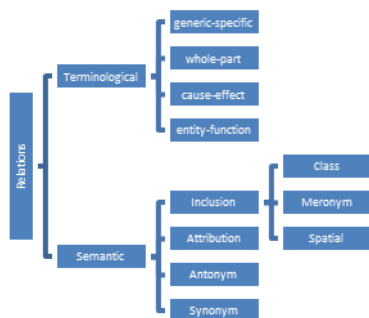


Fig 1. Classification of relations

4. SEMANTIC SIMILARITY VS SEMANTIC RELATEDNESS

Similarity has been defined in Merriam-Webster as “the quality or state of being similar” whereas the term relatedness has been defined as “connected by reason of an established or discoverable relation”. Semantic relatedness quantifies the degree of relatedness between two concepts in a taxonomy by

using all relations (which may be any kind of relations) between them. Semantic similarity measure uses only the hyponymy/hypernymy (i.e. is-a) relations. Like semantic similarity measures, semantic relatedness measures are used in information retrieval and Natural Language Applications.

5. SEMANTIC RELATEDNESS MEASURES

There are three different approaches to calculate the semantic similarity between concepts which can be extended to semantic relatedness as well. Figure 2 illustrates the approaches namely Path based, Information Content based and Text based measures and the major contributions in each category.

5.1 Path based Measures

The path based measures makes use of the number of nodes in the shortest path between two concepts as parameters for computing the relatedness. The simplest path based measure is of Rada et.al [18] which uses the concept of semantic distance. The semantic distance measure is inversely proportional to the relatedness measure. The semantic distance is computed as a function of node count and is given by the sum of nodes that lie in the shortest path between two concepts. Leacock and Chodorow [9] have proposed a similarity measure based on the path length and the maximum depth D of the taxonomy. This normalized measure is given by

$$lch(c_1, c_2) = -\log \frac{\text{length}(c_1, c_2)}{2D}$$

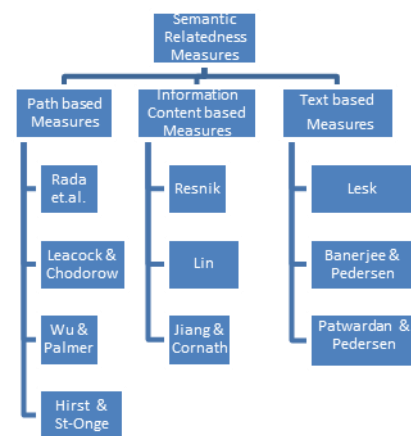


Fig 2. Classification of semantic relatedness measures

Wu and Palmer [23] use the depth of the two concepts and the depth of their Least Common Subsumer (LCS) to calculate the similarity score between the two concepts and is given by

$$wup(c_1, c_2) = \frac{\text{depth}(LCS(c_1, c_2))}{\text{depth}(c_1) + \text{depth}(c_2)}$$

Hirst and St-Onge [4] has proposed a relatedness measure that is based on the nature of the path between two concepts. If the two concepts are connected by a path that is not too long and with minimum changes in direction, then it signifies high degree of relatedness between the concepts.

The advantage of path based measures is their simplicity. But the limitations of path based measures are that they assume

only the semantic relations esp. *is-a* relation. The path based measures were hence used in finding the similarity between nouns only.

5.2 The Information Content based Measures

In this approach the similarity is computed as a function of the Information Content (IC) which is defined as the measure of the specificity of a concept. The extent to which the two concepts share the information is used to compute the similarity between the two concepts. The similarity measure proposed by Resnik [19] is given by

$$Sim_{res} = IC(LCS(c_1, c_2))$$

IC is defined as $IC = -\log(P(c))$ where $P(c)$ is the probability of encountering a concept 'c' in a large corpus.

The information content based approach for computing similarity proposed by Lin [12] is given by

$$Sim_{lin} = \frac{2 * IC(LCS(c_1, c_2))}{IC(c_1) + IC(c_2)}$$

Jiang and Cornath [6] have proposed a measure similar to Lin which computes the difference. The similarity score is inversely proportional to the distance score given by

$$dist_{j\&c} = IC(c_1) + IC(c_2) - 2 * IC(LCS(c_1, c_2))$$

Following these prominent works, variants of these measures have been reported in the literature.

The information content based measures are simple like path based measures and they are insensitive to the problem of varying link distances [19]. Since they act on both, the knowledge of the corpora and also the taxonomic structure it is possible to handle multiple contexts. The disadvantage of information content based measures is that of path based measures considering nouns and assuming *is-a* relation between concepts.

5.3 Text based Measures

The text based measures or corpus based measures make use of the probability of co-occurrence of words in a large text or a corpus.

5.3.1 Gloss Overlap measures

A classic text based relatedness measure is that of Lesk [11], who proposed semantic relatedness based on gloss overlap. An extended gloss overlap approach was proposed by Banerjee and Pedersen [1] in which the overlap score was computed by extending the gloss of the concepts under consideration so as to include the glosses of the related concepts in the hierarchy.

5.3.2 Context Vector measure

Patwardhan et.al. [16] uses the concept of Context vectors to measure the semantic relatedness of concepts. The contextual profile of concepts is represented as co-occurrence vectors. The relatedness of the concepts is measured using the cosine of the angle between the two vectors.

5.4 Applications of Semantic Relatedness measures

The applications of semantic relatedness have been witnessed in applications like NLP, spelling correction, Word Sense Disambiguation, Information Retrieval etc. A very few works

have been reported in the literature regarding the use of semantic relatedness measures in e-learning. The degree of relatedness has been computed between learning objects and in the following section discusses about them in detail.

6. SEMANTIC RELATEDNESS MEASURES IN E-LEARNING

In this section, the novel approaches used in calculating the semantic relatedness between Learning Objects has been presented.

In [7] the authors related learning objects based on their usage context and claim the hypothesis that similarity of usage indicates content similarity. In this work the learner's action are recorded and the pre and post context usage of a learning object is collected. A Usage Context Profile (UCP) is created for each learning object consisting of the learning objects used before and after this said learning object. The Jaccard similarity measure or cosine similarity measure is used to calculate the similarity between the elements of the usage contexts based on the treatment of the usage context as a set or bag respectively. The similarity or relatedness between two objects O1 and O2 is calculated using the arithmetic mean or median of the similarity of the UCP of O1 and O2. The Pearson correlation Coefficient is used to evaluate the correctness of the usage based similarity by correlating the results with the semantic meta-data based similarity calculated using the cosine similarity measure. The authors have also evaluated the effectiveness of the usage based similarity against the manual content-based similarity. The experiment carried out specifies that the usage based similarity is an indication of content based similarity.

In [2] the authors of the adaptive system *My online teacher* have developed a method for computing similarity between the concepts by calculating correspondence weight computation between concept attributes.

In [8] the authors have proposed a path based semantic relatedness measure using object properties in an ontology. The proposed measures combines the information theoretic approach to semantic similarity, the heterogeneity of relations and the classification of semantically correct and incorrect paths while calculating the semantic relatedness between the concepts.

In [13] the authors have used a metadata based approach to determine the similarity among Learning Objects based on information represented in their metadata. The XML structures used to represent the LO metadata are subjected to similarity evaluation using a fuzzy approach. A search component designed as a part of a Learning Object Management system uses the similarity score computed using the proposed metric facilitating searching based on Semantic Query, Learning Objects Patterns and Query by Examples.

In [15] the authors have proposed an algorithm namely Hofmethode based on occurrence context of words to compute semantic similarity using the text in the corpus. The computed semantic similarity scores are used to construct a semantic map of e-learning products.

7. INFERENCE FROM THE STUDY

The study of literature reveals the fact that the different semantic relatedness measures have been proposed can be classified as knowledge based and corpus based measures. Though these measures claim to score the relatedness between the concepts, they provide the score of semantic similarity between the concepts only. This is due to the fact that these measures consider a restricted set of relations – mostly *is-a* relation only. Only a few measures concentrate on heterogeneous relations. The term relatedness depends on the

type of relations which exists between concepts. The significance of the relation between the concepts are least considered in these measures. In the works related to e-learning also, the domain specific relations between LO are not considered and other strategies have been devised to derive the relatedness. Table 1 summarizes the results of the study. With respect to applications of semantic relatedness measures, the Semantic relatedness measures are used mainly in Information Retrieval, Natural Language Processing, Word Sense Disambiguation, Spelling Correction, Query expansion etc. E-learning is an excellent candidate for semantic relatedness computation and the semantic relatedness scores computed for learning object can be used in recommending learning objects, scaffolding, learning sequence generation, query answering etc.

Table 1. Summary of the semantic relatedness measures

| Classification | Approach | Mechanics | Advantage | Dis-advantage |
|---|-------------------------------------|---|---|--|
| Classic semantic relatedness measures | Path based measures | Uses the taxonomic structure | Simplicity | Only nouns and is-a relations are considered |
| | Information content based measures | Use the frequency of words in the corpus | | |
| | Text based measures | Uses taxonomic structure and probability of occurrence of words | | |
| Semantic relatedness measures in E-learning | Usage Context based measure | Based on the usage context of the LO | Simplicity | The information content and semantics not considered |
| | Meta data based measure | Based on standard meta data | Relies only on the object properties | The structural knowledge of the taxonomy is not considered |
| | Lexical co-occurrence based measure | Based on the occurrence context of words | Simplicity | |
| | Hofmethode | | | |
| | Path based measure | Combines IC and path based techniques | Corpus and taxonomic structure are considered | The significance of heterogeneous relations not considered |

8. CONCLUSION AND FUTURE WORK

In this paper a detailed study of the significant works that have been carried out in the computation of semantic relatedness between concepts has been presented. The advantages and disadvantages of those measures have been listed. E-learning is evolving as an interesting Internet application and is more open to mechanisms and strategies for enhancing the learner's learning experience. In that aspect the status quo of semantic relatedness in e-learning has been analysed. The study reveals the different works carried out in the computation of semantic relatedness between LO and also indicates that due implication to the different relations defined between LO has not been considered in the computation of the semantic relatedness between LO. Also there is a lacuna in employing the semantic relatedness measures for other applications of E-learning. Hence the future work is to propose a metadata based semantic relatedness measure combining the knowledge about the taxonomic structure considering the heterogeneous relations defined between LO and to apply the measure for different applications of e-learning to check the adeptness of the measure.

9. REFERENCES

- [1] Banerjee S, Pedersen T. 2002 An adapted Lesk algorithm for word sense disambiguation using WordNet. In Proceedings of the 3rd International conference on intelligent text processing and computational linguistics.
- [2] Cristea. A.I., De Mooji, 2003 Designer adaptation in adaptive hypermedia authoring . In proceedings of the International conference on Information technology.
- [3] Elizabeth Marshman, Julie L. Gariépy and Charissa Harms, University of Ottawa, 2012 Helping language professionals relate to terms: Terminological relations and term bases. Journal of Specialized translation. Issue 18.
- [4] Hirst G., St-Onge D. 1998 Lexical chains as representations of context for the detection and correction of malapropisms. C. Fellbaum (Ed.), WordNet: An electronic lexical database, MIT Press.
- [5] Hodgins, H.W. 2000. The future of learning objects in D.A. Wiley (Ed.). The Instructional Use of Learning Objects.
- [6] Jiang J, Conrath D. 1997 Semantic similarity based on corpus statistics and lexical taxonomy. In Proceedings of the 10th International conference on research in computational linguistics.
- [7] Katja Niemann, Maren Scheffel, Martin Friedrich, Uwe Kirschenmann, Hans-Christian Schmitz, Martin Wolpers, 2010, Usage-based Object Similarity, Journal of Universal Computer Science, vol. 16, no. 16.
- [8] Laurent Mazuel and Nicolas Sabouret , Semantic relatedness measure using object properties in an ontology , white Paper
- [9] Leacock C., Chodorow M. 1998 Combining local context and WordNet similarity for word sense identification. C. Fellbaum (Ed.), WordNet: An electronic lexical database, MIT Press.
- [10] Learning technology standards committee 2000 Available: <http://ltsc.ieee.org/>

- [11] Lesk. M, 1986, Automatic sense disambiguation using machine readable dictionaries. How to tell a pine comes for an ice cream cone. In proceedings of the SIGDOC conference.
- [12] Lin D. 1998 An information-theoretic definition of similarity. In Proceedings of the 15th International Conference on Machine Learning.
- [13] Menendez-Dominguez, V.H. ; Zapata, A. ; Prieto-Mendez, M.E. ; Romero, C, 2011, A similarity-based approach to enhance learning objects management systems , In proceedings of the 11th International Conference on Intelligent Systems Design and Applications .
- [14] McCarthy D, Keoling R, Weeds J, Carroll J. 2004 Finding predominant word senses in untagged text. In Proceedings of the 42nd meeting of the association for computational linguistics.
- [15] Oliver Michel, Damian Läge, 2009, The Hofmethode: Computing semantic similarities between elearning products. In proceedings of the Interactive conference on computer aided learning.
- [16] Patwardhan S, Pedersen T. 2006 Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In Proceedings of the EACL 2006 workshop.
- [17] Storey. Veda C. 1993, understanding Semantic Relationships. VLDB Journal Issue 2.
- [18] Rada R, Mili H, Bicknell E, Blettner M. 1989 Development and application of a metric on semantic nets. IEEE transactions on systems, man and cybernetics Volume 19 Issue 1.
- [19] Resnik P. 1995 Using information content to evaluate semantic similarity in a taxonomy. In Proceedings of the 14th International joint conference on artificial intelligence.
- [20] Rosenberg, Marc J. 2001 E-Learning: strategies for delivering knowledge in the digital age. McGraw-Hill Companies, Inc.,
- [21] The Herridge group, Learning objects and instructional design .White paper.
- [22] Wiley, David A. 2000 Learning Object design and sequencing theory. Doctoral thesis. Brigham Young University.
- [23] Wu. Z, Palmer. M, 1994 Verb semantics and lexical selection. In proceedings of the 32nd annual meeting of the associations for computational linguistics.