SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining

Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani

firstname.lastname@isti.cnr.it
Istituto di Scienza e Tecnologie dell’Informazione
Consiglio Nazionale delle Ricerche
Pisa, Italy

LREC 2010,
Malta, May 17–23, 2010
SentiWordNet

SentiWordNet is an automatically generated lexical resource that assigns to each synset of WordNet a triple of sentiment-related values: positivity, negativity, objectivity.

SentiWordNet has been first presented at LREC 2006, in Genova.

The new SentiWordNet 3.0 is aligned to the new WordNet 3.0.

SentiWordNet 3.0 is based almost on the same algorithms that generated SentiWordNet 1.0 and 2.0.

**SentiWordNet**

<table>
<thead>
<tr>
<th>Word</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimable(1)</td>
<td>deserving of respect or high regard</td>
</tr>
<tr>
<td>honorable(5)</td>
<td>good(4) respectable(2) estimable(2) deserving of esteem and respect; “all respectable companies give guarantees”; “ruined the family's good name”</td>
</tr>
<tr>
<td>computable(1)</td>
<td>estimable(3) may be computed or estimated; “a calculable risk”; “computable odds”; “estimable assets”</td>
</tr>
</tbody>
</table>
Gloss classification: **SentiWordNet 1.0**

Sentiment classification of a *WordNet* synset by classifying its gloss.

<table>
<thead>
<tr>
<th>Synset</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>good#a#3</td>
<td>morally admirable</td>
</tr>
<tr>
<td>bad#a#1</td>
<td>having undesirable or negative qualities</td>
</tr>
</tbody>
</table>

A committee of three-way gloss classifiers (positive/negative/objective) is generated by using a semi-supervised learning method.

The training set for a classifier is generated iteratively, starting from a small seed set of well-known positive, negative, and objective synsets, and adding new synsets by navigating the *WordNet* relations.

Each committee member uses different parameters (i.e., number of iterations, seed set, learner), making it more or less restrictive in recognizing subjectivity.

The triple of values for a synset is determined as the normalized count of votes produced by the committee members for each class.
Gloss classification: **SentiWordNet 1.0**

The classifiers of **SentiWordNet 1.0** use a traditional **bag of words** model to represent the glosses.

- Ambiguous terms in glosses, e.g., “estimable”, negatively impact on accuracy.

1.0 → 3.0

The classifiers of **SentiWordNet 3.0** use a **bag of synsets** model to represent the glosses.

The output of this process is **SentiWordNet 3.0-semi**.
Random walk: **SentiWordNet 2.0**

Improving **SentiWordNet 1.0** by reassigning values to synsets based on the output of a PageRank random walk algorithm applied to a graph of synsets:

- synsets are the node of the graph;

- a link between a $s_i$ and $s_j$ exists iff $s_i$ appears in the gloss of $s_j$ (definiens $\rightarrow$ definiendum).
  - If a synset is described/pointed mostly by negative synsets it is likely to be negative.

- the PageRank algorithm is used to let positivity flow into the graph, starting from an initial state determined by **SentiWordNet 1.0** positivity values (that same is separately done for negativity);

- the final PageRank values for positivity and negativity determine how the positivity and negativity values have to be reassigned to synsets.
Random walk: **SentiWordNet 2.0**

**eXtendedWordNet** is the source of the (automatically) disambiguated glosses for **WordNet 2.0**.

<table>
<thead>
<tr>
<th>Synset</th>
<th>{tidy#v#1, tidy_up#v#1, ...}</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordNet gloss</td>
<td>put (things or places) in order;</td>
</tr>
<tr>
<td>eXtendedWordNet gloss</td>
<td>put#v#1 (things#n#1 or places#n#6) in order#n#15</td>
</tr>
</tbody>
</table>

**2.0 → 3.0**

The manually disambiguated glosses are a more **reliable and complete** resource than **eXtendedWordNet**.

- The currently available release of **eXtendedWordNet** does not disambiguate the glosses of adverbs.
- We put links for all the senses.

The source for the initial values of the random walk algorithm is **SentiWordNet 3.0-semi**, instead of **SentiWordNet 1.0**.
Evaluation

Micro-WN(Op) is a corpus of 1105 human annotated synsets, using the same annotation of model SentiWordNet.

**Issue:** Micro-WN(Op) is aligned to *WordNet* 2.0.

We have *automatically* mapped it to *WordNet* 3.0 (Micro-WN(Op)-3.0) by using the publicly available synset mappings (available only for nouns and verbs) and a gloss similarity-based mapping heuristic.

The various SentiWordNet versions are evaluated by comparing how they rank the synsets of Micro-WN(Op) by positivity, or negativity, with respect to the ranking determined by human annotators.

**Evaluation measure:** \(p\)-normalized Kendall \(\tau\) distance

\[
\tau_p = \frac{n_d + p \cdot n_u}{Z} \tag{1}
\]

Lower values indicate higher agreement.
## Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Rankings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positivity</td>
</tr>
<tr>
<td><strong>SentiWordNet 1.0</strong></td>
<td>.349</td>
</tr>
<tr>
<td><strong>SentiWordNet 2.0</strong></td>
<td>.292</td>
</tr>
<tr>
<td><strong>SentiWordNet 3.0-semi</strong></td>
<td>.339</td>
</tr>
<tr>
<td><strong>SentiWordNet 3.0</strong></td>
<td><strong>.281</strong></td>
</tr>
</tbody>
</table>

Table 1: $\tau_p$ values for the positivity and negativity rankings derived from SentiWordNet 1.0, 2.0, 3.0-semi, and 3.0, as measured on Micro-WN(Op) and Micro-WN(Op)-3.0.

**SentiWordNet 3.0-semi improves over SentiWordNet 1.0.**

The relative improvement of **SentiWordNet 3.0** over **SentiWordNet 1.0** is -19.48% for positivity and -21.96% for negativity.

**SentiWordNet 2.0** obtains a better result on negativity, but **SentiWordNet 3.0** results are better balanced.
Online user’s feedback

SentiWordNet is generated by an automated process, it contains errors. It is common for a paper using SentiWordNet to report some of such errors.

‘‘for the term bad there is an entry with pos=0, neg=1, obj=0 and another entry with pos = 0.625, neg = 0.125, obj = 0.25 which are completely conflictive’’ [Denecke, 2009]

Collecting user feedback, why not?

User feedback will be released as public domain.
**Conclusion**

**SentiWordNet** 3.0 and Micro-WN(Op)-3.0 are available at:

http://swn.isti.cnr.it/

**SentiWordNet** 3.0 improves over the previous **SentiWordNet** versions:

- by using a bag-of-synsets for gloss representation in the semi-supervised learning step;
- by using manually disambiguated glosses in the random walk step.

The evaluation of **SentiWordNet** 3.0 is based on a gold standard that has been automatically aligned to **WordNet** 3.0.

- Adjectives and adverbs have been mapped by using a gloss similarity heuristic.

Collection of user feedback will allow to improve **SentiWordNet** and to develop a dedicated gold standard for **WordNet** 3.0.
Thank you. Questions?