SentiWordNet

WordNet is a lexical database that groups English words into synonym sets called synsets and provides short definitions, and records semantic relations between the synonym sets.

SentiWordNet 1.0 is a lexical resource operating on top of WordNet synsets, in which each synset s is classified based on three numerical scores $Neg(s)$, $Obj(s)$ and $Pos(s)$ specifying the synset's sentiment as either negative, objective or positive.

SentiWordNet 3.0 is newer, enhanced version of **SentiWordNet 1.0**

The method used to develop **SentiWordNet 1.0** relies on training a set of ternary classifiers, that is capable of deciding whether synset is positive, negative or objective. The ternary classifiers differs in the training set and the learning device used for training, therefore producing different classification results. Opinion-based scores for a synset are drawn by normalized portion of ternary classifiers that have assigned the corresponding label to it. If all ternary classifiers assign the same label to a synset, the label will have the maximum score, otherwise each label will have a score corresponding to number of assigned classifiers.

Ternary classifiers are generated using semi-supervised method, where only a small subset $L \in Tr$ of the training data has been manually labeled. The training data $U = Tr - L$ were left unlabelled as it is up to the process to label them automatically using L as an input, where $L = L_n \cup L_o \cup L_p$ defining L_n, L_o, and L_p as sets of Negative, Objective and Positive synsets. L_n and L_p are small sets defined by manually selecting deliberate synsets for 14 characteristic positive and negative terms. L_p and L_n are then expanded in K interations into the final training set Tr_p^K and Tr_n^K.

The main differences between **SentiWordNet 1.0** and its newer version **SentiWordNet 3.0** are:

- Version 3.0 consists of annotation of the newer **WordNet 3.0** instead of the **WordNet 2.0**.
- In 3.0 the result of this semi-supervised learning algorithm is only an intermediate step of the annotation process. It is then fed to an iterative random-walk process that is run to convergence.
- The 3.0 version in both semi-supervised learning process and the random-walk process uses manually disambiguated glosses from **Princeton WordNet Gloss Corpus** which is assumed to be more accurate than the ones from **ExtendedWordNet** used in version 1.0.

The semi-supervised learning step in **SentiWordNet 3.0** is identical to the process used in **SentiWordNet 1.0**. In the random-walk step **WordNet 3.0** is viewed as a graph that is randomly, iteratively walked and values $Pos(s)$, $Neg(s)$ and $Obj(s)$ are expected to change at each iteration. The random-walk process terminates when the iteration has converged.