ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-1634

Transductive Learning for the Identification of Word Sense Temporal Orientation

Mohammed Hasanuzzaman [1] and **Gaël Dias** [2] and **Stéphane Ferrari** [3]

Abstract.
The ability to capture the time information conveyed in natural language is essential to many natural language processing applications such as information retrieval, question answering, automatic summarization, targeted marketing, loan repayment forecasting, and understanding economic patterns. In this paper, we propose a graph-based semi-supervised classification strategy that makes use of WordNet definitions or 'glosses', its conceptual-semantic and lexical relations to supplement WordNet entries with information on the temporality of its word senses. Intrinsic evaluation results show that the proposed approach outperforms prior semi-supervised, non-graph classification approaches to the temporality recognition of word senses, and confirm the soundness of the proposed approach.

1 Introduction

There is considerable academic and commercial interest in processing time information in text, where that information is expressed either explicitly, or implicitly, or connotatively. Recognizing such information and exploiting it for Natural Language Processing (NLP) and Information Retrieval (IR) tasks are important features th and at can significantly improve the functionality of NLP/IR applications [5, 1].

Most text applications have been relying on rule-based time taggers such as HeidelTime [9] or SUTime [2] to identify and normalize time mentions in texts. Although interesting levels of performance have been seen, their coverage is limited to the finite number of rules they implement. Such systems would certainly benefit from the existence of a temporal resource enumerating a large set of possible time variants.

However, discovering the temporal orientation of words is a challenging issue even for humans if they intend to formalize them in a knowledge-base, despite the fact that they manage temporal information very naturally and efficiently during their everyday life. There are several explanations for this difficulty: (i) temporal connotations can be conveyed via a wide range of different mechanisms including grammar, aspect, and lexical semantic knowledge [8]. These properties need to be correctly identified, interpreted, and combined to derive the appropriate temporal orientation, (ii) another challenge arises from the fact that time can be expressed in countless manners and is not always expressed explicitly, rather implicitly and require interpretations or inferences derived from world knowledge, (iii) conventional knowledge acquisition approaches are usually driven by humans, which means that they are labor-intensive, time-consuming and troublesome, (iv) the data sparsity problem is aggravated by the fact that dictionary definitions or 'glosses' are very short, typically contains few words.

Whereas most of the prior computational linguistics and text mining temporal studies have focused on temporal expressions and events, there has been a lack of work looking at the temporal orientation of word senses/synsets. In this paper, we put forward a semi-supervised graph-based classification paradigm build on an optimization theory namely the max-flow min-cut theorem [7]. In particular, we propose minimum cut in a connected graph to time-tag each synset of WordNet [6] to one of the two dimensions: *temporal* and *atemporal*. Our methodology was evaluated intrinsically and outperformed prior approaches to the temporality recognition of word senses.

2 Methodology

The s-t mincut algorithm is based on finding minimum cuts in a graph, and uses pairwise relationships among examples in order to learn from both labeled and unlabeled data. In particular, it outputs a classification corresponding to partitioning a graph in a way that minimizes the number of similar pairs of examples that are given different labels.

The formulation of our mincut strategy for temporal classification of synsets involves the following steps.

- **Step I.** We define two vertices s (source) and t (sink), which correspond to the *temporal* and *atemporal* categories, respectively. Vertices s and t are *classification vertices*, and all other vertices (labeled and unlabeled) are *example vertices*.
- **Step II.** The labeled examples are connected to the classification vertices they belong to via edges with high constant non-negative weight. The unlabeled examples are connected to the classification vertices via edges weighted with non-negative scores that indicate the degree of belonging to both the *temporal* and *atemporal* categories. Weights (i.e. individual scores) are calculated based on a supervised classifier learned from labeled examples. For the classification task, each synset from the labeled dataset is represented by its gloss encoded as a vector of word unigrams weighted by their frequency. Then, a two-class SVM classifier is built from the Weka platform.[4] and the SVM membership scores are directly mapped to edge weights.
- **Step III.** For all pairs of example vertices, for which there exists a listed semantic relation in WordNet, an edge is created. This one

[1] ADAPT Centre, School of Computing, Dublin City University, Dublin, Ireland
[2] Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
[3] Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

[4] `http://www.cs.waikato.ac.nz/ml/weka/`

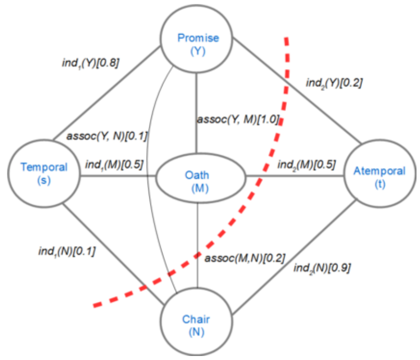

Figure 1: Example graph.

receives a non-negative score that indicates the degree of semantic relationship between both vertices (i.e. association score).

- **Step IV.** The max-flow theorem [7] is applied over the built graph to find the minimum s-t cut.[5]

2.1 Example

Figure 1 illustrates a classification problem with a set of three words {promise (Y), oath (M), chair (N)} belonging either to the *temporal* class (C_1) or the *atemporal* class (C_2) with the s-t mincut algorithm. Square brackets enclose edge weights (here probability scores). Table 1 presents all possible cuts and respective costs. The minimum cut (indicated by the dashed red line) places {promise, oath } in C_1 and {chair} in C_2.

C_1	C_2	$\sum_{x \in C_1} ind_2(x)$ + $\sum_{x \in C_2} ind_1(x)$	$\sum_{x_i \in C_1, x_k \in C_2} assoc(x_i, x_k)$	$cost(S,T)$
Y,M	N	0.2+0.5+0.1	0.1 + 0.2	1.1
none	Y,M,N	0.8+0.5+0.1	0	1.4
Y,M,N	none	0.2+0.5+0.9	0	1.6
Y	M,N	0.2+0.5+0.1	1.0+0.1	1.9
N	Y,M	0.8+0.5+0.9	0.1+0.2	2.5
M	Y,N	0.8+0.5+0.1	1.0+0.2	2.6
Y,N	M	0.2+0.5+0.9	1.0+0.2	2.8
M,N	Y	0.8+0.5+0.9	1.0+0.1	3.3

Table 1: Possible cuts for the illustrative case of Figure 1.

3 Experiments and Evaluation

We used a list that consists of 632 *temporal* synsets and an equal number of *atemporal* synsets provided by Dias et al. [3] as labeled data for our experiments.

Using our formulation in Section 2, we construct a connected graph by importing 1264 training set (632 *temporal* and 632 *atemporal* synsets), and 116394 unlabeled synsets[6]. We construct edge weights to *classification vertices, s (temporal)* and *t (atemporal)* by using the SVM classifier discussed above. WordNet relations for links between *example vertices* are weighted by non-negative constant value of 1.

In order to compare our approach to prior works, we adopted a similar evaluation strategy as proposed in Dias et al. (2014) and Hasanuzzaman et al. [4]. To assess human judgment regarding the temporal parts, inter-rater agreement with multiple raters (i.e. 3 human annotators with the 4th annotator being the classifier) was performed over a set of 398 randomly selected synsets. The free-marginal multirater kappa and the fixed-marginal multirater kappa values are reported in Table 2 and assess moderate agreement for previous versions of TempoWordNet (TWnL, TWnP and TWnH), while good agreement is obtained for the resources constructed by mincut. These agreement values provide a first and promising estimate of the improvement over the previous versions of TempoWordNet. We plan to confirm that in the future by comparing the systems to a true reference instead of observing the agreement between the systems and a multi-reference as we currently do.

Metric	TWnL	TWnP	TWnH	Mincut
Fixed-marginal κ	0.51	0.46	0.54	0.71
Free-marginal κ	0.52	0.55	0.59	0.82

Table 2: Inter-annotator agreement.

4 Conclusions

In this paper, we proposed a semi-supervised minimum cut framework to address the problem of associating word senses with their underlying temporal dimensions. Comparative evaluations are performed to measure the quality of the resource. The results confirm the soundness of the proposed approach.

As part of future work, we plan to investigate the effect of other graph construction methods, such as different weights to different WordNet relations to reflect the degree to which they are temporality preserving instead of using same for all. Another direction of future work is to fine tune the temporal part into *past, present,* and *future*. We would also like to explore the impact of the resource on more applied temporal information extraction task such as temporal relation annotation of TempEval-3 challenge.

REFERENCES

[1] Ricardo Campos, Gaël Dias, Alípio M. Jorge, and Adam Jatowt, 'Survey of temporal information retrieval and related applications', *ACM Computing Survey*, **47**(2), 15:1–15:41, (2014).

[2] Angel X Chang and Christopher D Manning, 'Sutime: A library for recognizing and normalizing time expressions.', in *LREC*, pp. 3735–3740, (2012).

[3] Gaël Harry Dias, Mohammed Hasanuzzaman, Stéphane Ferrari, and Yann Mathet, 'Tempowordnet for sentence time tagging', in *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, WWW Companion '14, pp. 833–838, Republic and Canton of Geneva, Switzerland, (2014). International World Wide Web Conferences Steering Committee.

[4] Mohammed Hasanuzzaman, Gaël Dias, Stéphane Ferrari, and Yann Mathet, 'Propagation strategies for building temporal ontologies', in *14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 6–11, (2014).

[5] Inderjeet Mani, James Pustejovsky, and Robert Gaizauskas, *The language of time: a reader*, volume 126, Oxford University Press, 2005.

[6] Georges A. Miller, 'Wordnet: a lexical database for english', *Communications of the ACM*, **38**(11), 39–41, (1995).

[7] Christos H Papadimitriou and Kenneth Steiglitz, *Combinatorial optimization: algorithms and complexity*, Courier Corporation, 1998.

[8] James Pustejovsky, 'Time and the semantic web', in *Temporal Representation and Reasoning, 2005. TIME 2005. 12th International Symposium on*, pp. 5–8. IEEE, (2005).

[9] Jannik Strötgen and Michael Gertz, 'Heideltime: High quality rule-based extraction and normalization of temporal expressions', in *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pp. 321–324, Stroudsburg, PA, USA, (2010). Association for Computational Linguistics.

[5] Max-flow algorithms show polynomial asymptotic running times and near-linear running times in practice.

[6] All synset of WordNet− labeled data