International Journal of Hybrid Information Technology
Vol.8, No.8 (2015), pp.303-314
http://dx.doi.org/10.14257/ijhit.2015.8.8.31

Concept Similarity Measure with Hierarchy Structure and Information

Cong Dai[1], Dongmei Li[1], Hui han[1*], Qichen Han[2] and Jiajia Hou[3]

[1]School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China
[2]University of Chinese Academy of Sciences, College of Engineering & Information Technology, Beijing, 100049, China
[3]School of Information, Renmin University of China, Beijing100872, China
hanhuie@126.com

Abstract

Calculating semantic similarity between concepts in ontology is an important issue in natural language processing and so on, so how to measure the similarity becomes a hot topic among many scholars. However, most existing methods cannot distinguish the similarity further. Confronting this problem, we propose a new semantic similarity method combining hierarchy structure of ontology and information content of two concepts based on domain ontology, which highlights the semantic information of leaves in the ontology structure. Our experiment demonstrates that, compared with other available methods, our proposal can improve the accuracy between two leaves and between leaf and non-leaf.

Keywords: *domain ontology; hierarchy structure; information content of concept; concept similarity*

1. Introduction

With the rapid development of Internet, the quantity of data is set to explode in terms of growth. To solve the problem that data are stored out of order, more data are stored in form of ontology. The measurement of semantic similarity between concepts in ontology is exploited in several research fields such as natural language processing, information retrieval and other related applications. Nowadays, ontology-based methods of calculating the semantic similarity between concepts are widely applied in many natural language processing tasks such as semantic annotation [1], word sense disambiguation [2], information extraction [3,4] or relation extraction [5]. Domain ontology possesses specific knowledge and provides a centralized introduction. Hence, calculating concept similarity in domain ontology is meaningful.

Currently, there exist many ontology-based methods to calculate concept similarity [6, 7], mainly including hierarchy structure-based measures, information content-based measures, feature-based measures and gloss-based measures. We focus on the former two methods below.

The depth of two concepts in the ontology structure and the edge connecting two concepts, especially the length of the shortest path between two concepts are regarded as factors. Wu proposed a measure that considers the impact of two compared concepts and their Nearest Common Ancestor (NCA) [8]. In [9], a measure is derived considering the depth of two concepts and the shortest path using is-a hierarchy for concepts in WordNet. Hao thought that the similarity between two concepts is determined by the depth of NCA and the shortest path [10]. However, the methods presented above agree the weight of every edge is the same, just thinking about the hierarchy of concepts but ignore the

ISSN: 1738-9968 IJHIT

intrinsic information in the ontology structure. The link strength of edge is without consideration, either. Therefore, the similarity cannot be distinguished between two concepts with the same depth accurately by using only hierarchy structure-based measures.

The main idea of information content-based approaches is combining concepts with ontologies, using information of concepts to measure the similarity. Resnik proposed that the similarity was depended on how much common information they shared, which was represented by the information content of their NCA [11]. Meanwhile, Resnik used the frequency of a concept to measure its information content in a given corpus. However, Resnik's metric is strict with the style of corpus and needs a large and comprehensive corpus for a more accurate result. Lin extended the Resnik's work, the modification consisted of the combination of information content of compared concepts and their NCA's [12]. It is advocated in [13] that the number of the concepts which concept *c* subsumes and the number of all the concepts in a taxonomy could contribute to getting similarity between concepts. Sánchez [14] reflected the number of leaves that two compared nodes subsume into the calculation. One of the problems of the metrics above ignores the hierarchy structure of ontology so that they cannot distinguish two concepts with the same information content but in different depth precisely by using only information content-based measures.

In fact, either method above has its advantages and disadvantages, so we propose a new efficient approach combining the two factors, synthetically considering the depth and the information content of compared concepts to make up for each other's limitation. Zhou presented a comprehensive metric which not only considered the information content of two concepts and their NCA, but also used depth to distinguish different concepts [15]. Hadj proposed that the hyponyms of a given concept should be reckoned into the computation of information content, meanwhile, the depth in a taxonomy does contribute a lot to the similarity [16]. Meng's metric is inspired by Seco's and Zhou's work, improving the depth's function in calculating the similarity [17]. The shortest path, the hierarchy structure of ontology and the superior concepts coincidence degree were proposed to compute similarity in [18]. Owing to the fact that leaves are the lowest in ontology, they own more semantic information than other nodes, it is assumed in [19] that the special function of leaves. Our proposal is based on this idea and Wu's method with the reference to local density in [20]. Moreover, we integrate the features of orchard pests and diseases domain ontology into the calculation. The purpose of this paper is to design a Concept Similarity Measure with Hierarchy Structure and Information (CSM_HSI), which improves the calculation of information content and considers the structure information more carefully. Finally, we verify the effectiveness of our algorithm by experiment.

2. Concept Similarity Measure with Hierarchy Structure and Information

2.1. Relevant Definitions

Definition 1 (Semantic Cluster) in the ontology, the semantic cluster consists of the leaves that concept *c* subsumes, denoted as *SC(C)*. Especially, if *C* is a leaf, $SC(c) \subset \varnothing$.

Definition 2 (Local Density) in the ontology, we define the number of the siblings of *C* plus 1 as *LD(C)*.

Definition 3 (Semantic Probability) in the ontology, we denote *P(C)* to be the sematic probability of concept *C*.

Definition 4 (Ancestor Collection) in the ontology, the ancestor collection consists of all the concepts that subsume concept *C*, denoted as *A(C)*. Especially, if concept c is a root node, $A(C) \subset \varnothing$.

Definition 5 (Generation Collection) in the ontology, the generation collection consists of all the concepts that concept *C* subsumes, denoted as *G(C)*. Especially, if concept c is a leaf node, $G(C) \subset \varnothing$.

Definition 6 (Limb Collection) in the ontology, the limb collection consists of concept C and its *A(C)*, *G(C)*, denoted as *Limb(C).*

Definition 7 (Depth Section) in the ontology, we denote the depth of the root node is 1. We define $DS(C_i,NCA)$ as the depth section of C_i and $DS(C_j, NCA)$ as the depth section of C_j.

2.2. Hierarchy Structure in Ontology

We use *DS(C,NCA)* to express the intrinsic features of a concept in the ontology structure. Among the structure-based measures, calculating the shortest path and the path from compared concepts to their NCA is the commonest (*e.g.*, [21,8]). There still has room for improvement. *DS(C,NCA)* is presented below:

$$DS(C, NCA) = dep(C) - dep(NCA) \tag{1}$$

where dep(C) and dep(NCA) are the depth of concept C and NCA in ontology.

And the modification of the previous method based on structures is as follows:

$$Sim_{Depth}(C_i, C_j) = 2 \times dep(NCA) / (DS(Ci, NCA) + DS(Cj, NCA) + 2 \times dep(NCA)) \tag{2}$$

2.3. Information Content in Ontology

We adopt the idea of Resnik that *IC* value is calculated by negative log likelihood equation (3). From the formula, it’s not hard to find that the result is closely depended on the accuracy of *P(C)*. The traditional approaches used to get similarity by calculating the frequency of a concept in a corpus. So corpus-based methods are limited by the type of the corpus. Li considered the ontology’s inner structure, proposing a method called Bottom-up concept probability computation method (B-U), which took the semantic information of leaves into account [19]. The computation equation is as follows:

$$IC(C) = -\log P(C) \tag{3}$$

$$p(p) = \begin{cases} \dfrac{1}{l_{count}} & p \text{ is a leaf} \\ \sum_{i=1}^{c(p)} p(c_i) & p \text{ is not a leaf} \end{cases} \tag{4}$$

where *p(p)* is the probability of concept *p* and l_{count} is the sum of the leaves in the ontology. *c* is the direct hyponym of *p* and *i* represent the sequence of *p*’s children. $p(c_i)$ refers to the probability of c_i and *c(p)* is the amount of *p*’s children. But there is a problem that, the *IC* of different leaves is the same, so that the method can neither distinguish two leaves or different leaves couples with the same NCA.

The modification makes full use of the semantic information of leaves in the ontology, including the local density and *SC(NCA)*, in order to differ the similarity between leaves. The *P(C)* formula is as follows:

$$P(C)=\begin{cases}\alpha\dfrac{LD(C)}{nc}+\beta\dfrac{1}{n} & C \text{ is a leaf}\\ \dfrac{nc}{n} & C \text{ is not a leaf}\end{cases} \tag{5}$$

where *LD(C)* represents the local density of *C* and *nc* refers to the sum of members in *SC(NCA).n* describes the total number of leaves. The parameter α , β are used to adjust the weight in the formula. Given the fact that the information of leaves in local position is more concrete than in the overall position, we make the value of α is bigger than β . Moving from top to bottom, the semantic information of a node is gradually increasing. Therefore, leaves contain the richest information in *Limb(C)*, concluding the concrete semantic information and representing the semantic features in the given domain ontology.

The value of the difference between two concepts measures how diverse they are, so we describe $diff(C_{i,},C_j)$ as the d-value of compared concepts below:

$$diff\left(C_i,C_j\right)=IC\left(C_i\right)-IC\left(C_j\right) \tag{6}$$

As we can see, the estimation of similarity between two compared concepts is related with their shared information and differences. The more shared information, the more similar they are. The more differences, the less similar they are. We know that *IC(NCA)* presents the information two concepts shares and $diff(C_i,NCA)$,$diff(C_j,NCA)$ show the differences between them. In order to avoid a zero denominator, let γ >0 to be a smoothing factor. We set $Sim_{IC}(C_{i,},C_j)$ as :

$$Sim_{IC}(C_i,C_j)=\frac{IC\left(NCA\right)+\gamma}{diff\left(C_i,NCA\right)+diff\left(C_j,NCA\right)+IC\left(NCA\right)+\gamma} \tag{7}$$

We propose an algorithm called Concept Information Content Similarity (CIC_Sim), which describes the concrete process of the calculation based on information content.

Algorithm 1 CIC_Sim algorithm

Input: two concepts

Output: IC-based similarity of two concepts

Step1. Iterate the orchard pests and diseases domain ontology, and then find the location of two compared concepts.

Step2. Find the parent nodes of two concepts respectively, and then get their NCA by using the queue.

Step3. Judge whether the two nodes are leaves or not, and then use equation (5) to get the value of $P(C_i)$, $P(C_j)$ and *P(NCA)*.

Step4. Put the value obtained from step4 into equation (1) to gain the IC of two concepts and NCA.

Step5. Put the value from step4 into equation (6), and then get the value of $diff(C_i,NCA)$ and $diff(C_j,NCA)$.

Step6. Use $IC(C_i)$, $IC(C_j)$,*IC(NCA)*, $diff(C_i,NCA)$ and $diff(C_j,NCA)$ to calculate similarity via equation (7).

2.4. Integration of Hierarchy Structure and Information Content

As for calculating similarity between concepts, if we only think of the structure of concepts, the intrinsic information of concepts is ignorant. And if we only take *IC* into consideration, we cannot distinguish concepts of different hierarchy precisely. Therefore, we need a metric with a combination of intrinsic and external factors. To make up for each other, this paper proposes a method with a weighted integration of depth and information content of concepts. It is defined as follows:

$$Sim(C_i, C_j) = \varepsilon \times Sim_{IC}(C_i, C_j) + \delta \times Sim_{Depth}(C_i, C_j) \quad (8)$$

where ε, δ are tuning factors to adjust the influence of each factor so that we can get better results . $\forall \varepsilon > 0, \delta > 0, \ \varepsilon + \delta = 1$. In this formula,

$\forall C_i, C_j \in T, \exists Sim(C_i, C_j) \in [0,1]$

Proof. Let's see equation (2) first:

- Supposed that compared concepts are the same.

  $DS(C_i, NCA) = dep(C_i) - dep(NCA) = 0$,

  $$\therefore Sim_{Depth}(C_i, C_i) = \frac{2 \times dep(NCA)}{2 \times dep(NCA)} = 1$$

- Supposed that compared concepts $C_{i,}, C_j$ have great differences.

  $DS(C_i, NCA) \to \infty, \ DS(C_j, NCA) \to \infty, dep(NCA) > 0$,

  We cannot decide whether the value of depth(NCA) is big or not, because it depends on the position where two concepts are.

  $\therefore Sim_{Depth}(C_i, C_j) \in (0,1]$.

  Let's review the formula (7):

- If the compared concepts are the same.

  $IC(C_i) = IC(NCA)$, and we can easily predict that,

  $diff(C_i, NCA) = 0$. Put the values into formula(7) and we can get the similarity.

  $$Sim_{IC}(C_i, C_j) = \frac{\gamma}{\gamma} = 1$$

- If the compared concepts $C_{i,}, C_j$ have great differences.

  $IC(NCA) \to 0, diff(C_i, NCA) \to \infty, \ diff(C_j, NCA) \to \infty$,

  $Sim_{IC}(C_i, C_j) \to 0, \therefore Sim_{IC}(C_i, C_j) \in [0,1]$.

  In a summary, $\forall \varepsilon > 0, \delta > 0$ 且 $\varepsilon + \delta = 1$,
  $\forall C_i, C_j \in T, \exists Sim(C_i, C_j) \in [0,1]$

□

Basd on formula (8), our CSM_HIS algorithm can be described as below:

Algorithm 2 CSM_HIS algorithm

Input: two concepts

Output: hierarchy and IC-based similarity of two concepts

Step1. Iterate the orchard pests and diseases domain ontology, and then find the location of two compared concepts.

Step2. Gain the depth and the IC of two compared concepts and their NCA.

Step3. Use $dep(C_i)$,$dep(C_j)$ and $dep(NCA)$ to get DS(C_i,NCA) and DS(C_j,NCA).

Step4. Obtain the $Sim_{Depth}\ (C_i, C_j)$ and $Sim_{IC}\ (C_i, C_j)$ of the compared concepts.

Step5. Put the value obtained from step4 into equation (8) to get a weighted similarity.

3. Experiment

3.1. Test Datasets and Setting of Relevant Parameters

The orchard pests and diseases domain ontology is used as the datasets, which is constructed by us. Meanwhile, the values of tuning factors in different formulas are shown in Table 1.

Table 1. Setting of Parameters

Parameter	Value
α	0.002
β	0.001
γ	3
ε	0.7
δ	0.3

3.2. Evaluation and Analysis

In order to evaluate the efficiency of our proposal, we compare it with Yang's [18] method and WSim_OC in [19]. 15 concepts pairs are taken from the databases in Table 2. We number the concepts pairs from 1 to 15, and the 1,2,7,10,11 are leaf pairs among them. By testing these 15 concept pairs with three methods, we can verify the rationality of our method. Meanwhile, according to the results in Table 2, Figure 1 directly reflects the similarity between concepts by three approaches.

Table 2. Partial Results of Concept Similarity

NO	Concept 1	Concept 2	WSim_OC	Yang	CSM_HSI
1	Concave shape	Concentric wheel shape	0.24	0.352	0.221
2	Ancient clock shape	Zigzag shape	0.35	0.281	0.304
3	Fruit	1cm in diameter	0.475	0.234	0.388
4	Leaf spot	Circle	0.629	0.441	0.529
5	Fruit spot	Leaf spot	0.396	0.42	0.532
6	Root	Silk screen shape	0.68	0.388	0.637
7	Apple	Pear	0.657	0.655	0.733
8	Disease	Disease of peach	0.702	0.451	0.771
9	Natural abscission	Atrophic abscission	0.781	0.735	0.842
10	Antennal segment number	Seven Antennal segment	0.782	0.735	0.842
11	Dark grey	Light grey	0.805	0.775	0.852
12	Branch spot	Branch spot size	0.853	0.591	0.897
13	Root exsiccation	exsiccation	0.873	0.591	0.914

14	Abscission of fruit	Natural abscission	0.872	0.591	0.914
15	Grey	Light grey	0.889	0.632	0.92

From the results presented in Figure 1, it shows our metric can evaluate the differences between compared concepts more precisely. The distribution of the values is regular so that we can predict that our proposal has certain stability to some extent. For better observation of different distribution, we paint scatter diagram of three methods respectively as Figure 2, Figure 3 and Figure 4.

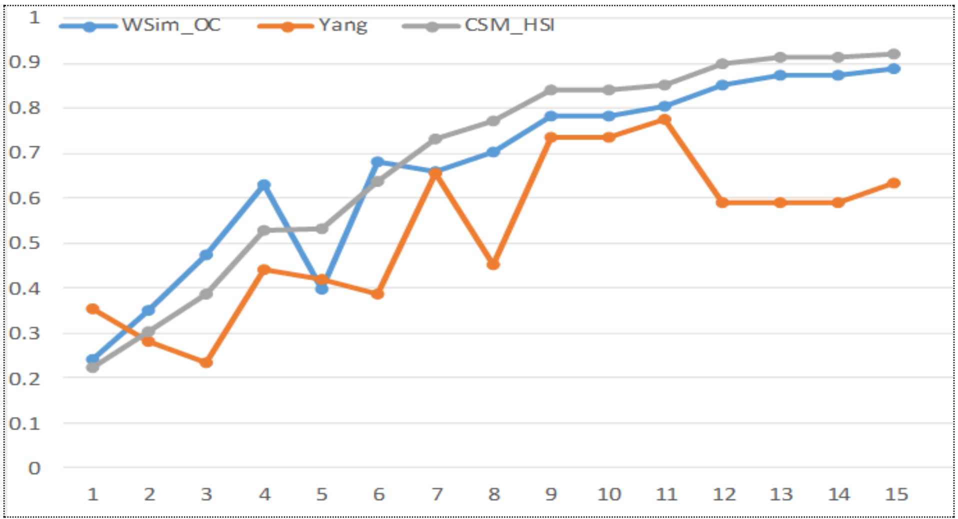


Figure 1. Comparison of Three Methods

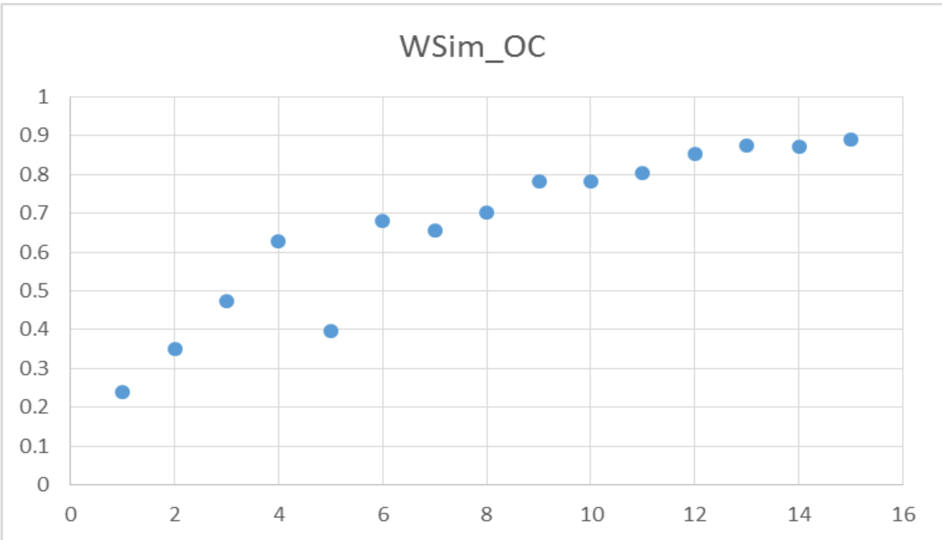


Figure 2. Similarity Distribution of WSim_OC

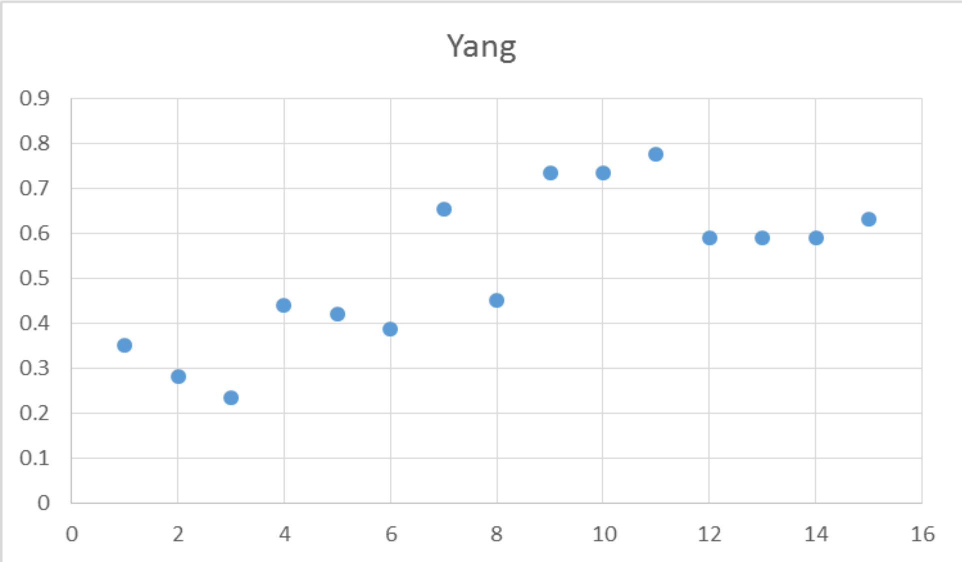


Figure 3. Similarity Distribution of Yang's Method

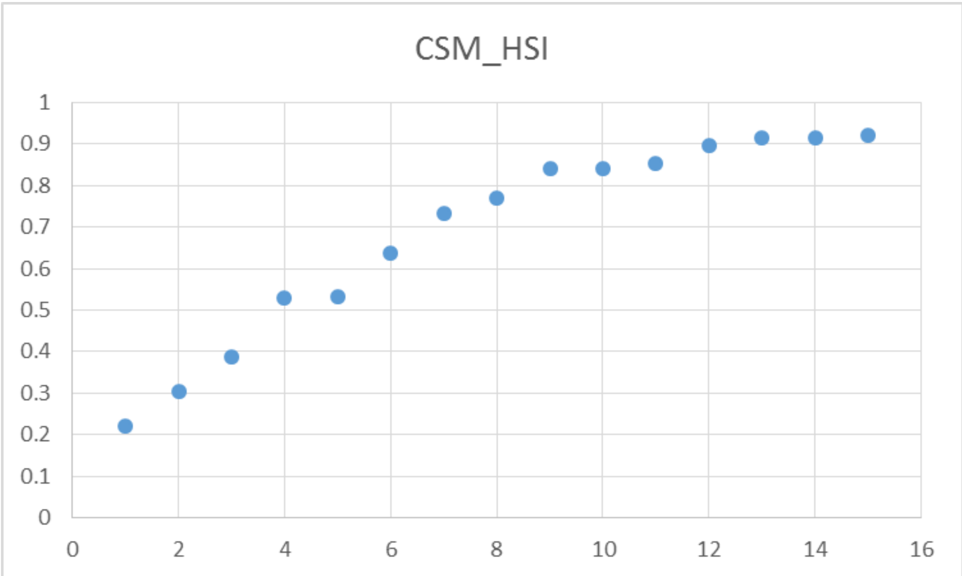


Figure 4. Similarity Distribution of Our Method

Moreover, some rules can be extracted from our experiment. Now we describe these rules with examples and analysis.

Rule 1. Assume that the d-value of two concept pairs' depths is 1. If there exists the parent-child relationship between them, they will be more similar than there does not exist parent-child relationship.

e.g. 1. We randomly choose two concept pairs A and B with different depth value. A exits the parent-child relationship, including "Antennal segment number" and "Seven antennal segments" whose depths are 4. B does not exit the parent-child relationship, including "Figure" and "Forewing shape" whose depths are 5. According to calculations by CSM_HSI, the former similarity is 0.914, and the latter similarity is 0.558, consistently with human judgment. So we can verify the validity of our rule and the rationality of our method.

Rule 2. Assume that compared concepts have the same depth. The nearer distance between two concepts and their NCA, the more similar they are. At the same time, the deeper their NCA is, the more similar they are.

e.g. 2. We randomly choose two concept pairs A and B whose depth are 6. A includes "Concave shape" and "Concentric wheel shape" and their NCA depth is 2. B includes "Radical pattern" and "Asterism pattern" and their NCA depth is 5. According to calculations by CSM_HSI, the former similarity is 0.221, and the latter similarity is 0.601, consistently with human judgment. The depth of A is less than B. With node moving from top to bottom, the semantic information becomes more concrete and contain more information content. Therefore, the information content of NCA between A concept pair is more than that between B. We can find the reason of generating differences via formula(3) and (5). So we can verify the validity of our rule and the rationality in our method.

Rule 3. Assume that two concepts are siblings. Generally, the deeper the two concepts are, the more similar they are.

e.g. 3. We randomly choose two concept pairs A and B with different depth value. A includes "Leaf spot" and "Fruit spot" whose depths are 4. B includes "Dark grey" and "Light grey" whose depths are 7. According to the results of CSM_HSI, the former similarity is 0.532, and the latter similarity is 0.852, consistently with human judgment. But it is acceptable that the similarity between deeper sibling concepts is less than the sibling concepts not very deep. According to CSM_HSI, we use a weighted method to calculate the similarity between two concepts. So, a concept not very deep may have rich information content. For example, the leaves may be not very deep in the ontology but may own much semantic information. The similarity of concept pair including "Apple" and "Pear" is 0.733, and the similarity of concept pair A is 0.532. Although the depth of "Apple" and "Pear" is 4, less than that of A, they have more information content. Apples and pears are really similar, in accordance with human judgment.

4. Conclusion

To distinguish the semantic similarity between concepts more accurately, we combine the hierarchy structure of the ontology and the information content of concepts. The intrinsic information content calculating is adopted in our metric and we also make some improvement using depth as a factor. The experiment demonstrates that our approach shows better results and improves the efficiency of calculating semantic similarity. Our proposal not only can be applied to orchard pests and diseases domain but other domain ontologies. In further research, there is still room for improving the influence of local density and depths in semantic similarity calculation.

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities.(No.TD2014-02), the National College Students' Training Programs of Innovation and Entrepreneurship(Research and Implementation of Computer Science Teaching Service Platform for Resource Sharing), Beijing Higher Education Reform(No. 2013-ms047) Beijing Forestry University Resources Sharing Course (Data Structure), and Beijing Forestry University Special Research of Campus Informatization (Construction of the Teaching Resources Sharing Platform for Program Design and Algorithms Courses).

References

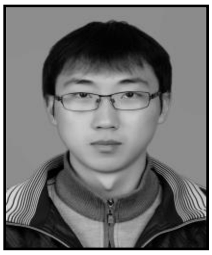
[1] D. Sánchez, D. Isern and M, "Millan. Content annotation for the semantic web: an automatic web-based approach", Knowledge and Information Systems, vol. 27, no. 3, **(2011)**, pp. 393-418.
[2] S. Patwardhan, S. Banerjee and T. Pedersen, "Using measures of semantic relatedness for word sense disambiguation", Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Text Processing(CICLing'03), **(2003)**, pp. 241-257; Berlin, Heidelberg.
[3] CHE Haiyan, SUN Jigui, JING Tao and Xi Bai, "An ontology-theme-based method of acquiring knowledge from Chinese natural language documents", Journal of Frontiers of Computer Science & Technology, vol. 1, no. 2, **(2007)**, pp. 206-215.
[4] J. Atkinson, A. Ferreira, E. Aravena, "Discovering implicit intention-level knowledge from natural-language texts", Knowledge-Based Systems, vol. 22, no. 7, **(2009),** pp. 502-508.
[5] D. T. Ouyang, J. F. Qu and Y. X. Ye, "Extending training set in distant supervision by ontology for relation extraction", Ruan Jian Xue Bao/Journal of Software, vol. 25, no. 9, **(2014)**, pp.2088-2101 (in Chinese). http://www.jos.org.cn/1000-9825/4638.htm
[6] L. Meng, R. Huang and J. Gu, "A review of semantic similarity measures in wordnet", International Journal of Hybrid Information Technology, vol. 6, no. 1, **(2013)**, pp. 1-12.
[7] M. Batet and D.Sánchez, A. Valls, "An ontology-based measure to compute semantic similarity in biomedicine", Journal of biomedical informatics, vol. 44, no. 1, **(2011)**, pp. 118-125.
[8] Z. Wu, M. Palmer, "Verbs semantics and lexical selection", Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics(ACL'94), (1994), pp. 133-138; Stroudsburg, PA.
[9] C. Leacock and M. Chodorow, "Combining Local Context and WordNet Similarity for Word Sense Identification", WordNet: An electronic lexical database, vol. 49, no. 2, **(1998)**, pp. 265-283.
[10] D. Hao, W. Zuo, T. Peng and F. He, "An approach for calculating semantic similarity between words using WordNet", Digital Manufacturing and Automation (ICDMA), 2011 Second International Conference on. IEEE, **(2011)**, pp. 177-180.
[11] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy", Proceedings of the 14th International Joint Conference on Artificial Intelligence(IJCAL'95), **(1995)**, pp. 448-453; San Francisco, CA.
[12] D. Lin, "An information-theoretic definition of similarity", Proceedings of the Fifteenth International Conference on Machine Learining(ICML'98), **(1998)**, pp. 296-304; San Francisco, CA.
[13] N. Seco, T. Veale, J. Hayes, "An intrinsic information content metric for semantic similarity in WordNet", Proceedings of 16th European Conference on Artificial Intelligence, including Prestigious Applicants of Intelligent Systems, **(2004)**, pp. 1089−1090; Valencia, Spain.
[14] D. Sánchez, M. Batet, D. Isern, "Ontology-based information content computation", Knowledge-Based Systems, vol. 24, no. 2,**(2011)**, pp. 297-303.
[15] Z. Zhou, Y. Wang, J. Gu, "A new model of information content for semantic similarity in WordNet", Proceedings of the International Conference on the Future Generation Communication and Networking Symposia(FGCNS'08), 2008 Second International Conference on. IEEE, **(2008)**, pp. 85-89.
[16] M. A. H. Taieb, M. B. Aouicha and A. B. Hamadou, "Ontology-based approach for measuring semantic similarity", Engineering Applications of Artificial Intelligence, vol. 36, **(2014)**, pp. 238-261.
[17] L. Meng, J. Gu, and Z. Zhou, " A new model of information content based on concept's topology for measuring semantic similarity in WordNet", International Journal of Grid and Distributed Computing, vol. 5, no. 3,**(2012)**, pp. 81-94.
[18] Y. Yuehua, D. Junping and H. Bowei, "A novel ontology-based semantic retrieval model for food safety domain", Chinese Journal of Electronics, vol. 22, no. 2, **(2013)**, pp. 247-252.
[19] L. Wenqing, S. Xin, Z. Changyou and F. Ye, "A Semantic Similarity Measure between Ontological Concepts", Acta Automatica Sinica, vol. 38, no. 2, **(2012)**, pp. 229-235.
[20] H. Z. Liu, H. Bao, D. Xu, "Concept Vector for Similarity Measurement Based on Hierarchical Domain Structure", Computing and Informatics, vol. 30, no. 5, **(2012)**, pp. 881-900.
[21] Y. Li, Z. A. Bandar, D. McLean, "An approach for measuring semantic similarity between words using multiple information sources", Knowledge and Data Engineering, IEEE Transactions on, vol. 15,no. 4, **(2003)**, pp. 871-882.

Authors

Cong Dai, she is an undergraduate student in School of Information and Technology at Beijing Forestry University. She joined Institute of Artificial Intelligence in Beijing Forestry University in 2013 as a research assistant. Her main research interests include intelligent information retrieval and semantic web.

Dongmei Li, she received her Ph.D. degree from Beijing Jiaotong University. Currently she is an associate professor in School of Information and Technology at Beijing Forestry University. Her main research interests include artificial intelligent, knowledge engineering and semantic web.

Hui Han, she received her Ph.D. degree from Tsinghua University. Currently she is a lecture in School of Information and technology at Beijing Forestry University. Her research interests include data mining and pattern recognition



Qichen Han, he received his bachelor degree from Beijing forestry University. Currently he is a master student in College of Engineering & Information Technology, University of Chinese Academy of Sciences. His main research interests include intelligent information retrieval and personalized recommendation.

Jiajia Hou, she received her bachelor degree from Beijing forestry University. Currently she is a master student in School of Information at Renmin University of China. Her main research interests include intelligent information processing and semantic web.

