Enriching Semantic Knowledge for WSD

Junpeng CHEN†[a], Nonmember and Wei YU††[b], Member

SUMMARY In our previous work, we proposed to combine ConceptNet and WordNet for Word Sense Disambiguation (WSD). The ConceptNet was automatically disambiguated through Normalized Google Distance (NGD) similarity. In this letter, we present several techniques to enhance the performance of the ConceptNet disambiguation and use this enriched semantic knowledge in WSD task. We propose to enrich both the WordNet semantic knowledge and NGD to disambiguate the concepts in ConceptNet. Furthermore, we apply the enriched semantic knowledge to improve the performance of WSD. From a number of experiments, the proposed method has been obtained enhanced results.

key words: ConceptNet disambiguation, Normalized Google Distance, semantic similarity measure, Word Sense Disambiguation

1. Introduction

In previous studies on knowledge-based WSD emphasis has been placed on how to exploit the knowledge resource, in which the most widely used knowledge resource was WordNet. However, the issue of applying WordNet-based methods on WSD usually achieved lower performance compared to supervised methods, mainly due the fact that the lexical and semantic knowledge contained in WordNet is not sufficient for WSD. One of the simplest techniques for improving performance is to enrich the knowledge of WordNet for WSD task. For example, Magnini & Cavaglia [1] adopted domain knowledge to assign domain labels to most WordNet synsets. Hwangetal [2] enriched semantic relations by means of the disambiguation of the glosses of WordNet. Since these WordNet-enriched methods are still limited in lexical and taxonomic knowledge, more sophisticated approaches have been proposed. C&R (2008) [3] extracted semantic relations from web to get more efficient and up-to-date semantic knowledge. P&N (2010) [4] exploited Wikipedia, a large collaborative web encyclopedia to extract knowledge for WSD.

In our previous work, we introduced the Common sense in ConceptNet [5] to enrich the knowledge resource for WSD task [6]. ConceptNet is a large collaborative web knowledgbase which focuses on psychological aspects of everyday life. Since it contains many ambiguous concepts,

ConceptNet cannot be directly used in WSD. Thus, we developed an automatic disambiguation approach based on NGD in ConceptNet disambiguation.

In this letter, we propose an extended semantic similarity measure to improve the performance of sense disambiguation on ConceptNet. One of the significant drawbacks of the NGD method in ConceptNet disambiguation in the previous work is that only the web knowledge has been considered, and thus there was too much noise which reduced the performance. For the purpose of alleviating this problem, we adopt expert knowledge in WordNet to enrich the semantic knowledge to disambiguate ConceptNet. Then, we apply the enriched semantic knowledge to improve the performance of WSD. The experimented results demonstrate that the proposed technique obtains enhanced results on the SemEval Coarse Word Sense Disambiguation task.

2. NGD-based ConceptNet Disambiguation

The ConceptNet contains nearly one million of assertions represented as triples like (concept1, relation, concept2) to define the concrete semantic relations between two specific concepts. However, ConceptNet cannot be directly used for WSD purpose due to the existence of polysemy and synonymy of the concept in it. Normalized Google Distance (NGD) [7] was proposed to measure semantic relatedness between two terms using the vast available knowledge on the Web. Concretely, NGD takes advantage of the number of hits returned by search engine such as Google. In this section, we briefly describe ConceptNet disambiguation based on NGD measure.

A simulation of the human’s processing of disambiguating is carried out by three steps to disambiguate ConceptNet. Given a ConceptNet assertion (c, relation, d), where concept c is ambiguous (the cases of d being ambiguous or both c and d being ambiguous are the similar). In order to disambiguate c in this assertion, firstly, we construct a word sense profile (WSP) for each sense of c. A WSP is a set of terms (words) relating to a sense, and it describes the sense in a whole. Secondly, we measure the relatedness between the terms in WSP with d in the same assertion based on NGD. Thirdly, we filter out the noisy terms in WSP, which would decrease the performance of ConceptNet disambiguating.

As a result, we calculate the score of the WSP for each sense, and choose the sense with the lowest WSP score as appreciated one for the ambiguous concept in the assertion.
Therefore, for each ambiguous concept of every ConceptNet assertion, we can assign the appropriate sense to it according to the WSP scores; and the resulted ConceptNet can be used to extend WordNet.

To compute the distance between terms, Small NGD value indicates close relatedness, while large value suggests the opposite. Given a term pair \((x, y)\), the normalized Google distance between \(x\) and \(y\), \(NGD(x, y)\), can be obtained as follows:

\[
NGD(x, y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x, y)}{\log N - \min\{\log f(x), \log f(y)\}},
\]

where \(f(x)\) is the number of Google hits for the term \(x\), \(f(y)\) is the number of Google hits for the term \(y\), \(f(x, y)\) is the number of Google hits for both terms \(x\) and \(y\), and \(N\) is the number of web pages indexed by Google. The smaller the NGD score, the more related the two terms are. According to the definition, it is desirable to use NGD to measure the relatedness between any term in the WSP and \(d\).

Since it only need to consider the relatedness between \(d\) and the term in WSP(c), the NGD scores are computed in the context of ConceptNet assertion \((c, relation, d)\).

As mentioned above, there are some noisy terms which reduce the precision of disambiguation significantly when the NGD scores are used. Suppose we have an assertion \(\langle c, relation, d\rangle\), where concept \(c\) is ambiguous and has two senses \(c_1\) and \(c_2\). Corresponding to the assertion, the correct sense of \(c\) is \(c_1\), and a wrong sense of \(c\) is \(c_2\).

For the noise from WSP\((c_2)\), which has close relation to \(d\) whereas occur in the incorrect WSP, it is closely related to \(d\) so that its NGD score is usually low, and thus lower the score of the WSD of the wrong sense. This may lead to selecting the wrong sense as the appreciated one for ambiguous concept. These noisy terms are hardly to filter out for most of them are related to the specific concept sense, which keeps unknown before the calculation of WSP scores. Furthermore, there is also the dilemma due to the filtering since this kind of noise is more likely to be retained for it has lower NGD score.

In order to alleviate this problem, we propose an ensemble approach in disambiguation. Basically, the proposed approach integrates both experts knowledge in WordNet and web knowledge in NGD.

3. Ensemble Semantic Measure of WordNet

To enrich semantic knowledge for noise filtering, we introduce the semantic similarity measures of WordNet in the ConceptNet disambiguation. There are mainly four kinds of semantic similarity measures of WordNet we has used, including Leacock & Chodorow [8], Jiang & Conrath [9], Lin [10], and Adopted Lesk [11]. We integrate these semantic similarity measures with NGD measure to improve the performance of ConceptNet disambiguation.

Let \(C_i\) represents a concept in ConceptNet, and a sense \(S_c\) of \(C_i\) represents a synset defined in WordNet, \(C'\) is another concept whose semantic similarity with \(C_i\) is to be calculated.

- Leacock & Chodorow (Lch)
  This measure develops a measure based on \(d(S_c, S_{c'})\), which is the distance of the shortest path between the two synsets \(S_c\) and \(S_{c'}\).

\[
Score_{\text{Lch}}(S_c, S_{c'}) = -\log \frac{d(S_c, S_{c'})}{2D}
\]

- Jiang & Conrath (Jcn)
  The measure takes into account the information content of the two senses, as well as that of their most specific ancestor in the taxonomy.

\[
Score_{\text{Jcn}}(S_c, S_{c'}) = \left( \frac{S_{c}}{S_{c'}} \right)^2 - S_c - S_{c'}
\]

- Lin
  Lin’s measure is based on similarity arbitrary objects.

\[
Score_{\text{Lin}}(S_c, S_{c'}) = \frac{2}{\log(p(S_c/S_{c'})) + \log(p(S_{c'}/S_c))}
\]

- Adopted Lesk
  This measure works by finding overlay in the glosses of two synsets. The relatedness score is the sum of the squares of the overlap lengths.

Let \(Score(M_i, S_j)\) is the score which a method \(M_i\) gives to concept sense \(S_j\). The predominant sense calculated by \(M_i\) for a concept \(C\) is then determined by

\[
PS(M_i, C) = \arg \max_{S \in \text{Senses}(C)} Score(M_i, S)
\]

Each ensemble component has one vote for the predominated sense, and the sense with the most votes is chosen. The scoring function for the voting ensemble is defined as:

\[
\text{Score}_{\text{voting}}([M_i]_1^k, s) = \sum_{i=1}^k \text{eq}[s, PS(M_i, C)]
\]

where \(\text{eq}[s, PS(M_i, C)] = \begin{cases} 1 & \text{if } s = PS(M_i, C) \\ 0 & \text{otherwise} \end{cases} \)

In our ensemble approach, NGD measure is also included.

4. Experiments

To prove the effectiveness of the proposed method, several experiments on WSD were conducted. Semeval-2007
coarse-grained all-words WSD dataset was selected to perform the comparative experiments [12]. This testbed examines the traditional WSD task in a coarse-grained way, dramatically reduces the number of word senses, allowing for higher inter-annotator agreement.

4.1 Performance of Disambiguated ConceptNet on WSD

To investigate the usefulness of ConceptNet disambiguated through different semantic measure. The Chen’s WSD algorithm [13] was implemented in this experiment. The WSD algorithm is not the core of our work for our aim is to evaluate the impact of the disambiguated knowledge resource on WSD performance.

The performance of the ConceptNet which is disambiguated through different semantic similarity measures on WSD is shown in Table 1. The four semantic similarity measures based on WordNet mentioned in Sect. 3, as well as the NGD measure are all adopted in ConceptNet disambiguation, respectively. The ensemble method is also evaluated. As expected, the ensemble method, which is represented by vote, gets the highest Recall and F1 compared to other individual measures. Lesk measure gets the highest Precision that reaches 89.1%. Jcn measure performances better than NGD and gets higher Precision and Recall. Lin and Lch perform worse than NGD, whereas Lin is not significantly different from NGD. This experiment results show that the ConceptNet disambiguated through the ensemble method can perform better in the WSD task than only using NGD or other individual measures in WordNet.

We also examined the relative contribution of each component to the overall performance.

Table 2 displays the drop in performance eliminating any particular component from the ensemble method. The measure that contributes the most to the ensemble is NGD. Interestingly Jcn and Lesk yield similar improvement in WSD F1 (2.0 and 2.1, respectively), when added to ensemble.

A close examination of the performance for the individual methods in the correct sense detection task shows that the F1 of all the methods is within a range of 7.2%. The actual words for which each algorithm gives the correct sense are very different. Table 3 shows the degree of overlap in assigning the appropriate correct sense among five measures. As can be seen, the largest amount of overlap is between Jcn and Lin, and this corresponds most of the words they correctly label. This means that each of these two methods gets only about twenty words right which the other labels incorrectly. The high overlap among Lch, Lin and Jcn is come from the IS-A hierarchy in the WordNet they all used in their score. However, both NGD and Lesk have low overlaps with Lch, Lin and Jcn. For NGD, it is because NGD is a measure calculated from the web, which can obtain different knowledge from the WordNet. For Lesk, it is since Lesk measure is based on the glosses of the two synsets in WordNet, not the IS-A hierarchy network. From Table 3, we see that there is a large amount of complementation between the measures, where the successes of one make up for the failures of the others. This suggests that the errors of the individual methods are sufficiently uncorrelated, and that some advantages can be gained by combining their predictions.

4.2 Evaluation of WSD Methods

In the next experiment, we evaluated the impact of combining disambiguated ConceptNet and WordNet. In addition to the Chen’s WSD algorithm, we also adopted GM algorithm [13], to examine the performance of the different knowledge resource on WSD. There are mainly three knowledge resources in this experiment, WordNet, WordNet+ConceptNet (disambiguated with NGD measure), and WordNet + ConceptNet (disambiguated with ensemble measure). We also used the random chosen sense (Random BL) and the most frequent sense (MFS BL) as baselines.

The evaluation of GM and Chen’s on Semeval-2007 coarse-grained task is given in Table 4, where ConceptNetNGD means the disambiguated ConceptNet through NGD measure, and ConceptNetensemble means the disambiguated ConceptNet through ensemble measure.

From the results, we can see that WordNet + ConceptNetensemble produce the highest Recall (78.8) and F1 (81.7) through Chen’s algorithm. This might be because the WordNet + ConceptNetensemble can act as enriched semantic knowledge resource in the WSD tasks, which combines both experts knowledge in WordNet and common knowledge in ConceptNet. And the ensemble disambiguated method can make the ConceptNet more precise and then perform better in WSD tasks.

| Table 1 | Performance on Semeval-2007 coarse-grained all words WSD (nouns only subset, and ConceptNet only, as % of all words). |
|---------|-----------------------------------------------------------------------------------------------------------------------------------|
| Method  | P  | R  | F1   |
| NGD     | 84.3 | 45.7 | 59.3 |
| Lch     | 84.0 | 45.7 | 57.5 |
| Lin     | 84.5 | 45.4 | 59.1 |
| Jcn     | 85.6 | 47.2 | 60.8 |
| Lesk    | 89.1 | 38.3 | 53.6 |
| Vote    | 88.2 | 49.3 | 63.3 |

| Table 2 | Decrease in F1 as a result of removal of each method from the rank-based ensemble (as % of all words). |
|---------|-----------------------------------------------------------------------------------------------|
| Ensemble| F1  |
| Vote    | 63.3 |
| NGD     | 60.0 (-3.3) |
| Lch     | 62.7 (-0.5) |
| Lin     | 62.8 (-0.4) |
| Jcn     | 61.3 (-2.0) |
| Lesk    | 61.2 (-2.1) |

| Table 3 | Semantic similarity measures’ pairwise agreement in correctly disambiguating the words in the dataset (as % of all words). |
|---------|------------------------------------------------------------------------------------------------|
|         | Leh | Lin | Jcn | Lesk |
| NGD     | 36.5 | 38.3 | 40.4 | 35.3 |
| Lch     | 41.7 | 42.0 | 33.4 |     |
| Lin     |     | 44.9 | 34.3 |     |
| Jcn     |     |     | 34.4 |     |
Table 4: Performance on Semeval-2007 coarse-grained all words WSD (nouns only subset, as % of all words).

| Resource          | Method | P   | R   | F1  |
|-------------------|--------|-----|-----|-----|
| WordNet           | GM     | 86.9| 55.0| 67.4|
|                   | Chen   | 87.1| 71.0| 78.2|
| WordNet + ConceptNet ensemble | GM     | 83.7| 73.6| 78.3|
|                   | Chen   | 83.1| 77.1| 79.9|
| WordNet + ConceptNet ensemble | GM     | 84.0| 74.2| 78.8|
|                   | Chen   | 84.8| 78.8| 81.7|
| MPS BL            |        | 77.4| 77.4| 77.4|
| Random BL         |        | 63.5| 63.5| 63.5|

The WordNet + ConceptNet ensemble also improves the results of GM algorithm. Compared to the performance through the WordNet + ConceptNet ensemble, both of the Precision and Recall of the GM algorithm increase when adopting the WordNet + ConceptNet ensemble, and reach 84.0 and 74.2, respectively. And the Recall and F1 of the GM algorithm are higher through WordNet + ConceptNet ensemble than only using WordNet. Besides, the two algorithms enriched with knowledge resources all perform better than the two baselines.

4.3 Examples

To show how ConceptNet is disambiguated to improve WSD, we give some examples in this section. For example, the word “plane” is polysemous both in ConceptNet and WordNet. WordNet gives five word senses for “plane” used as noun: plane_1 (an aircraft), plane_2 (an unbounded two-dimensional shape), plane_3 (a level of existence or development), plane_4 (a power tool), and plane_5 (a carpenter’s hand tool). ConceptNet gives about one hundred semantic relations about “plane”, but in a non-disambiguated way. Suppose there is a sentence “Plane crashes into building on taxiway”. If we only use WordNet to disambiguate the word “plane” in this sentence, we can hardly find any direct semantic relations from WordNet related to “plane” in the sentence. If we used the non-disambiguated ConceptNet in this sentence directly, we will find two semantic relations: “plane CapableOf crash” and “taxiway UsedFor plane”, but the non-disambiguated semantic relations cannot help disambiguate the word “plane”. Hence, we must at first carry out disambiguation. As we have mentioned in Sect. 3, we used four semantic similarity measures in WordNet, as well as NGD measure to disambiguate the ConceptNet. For example, given a ConceptNet semantic relation “taxiway UsedFor plane”, we can calculate the similarities between the five different word senses of “plane” defined by WordNet and the word sense of “taxiway” defined by WordNet as shown in Table 5.

| Resource          | Method | Similarity |
|-------------------|--------|------------|
| plane_1           |        | 1.153      |
| plane_2           |        | 0.998      |
| plane_3           |        | 1.335      |
| plane_4           |        | 1.335      |

“plane_1” easily through the disambiguated semantic relations in ConceptNet: “plane_1 CapableOf crash_1” and “taxiway_1 UsedFor plane_1”.

5. Conclusion

In this letter, we have proposed several approaches to improve the performance of WSD using the enriched semantic knowledge. In the proposed method, we combined four WordNet semantic similarity measures and NGD measure to get a disambiguated ConceptNet. Finally, the disambiguated ConceptNet was adopted in the WSD task. A number of experiments have proven that enriched by the Web knowledge (NGD) and experts knowledge (four WordNet semantic similarity measures), the ConceptNet can enhance the results in WSD task.

In future work, research on the adoption of the enriched ConceptNet is required to further improve the quality of the NLP techniques.

Acknowledgements

The work is supported by the National Social Science Foundation of China (Grant No. 12CTQ009), National Natural Sciences Foundation of China (Grant No. 71273126), the Major Projects of National Social Science Fund of China (Grant No. No. 13&ZD174).

References

[1] B. Magnini and G. Cavagli, “Integrating subject field codes into wordnet,” Proc. second International Conference on Language Resources and Evaluation, pp.1413–1418, 2000.
[2] M. Hwang, C. Choi, and P.K. Kim, “Automatic enrichment of semantic relation network and its application to word sense disambiguation,” IEEE Trans. Knowl. Data Eng., vol.23, no.6, pp.845–858, 2011.
[3] C. Montse and R. German, “Knownet: Building a large net of knowledge from the web,” Proc. 22nd International Conference on Computational Linguistics - vol.1, Manchester, United Kingdom, 2008.
[4] S.P. Ponzetto and R. Navigli, “Knowledge-rich word sense disambiguation rivaling supervised systems,” Proc. 48th Annual Meeting of the Association for Computational Linguistics, pp.1522–1531, Uppsala, Sweden, 2010.
[5] C. Havasi and R.S.A.J. Alonso, “Conceptnet 3: A flexible, multilingual semantic network for common sense knowledge,” Proc. Recent Advances in Natural Language Processing, 2007.
[6] J. Chen and J. Liu, “Combining conceptnet and wordnet for word sense disambiguation,” Proc. 5th International Joint Conference on Natural Language Processing, pp.686–694, Chiang Mai, Thailand, Nov. 2011.
[7] R.L. Cilibrasi and P.M.B. Vitanyi, “The Google similarity distance,”
IEEE Trans. Knowl. Data Eng., vol.19, no.3, pp.370–383, 2007.

[8] C. Leacock and M. Chodorow, Combining local context and WordNet similarity for word sense identification, pp.265–283, MIT Press, Cambridge, Massachusetts, 1998.

[9] J. Jiang and D. Conrath, “Semantic similarity based on corpus statistics and lexical taxonomy,” Proc. International Conference Research on Computational Linguistics ROCLING, Taiwan, 1997.

[10] D. Lin, “An information-theoretic definition of similarity,” Proc. Fifteenth International Conference on Machine learning ICML-98, Madison, Wisconsin, 1998.

[11] S. Banerjee and T. Pedersen, “Extended gloss overlaps as a measure of semantic relatedness,” Proc. Eighteenth International Joint Conference on Artificial Intelligence, pp.805–810, 2003.

[12] R. Navigli, K.C. Litkowski, and O. Hargraves, “Semeval-2007 task 07: Coarse-grained english all-words task,” Proc. 4th International Workshop on Semantic Evaluations, pp.30–35, Prague, Czech Republic, 2007.

[13] J. Chen, J. Liu, W. Yu, and P. Wu, “Combining lexical stability and improved lexical chain for unsupervised word sense disambiguation,” Proc. 2009 Second International Symposium on Knowledge Acquisition and Modeling - vol.01, pp.430–433, 2009.