Propagation Strategies for Building Temporal Ontologies

Md. Hasanuzzaman
Normandie University
GREYC UMR 6072
Caen, France

Gaël Dias
Normandie University
GREYC UMR 6072
Caen, France

Stéphane Ferrari
Normandie University
GREYC UMR 6072
Caen, France

Yann Mathet
Normandie University
GREYC UMR 6072
Caen, France

Abstract

In this paper, we propose to build temporal ontologies from WordNet. The underlying idea is that each synset is augmented with its temporal connotation. For that purpose, temporal classifiers are iteratively learned from an initial set of time-sensitive synsets and different propagation strategies to give rise to different TempoWordNets.

1 Introduction

Temporality has recently received increased attention in Natural Language Processing (NLP) and Information Retrieval (IR). Initial works have been proposed in NLP and are exhaustively summarized in (Mani et al., 2005). More recently, the introduction of the TempEval task (Verhagen et al., 2009) in the Semantic Evaluation workshop series has clearly established the importance of time to deal with different NLP tasks. The ultimate aim of research in this area is the automatic identification of temporal expressions (timexes), events and temporal relations within a text in the TimeML format (Pustejovsky et al., 2005).

In IR, the time dimension has also received particular attention for the past few years. According to (Metzger, 2007), time is one of the key five aspects that determine a document credibility besides relevance, accuracy, objectivity and coverage. So, the value of information or its quality is intrinsically time-dependent. As a consequence, a new reasearch field called Temporal Information Retrieval (T-IR) has emerged (Alonso et al., 2011) and deals with all classical IR tasks such as crawling (Kulkarni et al., 2011), indexing (Anand et al., 2012) or ranking (Kanhabua et al., 2011) from the time viewpoint.

However, both NLP and IR evidence the lack of temporal lexical resources. For example, automatic temporal ordering of events in text is usually performed via various linguistic mechanisms

including the use of time expressions such as "before","after" or "during" that explicitly assert a temporal relation. In particular, (Derczynski and Gaizauskas, 2012) investigate the role of temporal signals in temporal relation extraction over the TimeBank annotated corpus. However, the list of such expressions is limited. From the IR viewpoint, most methodologies rely on the presence of explicit timexes and hardly bridge the gap when no explicit mention of time is available. One recent exception is proposed in (Jatowt et al., 2013) where text time-tagging is seen as a classification task, but no use of specific temporal clues is introduced or proposed.

Inspired by SentiWordNet (Esuli and Sebastiani, 2006), we propose to introduce the temporal connotation of each synset in WordNet (Miller, 1995) by iteratively learning temporal classifiers from an initial set of time-sensitive synsets and a given propagation strategy. As such, each synset is automatically time-tagged with four dimensions i.e. *atemporal*, *past*, *present* and *future*, thus giving rise to different TempoWordNets depending on the propagation strategy.

TempoWordNets are evaluated both manually and automatically. First, results show that manual annotation of time-tagged synsets is a hard task for humans. Second, automatic evaluation based on sentence temporal classification shows that the introduction of time-augmented lexical knowledge bases (TempoWordNets) allows 3.9% improvements of F_1-measure against the vector space model representation and 4.2% against the semantic vector space model obtained with the existing WordNet time subtree.

2 Related Work

A great deal of works have been proposed in temporal NLP. Most recent studies have been developed in the context of the TempEval evaluation contests, which were initiated by (Verhagen et al.,

2007). TempEval was initially divided into three challenges: (A) identifying temporal relations between events and time expressions, (B) identifying temporal relations between events and the document creation time and (C) identifying the temporal relations between contiguous pairs of matrix verbs. In TempEval-2 (Pustejovsky and Verhagen, 2009), the best performing systems were based on conditional random fields mixed with parsing methodologies (UzZaman and Allen, 2010). More recently, in TempEval-3 (UzZaman et al., 2013), new systems have been performing at high level of performance for all three tasks such as the rule-based multilingual temporal tagger Heideltime (Strötgen and Gertz, 2013). In IR, the work of (Baeza-Yates, 2005) defines the foundations of T-IR. Since, research have been tackling several topics such as query understanding (Metzler et al., 2009), temporal snippets generation (Alonso et al., 2007), temporal ranking (Kanhabua et al., 2011), temporal clustering (Alonso et al., 2009), future retrieval (Radinsky and Horvitz, 2013) or temporal image retrieval (Dias et al., 2012).

As expressed in (Strötgen and Gertz, 2013), time taggers usually contain pattern files with words and phrases, which are typically used to express temporal expressions in a given language (e.g. names of months). In fact, most temporal NLP tasks rely on a time-sensitive vocabulary. On the contrary, T-IR systems usually do not use information about time in language although they could benefit from it when facing the recurrent problem of missing explicit timexes.

WordNet is a good place to start to find time-sensitive concepts. Indeed, one can list a set of 21 temporal synsets by iteratively following the hyponym relation from the concept of time (synset # 00028270) represented by the following gloss: *the continuum of experience in which events pass from the future through the present to the past*. However, likewise the tennis problem evidenced in (Fellbaum, 1998), most temporal words are not under the concept of time. For example, concepts such as "prediction", "remember", "ancient" or "fresh" clearly have a time dimension although they are not listed under the time subtree of WordNet. Based on the intial ideas of (Moens and Steedman, 1987) on temporal ontologies and inspired by SentiWordNet (Esuli and Sebastiani, 2006), we propose to enrich all WordNet synsets with their temporal connotation.

3 TempoWordNet as SentiWordNet

In (Dias et al., 2014), we first proposed to build TempoWordNet based on the idea of (Esuli and Sebastiani, 2006) for SentiWordNet. Each synset is automatically time-tagged with four dimensions i.e. *atemporal*, *past*, *present* and *future* by performing a two-step process.

A first temporal classifier is built based on a set of manually selected seed synsets and their corresponding glosses tagged as *past*, *present* and *future*. This process is then iterated based on the repetitive lexico-semantic expansion[1] of the initial seeds lists until cross-validation accuracy drops. This first step results in a three-class temporal classifier and an expanded list of temporal synset candidates.

A second temporal classifier is then learned to time-tag synsets as *atemporal* or *temporal*. This process is obtained by taking the final list of expanded seed synsets from the previous learning problem and randomly choosing a balanced number atemporal synsets. A 10-fold cross-validation is then used to learn the model.

TempoWordNet is finally obtained by (1) classifying all WordNet synsets as *atemporal* or *temporal* with the second classifier and (2) the resulting temporal synsets are tagged as *past*, *present* and *future* by the first classifier. This step is detailed in (Dias et al., 2014) and all materials can be found at http://tempowordnet.greyc.fr.

4 Diversified Expansion Strategies

The initial strategy proposed in the previous section evidences a clear lack. As the expansion process is semantically driven, the temporal connotation is highly depend on the initial seeds lists and as a consequence may not spread over a wide range of concepts in WordNet. As such, we propose two different strategies of expansion: (1) the probabilistic expansion and (2) the hybrid (probabilistic combined with semantic) expansion.

Probabilistic Expansion: We first learn a *temporal* vs. *atemporal* classifier based on the initial hand-crafted set of seeds proposed in (Dias et al., 2014). In particular, the seeds defined as *past*, *present* and *future* are markers of temporality, while the list of *atemporal* synsets is the obvious counterpart. Based on this list of *tempo-*

---

[1]Only exisiting lexico-semantic links in WordNet are used to propagate the temporal connotation.

ral and *atemporal* synsets, a 10-fold cross validation process is performed to learn the *temporal* vs. *atemporal* model, which is used to time-tag the whole WordNet. The synsets (or glosses) with highest *temporal* and *atemporal* values in Word-Net are then used for the expansion process of the seeds lists. The process is iteratively performed and stops when accuracy drops.

After building the *temporal* vs. *atemporal* classifier, WordNet is divided into two subsets: *temporal* synsets and *atemporal* ones. In order to fine tune the *temporal* part of WordNet, we learn a three-class classifier (i.e. *past*, *present* and *future*) based on the initial *past*, *present* and *future* seeds lists and the probabilistic expansion exclusively[2] within the temporal part of WordNet. So, a 10-fold cross validation process is iteratively performed until accuracy drops.

The results of the probabilistic expansion are presented in Table 1 and Table 2, when the expansion is based on the maximum probability value[3].

Steps	1	2	3
Precision	87.3	**100**	100
Recall	86.7	**100**	100
F_1-measure	86.9	**100**	100

Table 1: Cross validation for *temporal* vs. *atemporal* at each iteration. Probabilistic Expansion.

Steps	1	2	3
Precision	80.0	**99.7**	99.6
Recall	80.1	**99.7**	99.6
F_1-measure	80.0	**99.7**	99.6

Table 2: Cross validation for *past*, *present* and *future* at each iteration. Probabilistic Expansion.

Note that in our experiment, Support Vector Machines (SVM) with a linear kernel[4] over the vector space model representation of the synsets (i.e. each synset is represented by its gloss encoded as a vector of unigrams weighted by their frequency) have been used to classify all the synsets of WordNet. The results show that in both cases the expansion process stops at iteration 2.

Hybrid Expansion: Choosing synsets from WordNet with highest probability assigned by a classifier learned on the glosses of initial seeds lists can lead to the well-known semantic shift problem. So, the idea of the hybrid expansion is to control the expansion process so that the most probable time-sensitive synsets are also chosen based on their semantic distance with the expanded seed synsets at the previous iteration. The process is straightforward when compared to the probabilistic expansion.

First, a two-class (*temporal* vs. *atemporal*) text classifier is trained based on the glosses of each synsets contained in the initial seed lists to classify all the synsets of WordNet. Thereafter, WordNet synsets with highest probability are selected as candidates for expansion. From these candidates, only the ones that present the maximum semantic similarity to the previous seeds lists are chosen for expansion. Note that the semantic similarity is calculated between the candidate synset and all synsets in the previous expanded seeds lists. Once candidates for expansion have been chosen, a 10-fold cross validation process is iteratively performed until accuracy becomes steady.

Second, a three-class (*past*, *present* and *future*) classifier is learned over the *temporal* part of WordNet with the hybrid expansion process in the same exact manner as explained for the previous probabilistic expansion. Results for the expansion process are presented in the Table 3 and Table 4 for the same experimental setups as for the probabilistic expansion and using the (Leacock et al., 1998) semantic similarity measure[5].

Steps	1	2	...	25	26	27
Precision	87.3	94.1	...	96.0	**97.2**	96.6
Recall	86.7	93.2	...	95.5	**97.0**	96.3
F_1-measure	86.9	93.6	...	95.7	**97.1**	96.4

Table 3: Cross validation for *temporal* vs. *atemporal* at each iteration. Hybrid Expansion.

Steps	1	2	...	15	16	17
Precision	80.0	75.7	...	95.7	**96.4**	95.6
Recall	80.1	74.3	...	95.1	**96.0**	95.0
F_1-measure	80.0	74.9	...	95.4	**96.2**	95.3

Table 4: Cross validation for *past*, *present* and *future* at each iteration. Hybrid Expansion.

---

[2]Only temporal synsets are classified as *past*, *present* or *future* and used for the expansion process. Note that unbalanced sets can be formed.

[3]That means that all the synsets getting the highest value produced by the classifier are used to expand the initial seeds lists.

[4]We used the Weka implementation SMO with default parameters.

---

[5]Different configurations as well as different similarity metrics have been tested but these experiments are out-of-the-scope of this paper.

Representation	Uni.+SW	Uni.+SW+Wn	Uni.+SW+TWnL	Uni.+SW+TWnP	Uni.+SW+TWnH
Precision	85.8	85.6	87.8	**89.8**	89.5
Recall	85.7	85.3	87.8	**89.5**	89.4
F_1-measure	85.8	85.4	87.8	**89.6**	89.4

Table 5: Evaluation results for sentence classification with different TempoWordNets. Balanced corpus: 346 sentences for *past*, 346 sentences for *present* and 346 sentences for *future*.

Evaluation: In order to intrinsically evaluate the time-tagged WordNets (TempoWordNets), we first performed an inter-annotation process over samples of 50 automatically time-tagged WordNet synsets. In particular, three different annotators were presented with *temporal* synsets and their respective glosses, and had to decide upon their correct classification (*temporal* vs. *atemporal*). The results of the multirater agreement evaluation are presented in Table 6. In particular, we processed the free-marginal multirater kappa values (Randolph, 2005) and the fixed-marginal multirater kappa (Siegel and Castellan, 1988) as no bias is present in the data. Overall figures assess moderate agreement for the three TempoWordNets: TWnL for the lexico-semantic expansion, TWnP for the probabilistic expansion and TWnH for the hybrid expansion.

Metric	TWnL	TWnP	TWnH
Fixed-marginal κ	0.5073	0.5199	0.4197
Free-marginal κ	0.5199	0.5199	0.4399

Table 6: Inter-annotator agreement.

These results evidence the difficulty of the task for humans as they do not agree on a great deal of decisions. This is particularly due to the fact that the temporal dimension of synsets is judged upon their glosses and not directly on their inherent concept. For example, "dinosaur" can be classified as *temporal* or *atemporal* as its gloss *any of numerous extinct terrestrial reptiles of the Mesozoic era* allows both interpretations.

So, we performed a new experiment based on those examples where human annotator agreement was 100%. From this dataset, we performed an inter-annotator agreement process with four annotators (three human annotators plus the classifier). The underlying idea is to understand to what extent the built TempoWordNets comply with the "easy" cases. Results are illustrated in Table 7 and clearly show the enhanced intrinsic quality of the hybrid expansion strategy with an almost adequate agreement for the free-marginal κ.

Metric	TWnL	TWnP	TWnH
Fixed-marginal κ	0.4133	0.4767	0.5655
Free-marginal κ	0.4242	0.5161	0.6896

Table 7: Inter-annotation for "easy" cases.

5 Sentence Temporal Classification

In order to evaluate TempoWordNets, we proposed to test their capability to enhance the external task of sentence temporal classification. For that purpose, we used the corpus developed by (Dias et al., 2014), which contains 1455 sentences distributed as follows: 724 for *past*, 385 for *present* and 346 for *future*. Different sentence representations have been used. First, we proposed to represent each sentence with the classical vector space model using the tf.idf weighting scheme for unigrams without stop-words removal (Uni.+SW). Then, we proposed a semantic vector space representation where each sentence is augmented with the synonyms of any temporal word contained in it. In particular, we proposed that the words were matched directly from the WordNet time subtree (Uni.+SW+Wn) or from TempoWordNet (Uni.+SW+TWnL, Uni.+SW+TWnP and Uni.+SW+TWnH) and weighted with tf.idf. The results of our experiments are reported in Table 5. The results evidence that the WordNet time subtree does not embody enough time-related information and the process of automatically time-tagging WordNet can improve the task of sentence temporal classification, especially with the probabilistic or the hybrid expansion.

6 Conclusion

In this paper, we proposed the first steps towards the automatic construction of temporal ontologies. In particular, we presented and evaluated different propagation strategies to time tag WordNet giving rise to different TempoWordNets. First results are promising and we deeply believe that such a resource can be important for time related applications both in NLP and IR. All resources can be found at http://tempowordnet.greyc.fr.

9

References

O. Alonso, R. Baeza-Yates, and M. Gertz. 2007. Exploratory search using timelines. In *Proceedings of the ACM SIGCHI Workshop on Exploratory Search and HCI*.

O. Alonso, M. Gertz, and R. Baeza-Yates. 2009. Clustering and exploring search results using timeline constructions. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, pages 97–106. ACM.

O. Alonso, J. Strötgen, R. Baeza-Yates, and M. Gertz. 2011. Temporal information retrieval: Challenges and opportunities. In *Proceedings of the 1st International Temporal Web Analytics Workshop (TWAW)*, pages 1–8.

A. Anand, S. Bedathur, K. Berberich, and R. Schenkel. 2012. Index maintenance for time-travel text search. In *Proceedings of the 35th International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 235–244.

Ricardo Baeza-Yates. 2005. Searching the future. In *Proceedings of the ACM SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval*, pages 1–6.

L. Derczynski and R. Gaizauskas. 2012. A corpus-based study of temporal signals. *arXiv:1203.5066*.

G. Dias, J.G. Moreno, A. Jatowt, and R. Campos. 2012. Temporal web image retrieval. In *Proceedings of the 19th Edition of the International Symposium on String Processing and Information Retrieval (SPIRE)*, pages 199–204. Springer.

G. Dias, Md. Hasanuzzaman, S. Ferrari, and Y. Mathet. 2014. Tempowordnet for sentence time tagging. In *Proceedings of the 4th ACM Temporal Web Analytics Workshop (TEMPWEB)*.

A. Esuli and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC)*, pages 417–422.

C. Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.

A. Jatowt, C.-M. Au Yeung, and K. Tanaka. 2013. Estimating document focus time. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2273–2278.

N. Kanhabua, R. Blanco, and M. Matthews. 2011. Ranking related news predictions. In *Proceedings of the 34th International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 755–764.

A. Kulkarni, J. Teevan, K.M. Svore, and S. Dumais. 2011. Understanding temporal query dynamics. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 167–176.

C. Leacock, G.A. Miller, and M. Chodorow. 1998. Using corpus statistics and wordnet relations for sense identification. *Computational Linguisics*, 24(1):147–165.

I. Mani, J. Pustejovsky, and R. Gaizauskas. 2005. *The language of time: a reader*, volume 126. Oxford University Press.

M.J. Metzger. 2007. Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13):2078–2091.

D. Metzler, R. Jones, F. Peng, and R. Zhang. 2009. Improving search relevance for implicitly temporal queries. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 700–701.

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

M. Moens and M. Steedman. 1987. Temporal ontology in natural language. In *Proceedings of the 25th Annual Meeting on Association for Computational Linguistics (ACL)*, pages 1–7.

J. Pustejovsky and M. Verhagen. 2009. Semeval-2010 task 13: evaluating events, time expressions, and temporal relations (tempeval-2). In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 112–116.

J. Pustejovsky, B. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, and I. Mani. 2005. The specification language timeml. *The language of time: A reader*, pages 545–557.

K. Radinsky and E. Horvitz. 2013. Mining the web to predict future events. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 255–264.

J.J. Randolph. 2005. Free-marginal multirater kappa (multirater κfree): an alternative to fleiss' fixed-marginal multirater kappa. *Joensuu Learning and Instruction Symposium*.

N. Siegel and J.N. Castellan. 1988. *Nonparametric Statistics for the Social Sciences*. Mcgraw-hill edition.

J. Strötgen and M. Gertz. 2013. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation (LRE)*, 47(2):269–298.

N. UzZaman and J.F. Allen. 2010. Trips and trios system for tempeval-2: Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 276–283.

N. UzZaman, H. Llorens, L. Derczynski, M. Verhagen, J. Allen, and J. Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval)*.

M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, G. Katz, and J. Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 75–80.

M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, J. Moszkowicz, and J. Pustejovsky. 2009. The tempeval challenge: Identifying temporal relations in text. *Language Resources and Evaluation (LRE)*, 43(2):161–179.