PageRanking WordNet Synsets: An Application to Opinion Mining*

Andrea Esuli and Fabrizio Sebastiani
Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche
Via Giuseppe Moruzzi, 1 – 56124 Pisa, Italy
{andrea.esuli,fabrizio.sebastiani}@isti.cnr.it

Abstract

This paper presents an application of PageRank, a random-walk model originally devised for ranking Web search results, to ranking WordNet synsets in terms of how strongly they possess a given semantic property. The semantic properties we use for exemplifying the approach are positivity and negativity, two properties of central importance in sentiment analysis. The idea derives from the observation that WordNet may be seen as a graph in which synsets are connected through the binary relation "a term belonging to synset s_k occurs in the gloss of synset s_i", and on the hypothesis that this relation may be viewed as a transmitter of such semantic properties. The data for this relation can be obtained from eXtended WordNet, a publicly available sense-disambiguated version of WordNet. We argue that this relation is structurally akin to the relation between hyperlinked Web pages, and thus lends itself to PageRank analysis. We report experimental results supporting our intuitions.

1 Introduction

Recent years have witnessed an explosion of work on *opinion mining* (aka *sentiment analysis*), the discipline that deals with the quantitative and qualitative analysis of text for the purpose of determining its opinion-related properties (ORPs). An important part of this research has been the work on the automatic determination of the ORPs of *terms*, as e.g., in determining whether an adjective tends to give a positive, a negative, or a neutral nature to the noun phrase it appears in. While many works (Esuli and Sebastiani, 2005; Hatzivassiloglou and McKeown, 1997; Kamps et al., 2004; Takamura et al., 2005; Turney and Littman, 2003) view the properties of positivity and negativity as categorical (i.e., a term is either positive or it is not), others (Andreevskaia and Bergler, 2006b; Grefenstette et al., 2006; Kim and Hovy, 2004; Subasic and Huettner, 2001) view them as *graded* (i.e., a term may be positive to a certain degree), with the underlying interpretation varying from fuzzy to probabilistic.

Some authors go a step further and attach these properties not to terms but to term *senses* (typically: WordNet synsets), on the assumption that different senses of the same term may have different opinion-related properties (Andreevskaia and Bergler, 2006a; Esuli and Sebastiani, 2006b; Ide, 2006; Wiebe and Mihalcea, 2006).

In this paper we contribute to this latter literature with a novel method for ranking the *entire* set of WordNet synsets, irrespectively of POS, according to their ORPs. Two rankings are produced, one according to positivity and one according to negativity. The two rankings are independent, i.e., it is *not* the case that one is the inverse of the other, since e.g., the least positive synsets may be negative or neutral synsets alike.

*This work was partially supported by Project ONTOTEXT "From Text to Knowledge for the Semantic Web", funded by the Provincia Autonoma di Trento under the 2004–2006 "Fondo Unico per la Ricerca" funding scheme.

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 424–431,
Prague, Czech Republic, June 2007.

The main hypothesis underlying our method is that the positivity and negativity of WordNet synsets can be determined by mining their glosses. It crucially relies on the observation that the gloss of a WordNet synset contains terms *that themselves belong to synsets*, and on the hypothesis that the glosses of positive (resp. negative) synsets will mostly contain terms belonging to positive (negative) synsets. This means that the binary relation $s_i \blacktriangleright s_k$ ("the gloss of synset s_i contains a term belonging to synset s_k"), which induces a directed graph on the set of WordNet synsets, may be thought of as a channel through which positivity and negativity flow, from the *definiendum* (the synset s_i being defined) to the *definiens* (a synset s_k that contributes to the definition of s_i by virtue of its member terms occurring in the gloss of s_i). In other words, if a synset s_i is known to be positive (negative), this can be viewed as an indication that the synsets s_k to which the terms occurring in the gloss of s_i belong, are themselves positive (negative).

We obtain the data of the $\blacktriangleright$ relation from eXtended WordNet (Harabagiu et al., 1999), an automatically sense-disambiguated version of WordNet in which every term occurrence in every gloss is linked to the synset it is deemed to belong to.

In order to compute how polarity flows in the graph of WordNet synsets we use the well known PageRank algorithm (Brin and Page, 1998). PageRank, a random-walk model for ranking Web search results which lies at the basis of the Google search engine, is probably the most important single contribution to the fields of information retrieval and Web search of the last ten years, and was originally devised in order to detect how authoritativeness flows in the Web graph and how it is conferred onto Web sites. The advantages of PageRank are its strong theoretical foundations, its fast convergence properties, and the effectiveness of its results. The reason why PageRank, among all random-walk algorithms, is particularly suited to our application will be discussed in the rest of the paper.

Note however that our method is not limited to ranking synsets *by positivity and negativity*, and could in principle be applied to the determination of other semantic properties of synsets, such as membership in a domain, since for many other properties we may hypothesize the existence of a similar "hydraulics" between synsets. We thus see positivity and negativity only as proofs-of-concept for the potential of the method.

The rest of the paper is organized as follows. Section 2 reports on related work on the ORPs of lexical items, highlighting the similarities and differences between the discussed methods and our own. In Section 3 we turn to discussing our method; in order to make the paper self-contained, we start with a brief introduction of PageRank (Section 3.1) and of the structure of eXtended WordNet (Section 3.2). Section 4 describes the structure of our experiments, while Section 5 discusses the results we have obtained, comparing them with other results from the literature. Section 6 concludes.

2 Related work

Several works have recently tackled the automated determination of term polarity. Hatzivassiloglou and McKeown (1997) determine the polarity of adjectives by mining pairs of conjoined adjectives from text, and observing that conjunctions such as and tend to conjoin adjectives of the same polarity while conjunctions such as but tend to conjoin adjectives of opposite polarity. Turney and Littman (2003) determine the polarity of generic terms by computing the pointwise mutual information (PMI) between the target term and each of a set of "seed" terms of known positivity or negativity, where the marginal and joint probabilities needed for PMI computation are equated to the fractions of documents from a given corpus that contain the terms, individually or jointly. Kamps et al. (2004) determine the polarity of adjectives by checking whether the target adjective is closer to the term good or to the term bad in the graph induced on WordNet by the synonymy relation. Kim and Hovy (2004) determine the polarity of generic terms by means of two alternative learning-free methods that use two sets of seed terms of known positivity and negativity, and are based on the frequency with which synonyms of the target term also appear in the respective seed sets. Among these works, (Turney and Littman, 2003) has proven by far the most effective, but it is also by far the most computationally intensive.

Some recent works have employed, as in the present paper, the glosses from online dictionaries for term polarity detection. Andreevskaia and Berger (2006a) extend a set of terms of known positivity/negativity by adding to them all the terms whose glosses contain them; this algorithm does not view glosses as a source for a graph of terms, and is based on a different intuition than ours. Esuli and Sebastiani (2005; 2006a) determine the ORPs of generic terms by learning, in a semi-supervised way, a binary term classifier from a set of training terms that have been given vectorial representations by indexing their WordNet glosses. The same authors later extend their work to determining the ORPs of WordNet synsets (Esuli and Sebastiani, 2006b). However, there is a substantial difference between these works and the present one, in that the former simply view the glosses as sources of textual representations for the terms/synsets, and not as inducing a graph of synsets as we instead view them here.

The work closest in spirit to the present one is probably that by Takamura et al. (2005), who determine the polarity of terms by applying intuitions from the theory of electron spins: two terms that appear one in the gloss of the other are viewed as akin to two neighbouring electrons, which tend to acquire the same "spin" (a notion viewed as akin to polarity) due to their being neighbours. This work is similar to ours since a graph between terms is generated from dictionary glosses, and since an iterative algorithm that converges to a stable state is used, but the algorithm is very different, and based on intuitions from very different walks of life.

Some recent works have tackled the attribution of opinion-related properties to word senses or synsets (Ide, 2006; Wiebe and Mihalcea, 2006)[1]; however, they do not use glosses in any significant way, and are thus very different from our method.

The interested reader may also consult (Mihalcea, 2006) for other applications of random-walk models to computational linguistics.

3 Ranking WordNet synsets by PageRank

3.1 The PageRank algorithm

Let $G = \langle N, L \rangle$ be a directed graph, with N its set of nodes and L its set of directed links; let $\mathbf{W_0}$ be the $|N| \times |N|$ *adjacency matrix* of G, i.e., the matrix such that $\mathbf{W}_0[i,j] = 1$ iff there is a link from node n_i to node n_j. We will denote by $B(i) = \{n_j \mid \mathbf{W_0}[j,i] = 1\}$ the set of the *backward neighbours* of n_i, and by $F(i) = \{n_j \mid \mathbf{W_0}[i,j] = 1\}$ the set of the *forward neighbours* of n_i. Let $\mathbf{W}$ be the *row-normalized adjacency matrix* of G, i.e., the matrix such that $\mathbf{W}[i,j] = \frac{1}{|F(i)|}$ iff $\mathbf{W_0}[i,j] = 1$ and $\mathbf{W}[i,j] = 0$ otherwise.

The input to PageRank is the row-normalized adjacency matrix $\mathbf{W}$, and its output is a vector $\mathbf{a} = \langle a_1, \ldots, a_{|N|} \rangle$, where a_i represents the "score" of node n_i. When using PageRank for search results ranking, n_i is a Web site and a_i measures its computed authoritativeness; in our application n_i is instead a synset and a_i measures the degree to which n_i has the semantic property of interest. PageRank iteratively computes vector $\mathbf{a}$ based on the formula

$$a_i^{(k)} \leftarrow \alpha \sum_{j \in B(i)} \frac{a_j^{(k-1)}}{|F(j)|} + (1-\alpha)e_i \qquad (1)$$

where $a_i^{(k)}$ denotes the value of the i-th entry of vector $\mathbf{a}$ at the k-th iteration, e_i is a constant such that $\sum_i e_{i=1}^{|N|} = 1$, and $0 \leq \alpha \leq 1$ is a control parameter. In vectorial form, Equation 1 can be written as

$$\mathbf{a}^{(k)} = \alpha \mathbf{a}^{(k-1)} \mathbf{W} + (1-\alpha)\mathbf{e} \qquad (2)$$

The underlying intuition is that a node n_i has a high score when (recursively) it has many high-scoring backward neighbours with few forward neighbours each; a node n_j thus passes its score a_j along to its forward neighbours $F(j)$, but this score is subdivided equally among the members of $F(j)$. This mechanism (that is represented by the summation in Equation 1) is then "smoothed" by the e_i constants, whose role is (see (Bianchini et al., 2005) for details) to avoid that scores flow and get trapped into so-called "rank sinks" (i.e., cliques with backward neighbours but no forward neighbours).

The computational properties of the PageRank algorithm, and how to compute it efficiently, have been widely studied; the interested reader may consult (Bianchini et al., 2005).

In the original application of PageRank for ranking Web search results the elements of $\mathbf{e}$ are usually taken to be all equal to $\frac{1}{|N|}$. However, it is possible

[1] Andreevskaia and Berger (2006a) also work on term senses, rather than terms, but they evaluate their work on terms only. This is the reason why they are listed in the preceding paragraph and not here.

to give different values to different elements in $\mathbf{e}$. In fact, the value of e_i amounts to an *internal source of score* for n_i that is constant across the iterations and independent from its backward neighbours. For instance, attributing a null e_i value to all but a few Web pages that are about a given topic can be used in order to bias the ranking of Web pages in favour of this topic (Haveliwala, 2003).

In this work we use the e_i values as internal sources of a given ORP (positivity *or* negativity), by attributing a null e_i value to all but a few "seed" synsets known to possess that ORP. PageRank will thus make the ORP flow from the seed synsets, at a rate constant throughout the iterations, into other synsets along the ▶ relation, until a stable state is reached; the final a_i values can be used to rank the synsets in terms of that ORP. Our method thus requires *two* runs of PageRank; in the first $\mathbf{e}$ has non-null scores for the positive seed synsets, while in the second the same happens for the negative ones.

3.2 eXtended WordNet

The transformation of WordNet into a graph based on the ▶ relation would of course be non-trivial, but is luckily provided by *eXtended WordNet* (Harabagiu et al., 1999), a publicly available version of WordNet in which (among other things) each term s_k occurring in a WordNet gloss (except those in example phrases) is lemmatized and mapped to the synset in which it belongs[2]. We use eXtended WordNet version 2.0-1.1, which refers to WordNet version 2.0. The eXtended WordNet resource has been automatically generated, which means that the associations between terms and synsets are likely to be sometimes incorrect, and this of course introduces noise in our method.

3.3 PageRank, (eXtended) WordNet, and ORP flow

We now discuss the application of PageRank to ranking WordNet synsets by positivity and negativity. Our algorithm consists in the following steps:

1. The graph $G = \langle N, L \rangle$ on which PageRank will be applied is generated. We define N to be the set of all WordNet synsets; in WordNet 2.0 there are 115,424 of them. We define L to contain a link from synset s_i to synset s_k iff the gloss of s_i contains *at least* a term belonging to s_k (terms occurring in the examples phrases and terms occurring after a term that expresses negation are not considered). Numbers, articles and prepositions occurring in the glosses are discarded, since they can be assumed to carry no positivity and negativity, and since they do not belong to a synset of their own. This leaves only nouns, adjectives, verbs, and adverbs.

2. The graph $G = \langle N, L \rangle$ is "pruned" by removing "self-loops", i.e., links going from a synset s_i into itself (since we assume that there is no flow of semantics from a concept unto itself). The row-normalized adjacency matrix $\mathbf{W}$ of G is derived.

3. The e_i values are loaded into the $\mathbf{e}$ vector; all synsets other than the seed synsets of renowned positivity (negativity) are given a value of 0. The α control parameter is set to a fixed value. We experiment with several different versions of the $\mathbf{e}$ vector and several different values of α; see Section 4.3 for details.

4. PageRank is executed using $\mathbf{W}$ and $\mathbf{e}$, iterating until a predefined termination condition is reached. The termination condition we use in this work consists in the fact that the cosine of the angle between $\mathbf{a}^{(k)}$ and $\mathbf{a}^{(k+1)}$ is above a predefined threshold χ (here we have set $\chi = 1 - 10^{-9}$).

5. We rank all the synsets of WordNet in descending order of their a_i score.

The process is run twice, once for positivity and once for negativity.

The last question to be answered is: "why PageRank?" Are the characteristics of PageRank more suitable to the problem of ranking synsets than other random-walk algorithms? The answer is yes, since it seems reasonable that:

1. If terms contained in synset s_k occur in the glosses of many positive synsets, and if the positivity scores of these synsets are high, then it is likely that s_k is itself positive (the same happens for negativity). This justifies the summation of Equation 1.

[2] `http://xwn.hlt.utdallas.edu/`

2. If the gloss of a positive synset that contains a term in synset s_k also contains many other terms, then this is a weaker indication that s_k is itself positive (this justifies dividing by $|F(j)|$ in Equation 1).

3. The ranking resulting from the algorithm needs to be biased in favour of a specific ORP; this justifies the presence of the $(1-\alpha)e_i$ factor in Equation 1).

The fact that PageRank is the "right" random-walk algorithm for our application is also confirmed by some experiments (not reported here for reasons of space) we have run with slightly different variants of the model (e.g., one in which we challenge intuition 2 above and thus avoid dividing by $|F(j)|$ in Equation 1). These experiments have always returned inferior results with respect to standard PageRank, thereby confirming the correctness of our intuitions.

4 Experiments

4.1 The benchmark

To evaluate the quality of the rankings produced by our experiments we have used the Micro-WNOp corpus (Cerini et al., 2007) as a benchmark[3]. Micro-WNOp consists in a set of 1,105 WordNet synsets, each of which was manually assigned a triplet of scores, one of positivity, one of negativity, one of neutrality. The evaluation was performed by five MSc students of linguistics, proficient second-language speakers of English. Micro-WNOp is representative of WordNet with respect to the different parts of speech, in the sense that it contains synsets of the different parts of speech in the same proportions as in the entire WordNet. However, it is not representative of WordNet with respect to ORPs, since this would have brought about a corpus largely composed of neutral synsets, which would be pretty useless as a benchmark for testing automatically derived lexical resources *for opinion mining*. It was thus generated by randomly selecting 100 positive + 100 negative + 100 neutral terms from the General Inquirer lexicon (see (Turney and Littman, 2003) for details) and including all the synsets that contained at least one such term, without paying attention to POS. See (Cerini et al., 2007) for more details.

The corpus is divided into three parts:

- Common: 110 synsets which all the evaluators evaluated by working together, so as to align their evaluation criteria.
- Group1: 496 synsets which were each independently evaluated by three evaluators.
- Group2: 499 synsets which were each independently evaluated by the other two evaluators.

Each of these three parts has the same balance, in terms of both parts of speech and ORPs, of Micro-WNOp as a whole. We obtain the positivity (negativity) ranking from Micro-WNOp by averaging the positivity (negativity) scores assigned by the evaluators of each group into a single score, and by sorting the synsets according to the resulting score. We use Group1 as a validation set, i.e., in order to fine-tune our method, and Group2 as a test set, i.e., in order to evaluate our method once all the parameters have been optimized on the validation set.

The result of applying PageRank to the graph G induced by the $\blacktriangleright$ relation, given a vector $\mathbf{e}$ of internal sources of positivity (negativity) score and a value for the α parameter, is a ranking of all the WordNet synsets in terms of positivity (negativity). By using different $\mathbf{e}$ vectors and different values of α we obtain different rankings, whose quality we evaluate by comparing them against the ranking obtained from Micro-WNOp.

4.2 The effectiveness measure

A ranking $\preceq$ is a partial order on a set of objects $N = \{o_1 \ldots o_{|N|}\}$. Given a pair (o_i, o_j) of objects, o_i may precede o_j ($o_i \preceq o_j$), it may follow o_i ($o_i \succeq o_j$), or it may be tied with o_j ($o_i \approx o_j$).

To evaluate the rankings produced by PageRank we have used the *p-normalized Kendall τ distance* (noted τ_p – see e.g., (Fagin et al., 2004)) between the Micro-WNOp rankings and those predicted by PageRank. A standard function for the evaluation of rankings with ties, τ_p is defined as

$$\tau_p = \frac{n_d + p \cdot n_u}{Z} \quad (3)$$

[3] `http://www.unipv.it/wnop/`

where n_d is the number of *discordant pairs*, i.e., pairs of objects ordered one way in the gold standard and the other way in the prediction; n_u is the number of pairs ordered (i.e., not tied) in the gold standard and tied in the prediction, and p is a penalization to be attributed to each such pair; and Z is a normalization factor (equal to the number of pairs that are ordered in the gold standard) whose aim is to make the range of τ_p coincide with the $[0, 1]$ interval. Note that pairs tied in the gold standard are not considered in the evaluation.

The penalization factor is set to $p = \frac{1}{2}$, which is equal to the probability that a ranking algorithm correctly orders the pair by random guessing; there is thus no advantage to be gained from either random guessing or assigning ties between objects. For a prediction which perfectly coincides with the gold standard τ_p equals 0; for a prediction which is exactly the inverse of the gold standard τ_p equals 1.

4.3 Setup

In order to produce a ranking by positivity (negativity) we need to provide an $\mathbf{e}$ vector as input to PageRank. We have experimented with several different definitions of $\mathbf{e}$, each for both positivity and negativity. For reasons of space, we only report results from the five most significant ones.

We have first tested a vector (hereafter dubbed $\mathbf{e}1$) with all values uniformly set to $\frac{1}{|N|}$. This is the $\mathbf{e}$ vector originally used in (Brin and Page, 1998) for Web page ranking, and brings about an unbiased (that is, with respect to particular properties) ranking of WordNet. Of course, it is not meant to be used for ranking by positivity or negativity; we have used it as a baseline in order to evaluate the impact of property-biased vectors.

The first sensible, albeit minimalistic, definition of $\mathbf{e}$ we have used (dubbed $\mathbf{e}2$) is that of a vector with uniform non-null e_i scores assigned to the synsets that contain the adjective good (bad), and null scores for all other synsets. A further, still fairly minimalistic definition we have used (dubbed $\mathbf{e}3$) is that of a vector with uniform non-null e_i scores assigned to the synsets that contain at least one of the seven "paradigmatic" positive (negative) adjectives used as seeds in (Turney and Littman, 2003)[4], and null scores for all other synsets.

We have also tested a more complex version of $\mathbf{e}$, with e_i scores obtained from release 1.0 of SentiWordNet (Esuli and Sebastiani, 2006b)[5]. This latter is a lexical resource in which each WordNet synset is given a positivity score, a negativity score, and a neutrality score. We produced an $\mathbf{e}$ vector (dubbed $\mathbf{e}4$) in which the score assigned to a synset is proportional to the positivity (negativity) score assigned to it by SentiWordNet, and in which all entries sum up to 1. In a similar way we also produced a further $\mathbf{e}$ vector (dubbed $\mathbf{e}5$) through the scores of a newer release of SentiWordNet (release 1.1), resulting from a slight modification of the approach that had brought about release 1.0 (Esuli and Sebastiani, 2007b).

PageRank is parametric on α, which determines the balance between the contributions of the $\mathbf{a}^{(k-1)}$ vector and the $\mathbf{e}$ vector. A value of $\alpha = 0$ makes the $\mathbf{a}^{(k)}$ vector coincide with $\mathbf{e}$, and corresponds to discarding the contribution of the random-walk algorithm. Conversely, setting $\alpha = 1$ corresponds to discarding the contribution of $\mathbf{e}$, and makes $\mathbf{a}^{(k)}$ uniquely depend on the topology of the graph; the result is an "unbiased" ranking. The desirable cases are, of course, in between. As first hinted in Section 4.1, we thus optimize the α parameter on the synsets in Group1, and then test the algorithm with the optimal value of α on the synsets in Group2. All the 101 values of α from 0.0 to 1.0 with a step of .01 have been tested in the optimization phase. Optimization is performed anew for each experiment, which means that different values of α may be eventually selected for different $\mathbf{e}$ vectors.

5 Results

The results show that the use of PageRank in combination with suitable vectors $\mathbf{e}$ almost always improves the ranking, sometimes significantly so, with respect to the original ranking embodied by the $\mathbf{e}$ vector.

For positivity, the rankings produced using PageRank and any of the vectors from $\mathbf{e}2$ to $\mathbf{e}5$ all improve on the original rankings, with a relative improvement, measured as the relative decrease in τ_p,

[4] The seven positive adjectives are good, nice, excellent, positive, fortunate, correct, superior, and the seven negative ones are bad, nasty, poor, negative, unfortunate, wrong, inferior.

[5] http://sentiwordnet.isti.cnr.it/

ranging from -4.88% ($\mathbf{e}5$) to -6.75% ($\mathbf{e}4$). These rankings are also all better than the rankings produced by using PageRank and the uniform-valued vector $\mathbf{e}1$, with a minimum relative improvement of -5.04% ($\mathbf{e}3$) and a maximum of -34.47% ($\mathbf{e}4$). This suggests that the key to good performance is indeed a combination of positivity flow *and* internal source of score.

For the negativity rankings, the performance of both SentiWordNet-based vectors is still good, producing a -4.31% ($\mathbf{e}4$) and a -3.45% ($\mathbf{e}5$) improvement with respect to the original rankings. The "minimalistic" vectors (i.e., $\mathbf{e}2$ and $\mathbf{e}3$) are not as good as their positive counterparts. The reason seems to be that the generation of a ranking by negativity seems a somehow harder task than the generation of a ranking by positivity; this is also shown by the results obtained with the uniform-valued vector $\mathbf{e}1$, in which the application of PageRank improves with respect to $\mathbf{e}1$ for positivity but deteriorates for negativity. However, against the baseline constituted by the results obtained with the uniform-valued vector $\mathbf{e}1$ for negativity, our rankings show a relevant improvement, ranging from -8.56% ($\mathbf{e}2$) to -48.27% ($\mathbf{e}4$).

Our results are particularly significant for the $\mathbf{e}4$ vectors, derived by SentiWordNet 1.0, for a number of reasons. First, $\mathbf{e}4$ brings about the best value of τ_p obtained in all our experiments (.325 for positivity, .284 for negativity). Second, the relative improvement with respect to $\mathbf{e}4$ is the most marked among the various choices for $\mathbf{e}$ (6.75% for positivity, 4.31% for negativity). Third, the improvement is obtained with respect to an already high-quality resource, obtained by the same techniques that, at the term level, are still the best performers for polarity detection on the widely used General Inquirer benchmark (Esuli and Sebastiani, 2005).

Finally, observe that the fact that $\mathbf{e}4$ outperforms all other choices for $\mathbf{e}$ (and $\mathbf{e}2$ in particular) was not necessarily to be expected. In fact, SentiWordNet 1.0 was built by a semi-supervised learning method that uses vectors $\mathbf{e}2$ as its only initial training data. This paper thus shows that, starting from $\mathbf{e}2$ as the only manually annotated data, the best results are obtained neither by the semi-supervised method that generated SentiWordNet 1.0, nor by PageRank, but by the concatenation of the former with the latter.

		Positivity		Negativity	
$\mathbf{e}$	PageRank?	τ_p	Δ	τ_p	Δ
$\mathbf{e}1$	before	.500		.500	
	after	.496	(-0.81%)	.549	(9.83%)
$\mathbf{e}2$	before	.500		.500	
	after	.467	(-6.65%)	.502	(0.31%)
$\mathbf{e}3$	before	.500		.500	
	after	.471	(-5.79%)	.495	(-0.92%)
$\mathbf{e}4$	before	.349		.296	
	after	**.325**	(-6.75%)	**.284**	(-4.31%)
$\mathbf{e}5$	before	.400		.407	
	after	.380	(-4.88%)	.393	(-3.45%)

Table 1: Values of τ_p between predicted rankings and gold standard rankings (smaller is better). For each experiment the first line indicates the ranking obtained from the original $\mathbf{e}$ vector (before the application of PageRank), while the second line indicates the ranking obtained after the application of PageRank, with the relative improvement (a negative percentage indicates improvement).

6 Conclusions

We have investigated the applicability of a random-walk model to the problem of ranking synsets according to positivity and negativity. However, we conjecture that this model can be of more general use, i.e., for the determination of other properties of term senses, such as membership in a domain. This paper thus presents a proof-of-concept of the model, and the results of experiments support our intuitions.

Also, we see this work as a proof of concept for the applicability of general random-walk algorithms (and not just PageRank) to the determination of the semantic properties of synsets. In a more recent paper (Esuli and Sebastiani, 2007a) we have investigated a related random-walk model, one in which, symmetrically to the intuitions of the model presented in this paper, semantics flows from the *definiens* to the *definiendum*; a metaphor that proves no less powerful than the one we have championed in this paper.

References

Alina Andreevskaia and Sabine Bergler. 2006a. Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'06)*, pages 209–216, Trento, IT.

Alina Andreevskaia and Sabine Bergler. 2006b. Sentiment tag extraction from WordNet glosses. In *Proceedings of* *the 5th Conference on Language Resources and Evaluation (LREC'06)*, Genova, IT.

Monica Bianchini, Marco Gori, and Franco Scarselli. 2005. Inside PageRank. *ACM Transactions on Internet Technology*, 5(1):92–128.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117.

Sabrina Cerini, Valentina Compagnoni, Alice Demontis, Maicol Formentelli, and Caterina Gandini. 2007. Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. In Andrea Sansò, editor, *Language resources and linguistic theory: Typology, second language acquisition, English linguistics*. Franco Angeli Editore, Milano, IT. Forthcoming.

Andrea Esuli and Fabrizio Sebastiani. 2005. Determining the semantic orientation of terms through gloss analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM'05)*, pages 617–624, Bremen, DE.

Andrea Esuli and Fabrizio Sebastiani. 2006a. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'06)*, pages 193–200, Trento, IT.

Andrea Esuli and Fabrizio Sebastiani. 2006b. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, pages 417–422, Genova, IT.

Andrea Esuli and Fabrizio Sebastiani. 2007a. Random-walk models of term semantics: An application to opinion-related properties. Technical Report ISTI-009/2007, Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale dellle Ricerche, Pisa, IT.

Andrea Esuli and Fabrizio Sebastiani. 2007b. SENTIWORDNET: A high-coverage lexical resource for opinion mining. Technical Report 2007-TR-02, Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, IT.

Ronald Fagin, Ravi Kumar, Mohammad Mahdiany, D. Sivakumar, and Erik Veez. 2004. Comparing and aggregating rankings with ties. In *Proceedings of ACM International Conference on Principles of Database Systems (PODS'04)*, pages 47–58, Paris, FR.

Gregory Grefenstette, Yan Qu, David A. Evans, and James G. Shanahan. 2006. Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes. In James G. Shanahan, Yan Qu, and Janyce Wiebe, editors, *Computing Attitude and Affect in Text: Theories and Applications*, pages 93–107. Springer, Heidelberg, DE.

Sanda H. Harabagiu, George A. Miller, and Dan I. Moldovan. 1999. WordNet 2: A morphologically and semantically enhanced resource. In *Proceedings of the ACL SIGLEX Workshop on Standardizing Lexical Resources*, pages 1–8, College Park, US.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL'97)*, pages 174–181, Madrid, ES.

Taher H. Haveliwala. 2003. Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):784–796.

Nancy Ide. 2006. Making senses: Bootstrapping sense-tagged lists of semantically-related words. In *Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING'06)*, pages 13–27, Mexico City, MX.

Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. 2004. Using WordNet to measure semantic orientation of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, volume IV, pages 1115–1118, Lisbon, PT.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING'04)*, pages 1367–1373, Geneva, CH.

Rada Mihalcea. 2006. Random walks on text structures. In *Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING'06)*, pages 249–262, Mexico City, MX.

Pero Subasic and Alison Huettner. 2001. Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems*, 9(4):483–496.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting emotional polarity of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 133–140, Ann Arbor, US.

Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4):315–346.

Janyce Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL'06)*, pages 1065–1072, Sydney, AU.