A modification of Wu and Palmer Semantic Similarity Measure

Djamel Guessoum
Dept. of Electrical Engineering
École de Technologie Supérieure,
Montréal, Canada
e-mail: djamel.guessoum.1@ens.etsmtl.ca

Moeiz Miraoui
university of Gafsa,
Gafsa, Tunisia
e-mail: moeizmiraoui@gmail.com

Chakib Tadj
Dept. of Electrical Engineering
École de Technologie Supérieure,
Montréal, Canada
e-mail: Chakib.Tadj@etsmtl.ca

***Abstract*-Context-aware applications are intended to facilitate the adaptation of services in a pervasive computing system. The semantic similarity between contexts and the application of a semantic similarity measure as a mechanism for service adaptation are topics that have yet to be thoroughly explored in the literature. The most developed semantic similarity measures are those applied to the ontological / taxonomic representation of the context. The Wu and Palmer semantic similarity measure is one of these measures that is characterized by its simplicity and high performance, but it can give inaccurate results because two concepts in the same hierarchy of an ontology may show a lower similarity than two concepts belonging to different hierarchies. In this work, we present a modification to improve the accuracy of this measure.**

Keywords-semantic similarity; Wu and Palmer; Taxonomy; context.

I. INTRODUCTION

Semantic similarity measures are used in various fields with different types of applications. In pervasive computing, the application of these measures is linked to the concept of "context" and its impact on the adaptation of services provided to the user. Several studies apply these measures to service recommendation systems [10], in which context is represented by the user's profile-related preferences. K. Ning and D. O'Sullivan [11] developed the similarity measure between ontological concepts of [3] by including context and allocating the weight of relations between concepts.

Mention should be given to applications of similarity measures in other domains that can be used in the field of pervasive computing, such as data mining [8], or the research of Slimani et al. [19], which improved the semantic similarity measure of Z. Wu and M. Palmer [22] by taking into account the context of the measure.

A pervasive computing system is designed to provide services to a user by minimizing his direct involvement, and to this end, the few studies applying semantic similarity measures have each given a particular definition to the context and its specific purpose. Examples include M. Kirsch-Pinheiro et al. [7], who proposed a dynamic adaptation of services to solve the problem of incomplete information in the process of choosing the adequate service in a particular context. Y. Benazzouz [1] used the same type of similarity measures for clustering data in order to determine the particular situations triggering a particular service.

The remainder of this paper is organized as follows. Section 2 introduces some applications of the semantic similarity between contexts in pervasive computing, while Section 3 introduces the semantic similarity measures and context variables. In Section 4, semantic similarity measures applied to ontologies are shown and finally, in Section 5 we introduce our proposed modification of the Wu Palmer measure. The conclusion is presented in Section 6.

II. RELATED WORK

The identification of the current context is defined by the contextual information related to the triggering of a service as well as a situation or "current context" in the set of current contextual information, similar to a known situation or context [1], with each identified situation being linked to one or more of the services to be provided. This identification forms the basis of the rule-based adaptation mechanism, which is a set of conditional rules with the form: if (contextual information I) then (service S).

Identifying a context is based on data mining techniques. Once identified, semantic similarity measures are applied in order to compare it with contexts with known services. P.Y Gicquel [4] modeled the spatio-temporal context of a museum visitor in an ontological form, with the semantic similarity measures being used to recommend artwork similar to the interests of the user by comparing the properties of two concepts in the knowledge base. The similarity measure is a modified version of the similarity proposed by G. Pirró, and J. Euzenat [12], which combines the similarity calculation based on Tversky's model with that of informational content.

Y. Benazzouz [1] and F. Ramparany et al. [14] applied semantic similarity measures to group data and "pure" contexts based on the measures of [13] [23].

 ISBN: 978-1-61208-505-0

A similar approach was proposed by M. Kirsch-Pinheiro et al. [6] for the adaptation of content found in an intelligent device with a Pervasive Computing System (PCS). The authors used semantic similarity measures to assess the degree of matching between the predefined profiles of situations and the current context of the user with the aim of prioritizing them, using a graph-modeled context [25].

Semantic similarities between contexts in a PCS are thus based on the collection of one or several elements of contextual data that are relevant to one or several services. The description and semantic relations of these services are described in an ontological form, thus allowing the application of known semantic similarity measures.

Many variations of the Wu and Palmer measure are present in the literature. We will mention the work of Slimani et al. [19] in which a penalty function is integrated in the measure to penalize concepts belonging to different hierarchies and the measures of C. Leacock, and M. Chodorow [8] and Y. Li et al. [9], where each trying to make adjustments on a particular aspect of the measure of Wu and Palmer. All these measures are difficult to implement and add an extra computational load to the original measure.

III. SEMANTIC SIMILARITY MEASURES AND CONTEXT VARIABLES

The most frequently cited definition of context is that of A. K. Dey [2] who defines context in the following manner: “any information that can be used to characterize the situation of an entity (person, object or physical computing.” This definition clearly resembles that of B. Schilit et al. [20] since the context is conceived as a set of information collected from a user environment (person), physical environment (physical object), or system environment, with the purpose of data collection being the characterization of these environments.

The data set that characterizes a context is collected from several sources of information, for example, physical sensors in the environment, intelligent devices, virtual sensors, Internet access, or even telecommunication service providers; this information is thus very heterogeneous. In accordance with several previous studies [5][10], The contextual information can be categorized in 3 classes, as shown below:

1. Quantitative variables are expressed in scalar or vector form (i.e., temperature, latitude, longitude, altitude).
2. Quantifiable variables are expressed in qualitative or ordinal form (i.e., large, small, first, second).
3. Categorical variables are not quantifiable. Variables of this type are described as a set of characteristics (e.g., standing, sitting).

The global approach to measuring the similarity between contexts is primarily based on calculating local similarities between attributes or context variables [16]. The global similarity (1) can then be calculated based on these local similarities by weighting each attribute:

$$Similarity(Context_{new}, Context_{old}) = \frac{\sum_{i=1}^{n} w_i \times Sim(a_i^{Context_{new}}, a_i^{Context_{old}})}{\sum_{i=1}^{n} w_i} \quad (1)$$

where w_i is the weight of the attribute a_i, $a_i^{Context_{new}}$ is the attribute i of the new context, and $a_i^{Context_{old}}$ is the attribute i of the existing context.

IV. NOTION OF SEMANTIC SIMILARITY

In pervasive computing, where the notion of context plays a very important role, the semantic similarity measure is a tool to evaluate the resemblance between instances of a context. It allows services to be chosen and classified according to their relevance to a given query, and a user’s profile and preferences

A. Semantic similarity measures applied to ontologies

The most developed semantic similarity measures in recent years, based on the ontological representation of knowledge and especially in its taxonomic form, were described by D. Sánchez et al. [17]. The authors categorised the semantic similarity measures on the counting of arcs, characteristics of concepts, and information content.

Semantic similarity measures based on the counting of arcs were introduced by R. Rada et al. [13]. The basic notion for these measures was the fewer the number of arcs separating two concepts, the greater their similarity.
Among the studies using this approach we find:

- *Rada measure*

It is based on the fact that we can calculate the semantic similarity between two concepts in a hierarchical structure (ontology) with links, such as "is-a" by calculating the shortest path between these concepts.

- *Wu and Palmer measure*

Several variants based on the Rada measure have been proposed to improve some aspects, such as Z. Wu, and M. Palmer [22] applied to an ontology O (Fig. 1), who considered the depth of ontology in the measure, because two concepts in lower levels of ontology are more specific and are more similar. This measure is given by:

$$Sim_{WP}(X,Y) = \frac{2 \times N}{N_1 + N_2} \quad (2)$$

where Sim_{WP} is Wu and Palmer similarity, N_1 and N_2 are the number of arcs between the concepts X , Y and the ontology root R and N is the number of arcs between the LCS and the ontology root R.

We chose to modify the semantic similarity measure proposed by Z. Wu, and M. Palmer [22] (2) because it is simple to implement in a pervasive computing system where the context is modeled using an ontology and gives realistic similarity results. Nevertheless, we modified the Wu and Palmer measure to eliminate an inherent disadvantage, in which two concepts in the same hierarchy may show a lower similarity than two concepts belonging to different hierarchies [16] [18] [19].

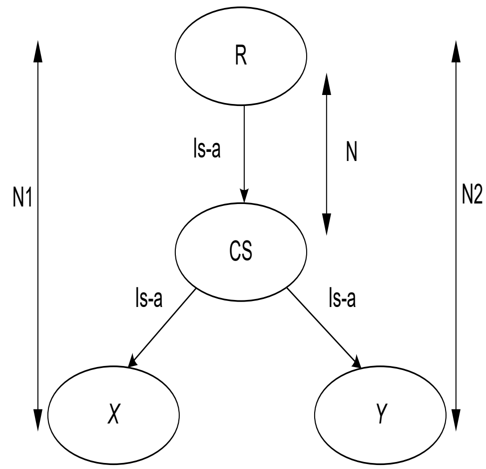

Figure 1. Wu and Palmer Ontology example

Several other measures were subsequently introduced by C. Leacock, and M. Chodorow [8] and Y. Li et al. [9], as the authors attempted to make adjustments for a particular aspect of Wu and Palmer's measure.

Semantic similarity measures based on the characteristics of concepts derive from the similarity model of [21], in which two concepts are more similar if 1) they share more common characteristics and 2) less non-common characteristics. However, the determination of the weighting parameters represents a major challenge for this type of measures.

Finally, semantic similarity measures based on the information content of a common concept involving two concepts to be compared were first introduced by P. Resnik [15]. Their dependency on the design of the ontology and their lack of consideration for the context are some of their limitations.

V. MODIFIED WU AND PALMER SIMILARITY MEASURE

As it was shown from the disadvantages of the Wu and Palmer semantic similarity measure, is that with this measurement one can obtain inaccurate results [16] [18] [19]. See the following example (Fig. 2):

$Sim_{WP}(c1, c2) = \frac{2*1}{(1+4)} = 0.4$ (LCS=Person, N=1, N1=1, N2=4)

$Sim_{WP}(c2, c3) = \frac{2*2}{(4+3)} = 0.57$ (LCS=Employee, N=2, N1=4, N2=3)

It is clear that the semantic similarity measures applied to the UnivBench ontology (Ontology from the educational field, used to describe data on universities and their departments [19] [24], $Sim_{WP}(c1, c2) < Sim_{WP}(c2, c3)$, despite the fact that the concepts c1 and c2 belong to the same hierarchy.

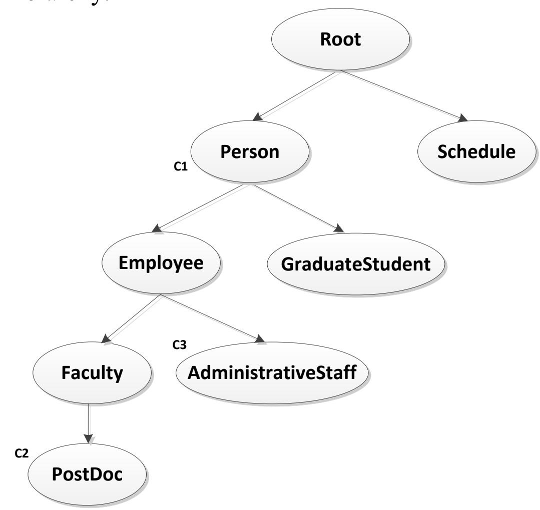

Figure 2. Extract from UniveBench. Ontology

The following modification is proposed to remedy this disadvantage (Fig. 3):

$$sim(c1, c2) = \begin{cases} Sim_{WP}(c1, c2) = \dfrac{2N}{N1 + N2} & if\ N1 \neq \mathrm{N}\ and\ N2 \neq \mathrm{N} \\ \dfrac{2N}{N2 - N}\ if\ (N1 = N)\ and\ \dfrac{2N}{N1 - N}\ if\ (N2 = N) \end{cases}$$

1- Two concepts belong to different hierarchies if : $N1 \neq \mathrm{N}\ and\ N2 \neq \mathrm{N}$ $sim(c1, c2) = Sim_{WP}(c1, c2)$

2- Two concepts belong to the same hierarchy if : N1=N or N2 =N,

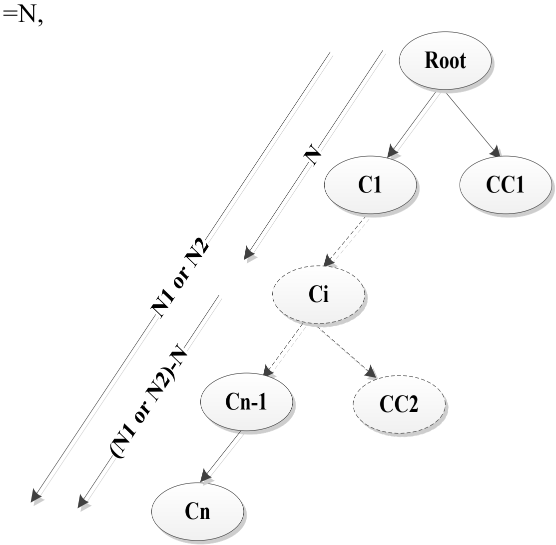

Figure 3. Modified Wu and Palmer Measure

Sim(Ci,Cj)> Sim(Ci,CC) ∀ i,j=1,...n

Where: Ci,Cj are two concepts of the same hierarcy and CC is a different hierarchy concept.

1- If N1=N, $\frac{2N}{N2-N} > \frac{2N}{N1+N2} \leftrightarrow N2 - N < N1 + N2$
2- If N2=N, $\frac{2N}{N1-N} > \frac{2N}{N1+N2} \leftrightarrow N1 - N < N1 + N2$

The proposed modification meets the four criteria of similarity measures: non-negativity, identity, symmetry, and uniqueness, as defined below:
1) non-negativity: Sim(A, B)≥0,
2) identity : Sim (A, A) = Sim (B,B)=1
3) symmetry: Sim (A, B) = Sim (B, A)
4) uniqueness : Sim (A, B)=1 → A=B

It is also clear that the semantic similarity between two concepts that belong to the same hierarchy is inversely proportional to the distance between these two concepts (N2-N or N1-N) and is always greater than the semantic similarity between a concept of that hierarchy and another concept from another hierarchy. It has all the advantages of the Wu and Palmer measure, namely its implementation simplicity and expressiveness.

This modified Wu and Palmer measure applied to the example of Fig. 2 gives the following results:

$Sim\ (c1, c2) = \frac{2*1}{(4-1)} = 0.66$
(LCS=Person, N=1, N1=1, N2=4)
$Sim\ (c2, c3) = \frac{2*2}{(4+3)} = 0.57$
(LCS=Employee, N=2, N1=4, N2=3)

VI. CONCLUSION

The proposed modification of the Wu and Palmer semantic similarity measure retains all the benefits of this measure namely its implementation simplicity and power to give close similarities to the reality unlike several other changes proposed in the literature. It also meets the criteria of semantic similarity measures namely the non-negativity, the identity, the symmetry and uniqueness. Its advantage is the fact that all the concepts in the same hierarchy must be more similar to each other than other concepts of a different hierarchy and the similarity between the concepts in the same hierarchy also depends on the distance between these concepts.

REFERENCES

[1] Y. Benazzouz, “Context discovery for the automatic adaptation of services in ambient intelligence,” (Doctoral dissertation, Ecole Nationale Supérieure des Mines de Saint-Etienne), pp. 71-98, 2011.

[2] A. K. Dey, “Understanding and using context,” Personal and Ubiquitous Computing, 5, 4-7, pp. 4, 2001..

[3] P. Ganesan, H. Garcia-Molina, and J. Widom, “Exploiting hierarchical domain structure to compute similarity,” ACM Transactions on Information Systems (TOIS), 21(1), pp. 64-93, 2003.

[4] P.Y Gicquel, “Semantic and contextual similarities for informal learning in mobility,” RJC EIAH’2012, pp. 45-50, 2012.

[5] K. C. Gowda and E. Diday, “Symbolic clustering using a new similarity measure,” IEEE Transactions on Systems, Man, and Cybernetics, 22(2), pp. 368-378, 1992.

[6] M. Kirsch-Pinheiro, M. Villanova-Oliver, J. Gensel, and H. Martin, “A personalized and context-aware adaptation process for web-based groupware systems,” In 4th International Workshop on Ubiquitous Mobile Information and Collaboration Systems, CAiSE’06 Workshop, pp. 884-898, 2006.

[7] M. Kirsch-Pinheiro, Y., Vanrompay, and Y. Berbers, “Context-aware service selection using graph matching,” In 2nd Non Functional Properties and Service Level Agreements in Service Oriented Computing Workshop (NFPSLA-SOC'08), ECOWS. CEUR Workshop proceedings, Vol. 411, pp. 10-12, 2008, November.

[8] C. Leacock, and M. Chodorow, “Combining local context and WordNet similarity for word sense identification,” In WordNet: An electronic lexical database, MIT Press, pp. 265-283, 1998.

[9] Y. Li, Z. Bandar, and D. McLean, “An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources,” IEEE Transactions on Knowledge and Data Engineering, 15, pp.871-882, 2003.

[10] L. Liu, F. Lecue,, N. Mehandjiev, and L. Xu, “Using context similarity for service recommendation,” In Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on , pp. 277-284, 2010, September.

[11] K. Ning and D. O'Sullivan. “Context modeling and measuring for context aware Knowledge Management,” International Journal of Machine Learning and Computing 2, no. 3, 2012.

[12] G. Pirró and J. Euzenat, “A feature and information theoretic framework for semantic similarity and relatedness,” In The Semantic Web–ISWC 2010, , Springer Berlin Heidelberg, pp. 615-630, 2010.

[13] R. Rada, H. Bicknell, E. Mili, and M. Blettner, “Development and application of a metric on semantic nets,” IEEE Transaction on Systems, Man, and Cybernetics, 1(19), pp. 17-30, 1989.

[14] F. Ramparany, Y. Benazzouz, J. Gadeyne, and P. Beaune, “Automated context learning in ubiquitous computing environments,” In SSN, pp. 9-21, 2011.

[15] P Resnik, “Using information content to evaluate semantic similarity,” In Proceedings of the 14th International Joint

Conference on Artificial Intelligence, August 20-25; Montréal Québec, Canada, pp. 448-453, 1995.

[16] M. Richter, Weber, Rosina “case based reasoning,” Textbook, , Springer Heidelberg New York Dordrecht London , p. 546, 2013,

[17] D. Sánchez, M. Batet, D. Isern, and A. Valls, , “Ontology-based semantic similarity: A new feature-based approach,” Expert Systems with Applications, 39(9), pp. 7718-7728, 2012.

[18] K. C. Shet and U. D. Acharya, “A New Similarity Measure for Taxonomy Based on Edge Counting,” arXiv preprint arXiv:1211.4709, 2012.

[19] T. Slimani, B. B. Yagahlane and K. Mellouli. “A new similarity measure based on edge counting,” Proceedings of the World Academy of Science, engineering and Technology 17, pp. 3, 2006.

[20] B. Schilit, N. Adams and R. Want, “Context-aware computing applications,” In IEEE Workshop on Mobile Computing Systems and Applications . Santa Cruz, CA, US, pp. 85-90, 1994.

[21] A. Tversky, “Features of similarity,” Psycological Review, 84(4), 1977.

[22] Z. Wu and M. Palmer, “Verbs semantics and lexical selection,” InProceedings of the 32nd annual meeting on Association for Computational Linguistics, pp. 133-138. Association for Computational Linguistics, 1994, June.

[23] J. Zhong, H. Zhu, J. Li, and Y. Yu, , “Conceptual graph matching for semantic search,” In Proceedings of the 10th International Conference on Conceptual Structures (ICCS), Springer-Verlag, London, pp. 92-196, 2002.

[24] M. Dean, G. Schreiber, S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, and L. A. Stein, “OWL web ontology language reference,” W3C Recommendation February, 10, 2004.

[25] D. Guessoum, M. Miraoui, and C. Tadj, “Survey of semantic similarity measures in pervasive computing,” International journal on smart sensing and intelligent systems, 8(1), pp. 125-158, 2015.