

Discovering entailment relations using “*textual entailment patterns*”

Fabio Massimo Zanzotto

DISCo, University of Milano-Bicocca,
Via Bicocca degli Arcimboldi 8, Milano, Italy,
zanzotto@disco.unimib.it

Maria Teresa Pazienza, Marco Pennacchiotti

DISP, University of Rome “Tor Vergata”,
Viale del Politecnico 1, Roma, Italy,
{pennacchiotti, pazienza}@info.uniroma2.it

Abstract

In this work we investigate methods to enable the detection of a specific type of textual entailment (*strict entailment*), starting from the preliminary assumption that these relations are often clearly expressed in texts. Our method is a statistical approach based on what we call *textual entailment patterns*, prototypical sentences hiding entailment relations among two activities. We experimented the proposed method using the entailment relations of WordNet as test case and the web as corpus where to estimate the probabilities; obtained results will be shown.

1 Introduction

Textual entailment has been recently defined as a common solution for modelling language variability in different NLP tasks (Glickman and Dagan, 2004). Roughly, the problem is to recognise if a given textual expression, the *text* (t), entails another expression, the *hypothesis* (h). An example is determining whether or not “*Yahoo acquired Overture* (t) entails *Yahoo owns Overture* (h)”. More formally, the problem of determining a textual entailment between t and h is to find a possibly graded truth value for the entailment relation $t \rightarrow h$.

Since the task involves natural language expressions, textual entailment has a more difficult nature with respect to logic entailment, as it hides two different problems: *paraphrase detection* and what can

be called *strict entailment detection*. Generally, this task is faced under the simplifying assumption that the analysed text fragments represent *facts* (f_t for the ones in the text and f_h for those in the hypothesis) in an assertive or negative way. *Paraphrase detection* is then needed when the hypothesis h carries a fact f that is also in the target text t but is described with different words, e.g., *Yahoo acquired Overture* vs. *Yahoo bought Overture*. On the other hand, *strict entailment* emerges when target sentences carry different facts, $f_h \neq f_t$. The challenge here is to derive the truth value of the entailment $f_t \rightarrow f_h$. For example, a strict entailment is “*Yahoo acquired Overture* \rightarrow *Yahoo owns Overture*”. In fact, it does not depend on the possible paraphrasing between the two expressions but on an entailment of the two facts governed by *acquire* and *own*.

Whatever the form of textual entailment is, the real research challenge consists in finding a relevant number of *textual entailment prototype relations* such as “*X acquired Y* entails *X owns Y*” or “*X acquired Y* entails *X bought Y*” that can be used to recognise entailment relations. Methods for acquiring such textual entailment prototype relations are based on the assumption that specific facts are often repeated in possibly different linguistic forms. These forms may be retrieved using their *anchors*, generally nouns or noun phrases completely characterising specific facts. The retrieved text fragments are thus considered alternative expressions for the same fact. This supposed equivalence is then exploited to derive textual entailment prototype relations. For example, the specific fact *Yahoo bought Overture* is characterised by the two anchors

{*Yahoo, Overture*}, that are used to retrieve in the corpus text fragments where they co-occur, e.g. “*Yahoo purchased Overture (July 2003)*.”, “*Now that Overture is completely owned by Yahoo!...*”. These retrieved text fragments are then considered good candidate for paraphrasing *X bought Y*.

Anchor-based learning methods have been used to investigate many semantic relations ranging from very general ones as the *isa* relation in (Morin, 1999) to very specific ones as in (Ravichandran and Hovy, 2002) where paraphrases of question-answer pairs are searched in the web or as in (Szpektor et al., 2004) where a method to scan the web for searching textual entailment prototype relations is presented. These methods are mainly devoted to induce entailment pairs related to the first kind of textual entailment, that is, *paraphrasing* as their target is mainly to look for the same “fact” in different textual forms. Incidentally, these methods can come across strict entailment relations whenever specific anchors are used for both a fact f_t and a *strictly* entailed fact f_h .

In this work we will investigate specific methods to induce the second kind of textual entailment relations, that is, *strict* entailment. We will focus on entailment between verbs, due to the fact that verbs generally govern the meaning of sentences. The problem we are facing is to look for (or verify) entailment relations like $v_t \rightarrow v_h$ (where v_t is the text verb and v_h the hypothesis verb). Our approach is based on an intuition: strict entailment relations among verbs are often clearly expressed in texts. For instance the text fragment “*Player wins \$50K in Montana Cash*” hides an entailment relation between two activities, namely *play* and *win*. If someone wins, he has first of all to play, thus, $win \rightarrow play$. The idea exploits the existence of what can be called *textual entailment pattern*, a prototypical sentence hiding an entailment relation among two activities. In the abovementioned example the pattern instance *player win* subsumes the entailment relation “ $win \rightarrow play$ ”.

In the following we will firstly describe in Sec. 2 our method to recognise entailment relations between verbs that uses: (1) the prior linguistic knowledge of these *textual entailment patterns* and (2) statistical models to assess stability of the implied relations in a corpus. Then, we will experiment our method by using the WordNet entailment relations

as test cases and the web as corpus where to estimate the probabilities (Sec. 3). Finally we will draw some conclusions (Sec. 4).

2 The method

Discovering entailment relations within texts implies the understanding of two aspects: firstly, how these entailment relations are usually expressed and, secondly, when an entailment relation may be considered stable and commonly shared. Assessing the first aspect requires the investigation of which are the prototypical textual forms that describe entailment relations. We will call them *textual entailment patterns*. These patterns (analysed in Sec. 2.2) will enable the detection of *point-wise entailment assertions*, that is, candidate verb pairs that still need a further step of analysis in order to be considered true entailment expressions. In fact, some of these candidates may be not enough stable and commonly shared in the language to be considered true entailments. To better deal with this second aspect, methods for statistically analysing large corpora are needed (see later in Sec. 2.3).

The method we propose may be used in either: (1) *recognising* if entailment holds between two verbs, or, (2) *extracting* from a corpus C all the implied entailment relations. In *recognition*, given a verb pair, the related textual entailment expressions are derived as instances of the *textual entailment patterns* and, then, the statistical entailment indicators on a corpus C are computed to evaluate the stability of the relation. In *extraction*, the corpus C should be scanned to extract textual expressions that are instances of the textual entailment patterns. The resulting pairs are sorted according to the statistical entailment indicators and only the best ranked are retained as useful verb entailment pairs.

2.1 An intuition

Our method stems from an observation: verb logical subjects, as any verb role filler, have to satisfy specific preconditions as the theory of *selectional restrictions* suggests. Then, if in a given sentence a verb v has a specific logical subject x , its selectional restrictions imply that the subject has to satisfy some preconditions p , that is, $v(x) \rightarrow p(x)$. This can be read also as: if x has the property of doing the action

v this implies that x has the property p . For example, if the verb is *to eat*, the selectional restrictions of *eat* would imply, among other things, that its subject is an *animal*. If the precondition p is “having the property of doing an action a ”, the constraint may imply that the action v entails the action a , that is, $v \rightarrow a$.

As for selectional restriction acquisition, the previous observation can enable the use of corpora as enormous sources of candidate entailment relations among verbs. For example “*John McEnroe won the match...*” can contribute to the definition of the selectional restriction $win(x) \rightarrow human(x)$ (since *John McEnroe* is a *human*), as well as to the induction (or verification) of the entailment relation between *win* and *play*, since *John McEnroe* has the *property of playing*. However, as the example shows, classes relevant for acquiring selectional preferences may be more explicit than active properties useful to derive entailment relations (i.e., it is easier to derive that *John McEnroe* is a human than that he has the property of playing).

This limitation can be overcome when *agentive nouns* such as *runner* play subject roles in some sentences. Agentive nouns usually denote the “doer” or “performer” of some action a . This is exactly what is needed to make clearer the relevant property of the noun playing the logical subject role, in order to discover entailment. The action a will be the one entailed by the verb heading the sentence. For example, in “*the player wins*”, the action *play* evoked by the agentive noun *player* is entailed by *win*.

2.2 Textual entailment patterns

As observed for the *isa* relations in (Hearst, 1992) local and simple inter-sentential patterns may carry relevant semantic relations. As we saw in the previous section, this also happens for entailment relations. Our aim is thus to search for an initial set of textual patterns that describe possible linguistic forms expressing entailment relations between two verbs (v_t, v_h) . By using these patterns, actual point-wise assertions of entailment can be detected or verified in texts. We call these prototypical patterns *textual entailment patterns*.

The idea described in Sec. 2.1 can be straightforwardly applied to generate textual entailment patterns, as it often happens that verbs can undergo an agentive nominalization (hereafter called *personifi-*

cation), e.g., *play* vs. *player*. Whether or not an entailment relation between two verbs (v_t, v_h) holds according to some writer can be verified looking for sentences with expressions involving the agentive nominalization of the hypothesis verb v_h . Then, the procedure to verify if entailment between two verbs (v_t, v_h) holds in a point-wise assertion is: whenever it is possible to personify the hypothesis v_h , scan the corpus to detect the expressions where the personified hypothesis verb is the subject of a clause governed by the text verb v_t .

Given the two investigated verbs (v_t, v_h) we will refer to this first set of textual entailment patterns as *personified patterns* $\mathcal{P}_{pers}(v_t, v_h)$. This set will contain the following textual patterns:

$$\mathcal{P}_{pers}(v_t, v_h) = \left\{ \begin{array}{l} \text{“pers}(v_h)|_{number:sing} \quad v_t|_{person:third,tense:present} \text{”}, \\ \text{“pers}(v_h)|_{number:plur} \quad v_t|_{person:third,tense:present} \text{”}, \\ \text{“pers}(v_h)|_{number:sing} \quad v_t|_{tense:past} \text{”}, \\ \text{“pers}(v_h)|_{number:plur} \quad v_t|_{tense:past} \text{”} \end{array} \right\}$$

where $pers(v)$ is the noun deriving from the personification of the verb v and elements such as $l|_{f_1, \dots, f_N}$ are the tokens generated from lemmas l by applying constraints expressed via the features f_1, \dots, f_N . For example, in the case of the verbs *play* and *win*, the related set of textual entailment expressions derived from the patterns will be $\mathcal{P}_{pers}(win, play) = \{ \text{“player wins”}, \text{“players win”}, \text{“player won”}, \text{“players won”} \}$. In the experiments hereafter described, the required verbal inflections (except personification) have been obtained using the publicly available morphological tools described in (Minnen et al., 2001) whilst simple heuristics have been used to personify verbs¹.

As the statistical measures introduced in the following section are those usually used for studying co-occurrences, two more sets of expressions, $\mathcal{F}_{pers}(v)$ and $\mathcal{F}(v)$, are needed to represent the single events in the pair. These are defined as:

$$\begin{aligned} \mathcal{F}_{pers}(v) &= \{ \text{“pers}(v)|_{number:sing} \text{”}, \text{“pers}(v)|_{number:plur} \text{”} \} \\ \mathcal{F}(v) &= \{ \text{“v}|_{person:third,tense:present} \text{”}, \\ &\quad \text{“v}|_{person:third,tense:present} \text{”}, \text{“v}|_{tense:past} \text{”} \} \end{aligned}$$

¹Personification, i.e. agentive nominalization, has been obtained adding “-er” to the verb root taking into account possible special cases such as verbs ending in “-y”. A form is retained as a correct personification if it is in WordNet.

2.3 Measures to estimate the entailment strength

The above textual entailment patterns define *point-wise* entailment assertions. In fact, if pattern instances are found in texts, the only conclusion that may be drawn is that someone (the author of the text) sustains the related entailment pairs. A sentence like “**Painter draws on old techniques but creates only decorative objects.**” suggests that *painting* entails *drawing*. However, it may happen that these correctly detected entailments are accidental, that is, the detected relation is only valid for that given text. For example, the text fragment “**When a painter discovers this hidden treasure, other people are immediately struck by its beauty.**” if taken in insulation suggests that *painting* entails *discovering*, but this is questionable. Furthermore, it may also happen that patterns detect wrong cases due to ambiguous expressions like “**Painter draws inspiration from forest, field**” where the sense of the verb *draw* is not the one expected.

In order to get rid of these wrong verb pairs, an assessment of point-wise entailment assertions over a corpus is needed to understand how much the derived entailment relations are shared and commonly agreed. This validation activity can be obtained by both analysing large textual collections and applying statistical measures relevant for the task.

Before introducing the statistical entailment indicators, some definitions are necessary. Given a corpus C containing samples, we will refer to the absolute frequency of a textual expression t in the corpus C with $f_C(t)$. The definition is easily extended to a set of expressions T as follows:

$$f_C(T) = \sum_{t \in T} f_C(t)$$

Given a pair v_t and v_h we may thus define the following *entailment strength indicators* $\mathcal{S}(v_t, v_h)$, related to more general statistical measures.

The first relevance indicator, $\mathcal{S}_f(v_t, v_h)$, is related to the probability of the textual entailment pattern as it is. This probability may be represented by the frequency, as the fixed corpus C makes constant the total number of pairs:

$$\mathcal{S}_f(v_t, v_h) = \log_{10}(f_C(\mathcal{P}_{pers}(v_t, v_h)))$$

where logarithm is used to contrast the effect of the Zipf’s law. This measure is often positively used in terminology extraction (e.g., (Daille, 1994)).

Secondly, another measure $\mathcal{S}_{mi}(v_t, v_h)$ related to point-wise mutual information (Fano, 1961) may be also used. Given the possibility of estimating the probabilities through maximum-likelihood principle, the definition is straightforward:

$$\mathcal{S}_{mi}(v_t, v_h) = \log_{10} \frac{p(\mathcal{P}_{pers}(v_t, v_h))}{p(\mathcal{F}_{pers}(v_t))p(\mathcal{F}(v_h))}$$

where $p(x) = f_C(x)/f_C(\cdot)$. The aim of this measure is to indicate the relatedness between two elements composing a pair. Mutual information has been positively used in many NLP tasks such as collocation analysis (Church and Hanks, 1989), terminology extraction (Damerau, 1993), and word sense disambiguation (Brown et al., 1991).

3 Experimental Evaluation

As many other corpus linguistic approaches, our entailment detection model relies partially on some linguistic prior knowledge (the expected structure of the searched collocations, i.e., the *textual entailment patterns*) and partially on some probability distribution estimation. Only a positive combination of both these two ingredients can give good results when applying (and evaluating) the model.

The aim of the experimental evaluation is then to understand, on the one side, if the proposed *textual entailment patterns* are useful to detect entailment between verbs and, on the other, if a statistical measure is preferable with respect to the other. We will here evaluate the capability of our method to *recognise* entailment between given pairs of verbs.

We carried out the experiments using the web as the corpus C where to estimate our two textual entailment measures (\mathcal{S}_f and \mathcal{S}_{mi}) and GoogleTM as a count estimator. The findings described in (Keller and Lapata, 2003) seem to suggest that count estimations we need in the present study over *Subject-Verb* bigrams are highly correlated to corpus counts.

As test bed we used existing resources: a non trivial set of controlled verb entailment pairs is in fact contained in WordNet (Miller, 1995). There, the entailment relation is a semantic relation defined at the synset level, standing in the verb subhierarchy. Each

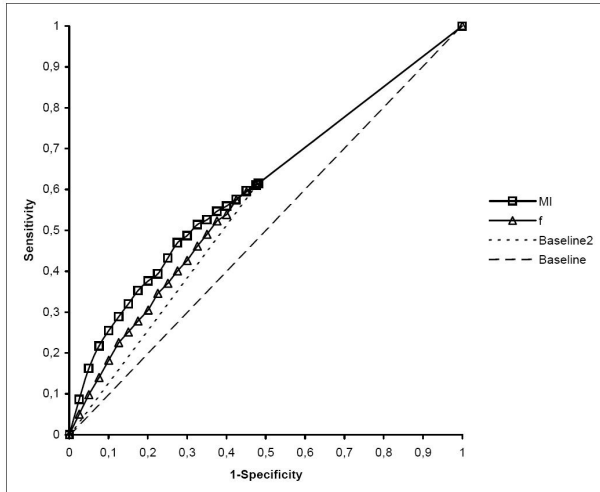


Figure 1: ROC curves

pair of synsets (S_t, S_h) is an oriented entailment relation between S_t and S_h . WordNet contains 415 entailed synsets. These entailment relations are consequently stated also at the lexical level. The pair (S_t, S_h) naturally implies that v_t entails v_h for each possible $v_t \in S_t$ and $v_h \in S_h$. It is then possible to derive from the 415 entailment synset a test set of 2,250 verb pairs. As the proposed model is applicable only when hypotheses can be personified, the number of the pairs relevant for the experiment is thus reduced to 856. This set is hereafter called the *True Set* (TS).

As the *True Set* is our starting point for the evaluation, it is not possible to produce a natural distribution in the verb pair space between entailed and not-entailed elements. Then, precision, recall, and f-measure are not applicable. The only solution is to use a ROC (Green and Swets, 1996) curve mixing *sensitivity* and *specificity*. What we then need is a *Control Set* (CS) of verb pairs that in principle are not in entailment relation. The *Control Set* has been randomly built on the basis of the *True Set*: given the set of all the hypothesis verbs H and the set of all the text verbs T of the *True Set*, control pairs are obtained randomly extracting one element from H and one element from T . A pair is considered a control pair if it is not in the *True Set*. For comparative purposes the *Control Set* has the same cardinality of the *True Set*. However, even if the intersection

between the *True Set* and the *Control Set* is empty, we are not completely sure that the *Control Set* does not contains any pair where the entailment relation holds. What we may assume is that this last set at least contains a smaller number of positive pairs.

Sensitivity, i.e. the probability of having positive answers for positive pairs, and *specificity*, i.e. the probability of having negative answers for negative pairs, are then defined as:

$$\begin{aligned} \text{Sensitivity}(t) &= p((v_h, v_t) \in TS | \mathcal{S}(v_h, v_t) > t) \\ \text{Specificity}(t) &= p((v_h, v_t) \in CS | \mathcal{S}(v_h, v_t) < t) \end{aligned}$$

where $p((v_h, v_t) \in TS | \mathcal{S}(v_h, v_t) > t)$ is the probability of a candidate pair (v_h, v_t) to belong to TS if the test is positive, i.e. the value $\mathcal{S}(v_h, v_t)$ of the entailment detection measure is greater than t , while $p((v_h, v_t) \in CS | \mathcal{S}(v_h, v_t) < t)$ is the probability of belonging to CS if the test is negative. The ROC curve (*Sensitivity* vs. $1 - \text{Specificity}$) naturally follows (see Fig. 1).

Results are encouraging as textual entailment patterns show a positive correlation with the entailment relation. Both ROC curves, the one related to the frequency indicator \mathcal{S}_f (f in figure) and the one related to the mutual information \mathcal{S}_{MI} (MI in figure), are above the *Baseline* curve. Moreover, both curves are above the second baseline (*Baseline2*) applicable when it is really possible to use the indicators. In fact, textual entailment patterns have a non-zero frequency only for 61.4% of the elements in the *True Set*. This is true also for 48.1% of the elements in the *Control Set*. The presence-absence in the corpus is then already an indicator for the entailment relation of verb pairs, but the application of the two indicators can help in deciding among elements that have a non-zero frequency in the corpus. Finally, in this case, mutual information appears to be a better indicator for the entailment relation with respect to the frequency.

4 Conclusions

We have defined a method to recognise and extract entailment relations between verb pairs based on what we call *textual entailment pattern*. In this work we defined a first kernel of *textual entailment patterns* based on subject-verb relations. Potentials of the method are still high as different kinds of textual

entailment patterns may be defined or discovered investigating relations between sentences and sub-sentences as done in (Lapata and Lascarides, 2004) for temporal relations or between near sentences as done in (Basili et al., 2003) for cause-effect relations between domain events. Some interesting and simple inter-sentential patterns are defined in (Chklovski and Pantel, 2004). Moreover, with respect to anchor-based approaches, the method we presented here offers a different point of view on the problem of acquiring textual entailment relation prototypes, as textual entailment patterns do not depend on the repetition of “similar” facts. This practically independent view may open the possibility to experiment co-training algorithms (Blum and Mitchell, 1998) also in this area. Finally, the approach proposed can be useful to define better probability estimations in probabilistic entailment detection methods such as the one described in (Glickman et al., 2005).

References

- Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. 2003. Inducing hyperlinking rules from text collections. In *Proceedings of the International Conference Recent Advances of Natural Language Processing (RANLP-2003)*, Borovets, Bulgaria.
- Avrim Blum and Tom Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *COLT: Proceedings of the Conference on Computational Learning Theory*. Morgan Kaufmann.
- Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1991. Word-sense disambiguation using statistical methods. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL)*, Berkely, CA.
- Timoty Chklovski and Patrick Pantel. 2004. VerbOCEAN: Mining the web for fine-grained semantic verb relations. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain.
- K.W. Church and P. Hanks. 1989. Word association norms, mutual information and lexicography. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada.
- Beatrice Daille. 1994. *Approche mixte pour l'extraction de terminologie: statistique lexicale et filtres linguistiques*. Ph.D. thesis, C2V, TALANA, Université Paris VII.
- F.J. Damerau. 1993. Evaluating domain-oriented multi-word terms from text. *Information Processing and Management*, 29(4):433–447.
- R.M. Fano. 1961. *Transmission of Information: a statistical theory of communications*. MIT Press, Cambridge, MA.
- Oren Glickman and Ido Dagan. 2004. Probabilistic textual entailment: Generic applied modeling of language variability. In *Proceedings of the Workshop on Learning Methods for Text Understanding and Mining*, Grenoble, France.
- Oren Glickman, Ido Dagan, and Moshe Koppel. 2005. Web based probabilistic textual entailment. In *Proceedings of the 1st Pascal Challenge Workshop*, Southampton, UK.
- D.M. Green and J.A. Swets. 1996. *Signal Detection Theory and Psychophysics*. John Wiley and Sons, New York, USA.
- Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 15th International Conference on Computational Linguistics (CoLing-92)*, Nantes, France.
- Frank Keller and Mirella Lapata. 2003. Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics*, 29(3), September.
- Mirella Lapata and Alex Lascarides. 2004. Inferring sentence-internal temporal relations. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Boston, MA.
- George A. Miller. 1995. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41, November.
- G. Minnen, J. Carroll, and D. Pearce. 2001. Applied morphological processing of english. *Natural Language Engineering*, 7(3):207–223.
- Emmanuel Morin. 1999. *Extraction de liens sémantiques entre termes à partir de corpus de textes techniques*. Ph.D. thesis, Université de Nantes, Faculté des Sciences et de Techniques.
- Deepak Ravichandran and Eduard Hovy. 2002. Learning surface text patterns for a question answering system. In *Proceedings of the 40th ACL Meeting*, Philadelphia, Pennsylvania.
- Idan Szpektor, Hristo Tanev, Ido Dagan, and Bonaventura Coppola. 2004. Scaling web-based acquisition of entailment relations. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain.