We aimed at constructing a reliable, sense-oriented sentiment lexicon, largely based on manual emotive annotation of WordNet.

2. To move away from handcrafted propagation rules over WordNet.
3. Automatic expansion of manual annotation on a large scale.
4. Exploiting a more complex WordNet structure.
5. More reliable evaluation using large, manually annotated part of WordNet.

Data and Methods

Current state of emotive annotation of plWordNet:
- More than 3k annotations covering more than 54k lexical units and 41k synsets.
- 22k polarity annotations different than neutral (1k of lexical units and 9k synsets).
- 1.5k synsets with different polarity across their units.
- Without neutral units, only 145 of synsets with varying polarity strength.
- Without neutral and ambiguous annotations, only 41 synsets with conflicting polarity (it’s only 3.6% of all polarized synsets).

Relation selection

We analysed the existing structure of plWordNet to select a subset of lexico-semantic relations for propagation process:
- Take a subset of relations which covers more than 95% of all relations in plWordNet.
- 13 relations, 2 directions, 2 word types and 2 levels:
  - Features: 104 bags-of-words as features.

Relation selection

We analysed the existing structure of plWordNet to select a subset of lexico-semantic relations for propagation process:
- Take a subset of relations which covers more than 95% of all relations in plWordNet.
- Our final relation set: 13 most frequent relations.

Polarity Transfer

The acquired statistics show that synsets are strongly homogeneous in terms of the units polarity, so we decided to:
- Move the annotations from unit-level to synset-level.
- Reduce from 5-degree scale to 3-degree scale.

Scale–reduction procedure

1. Assign the following weights:
   - weight 2 for strong variants of polarities,
   - weight 1 for weak variants,
   - weight 1 for neutral and ambiguous annotations.
2. Reduce the polarities of annotations found in the synset using assigned weights.
   - Example: synset with units like [strong negative, negative, strong positive, neutral]; we have value 3 for negative category, 2 for positive, 1 for neutral category.
3. If some of polarity classes have equal values, use conversion rules.

Classifer-based Polarity Propagation (CPP)

- Baseline - simple, iterative, rule-based propagation on a narrow set of relations: hypernymy, hyponymy, similarity and antonymy.
- CPP-Naive - automatic rules extraction with classifier trained on manually annotated part of plWordNet using extended set of relations.
- CPP-Sorted - CPP-Naive solution extended with on-line reordering: synsets sorted by node degree (descending order).

Results

- Propagation: near about 4k synsets annotated with sentiment polarity.
- 10-fold CV - almost 4k synsets as a seed and 4k as a test set.

Conclusions

- The proposed method performs better in almost all cases comparing to simple rule-based solution.
- Surprisingly the solution with sorting the synsets by node degree did not improve our results.
- The classifier, as it moves away from the seed, loses information about the other parts of plWordNet - solution: use another approach with re-training.

Future Work

- Concentrate on a self-training approach, with re-training after each iteration.
- Evaluate once again a sorted approach using new re-training scenario.
- Extend manual annotation for other Parts-of-Speech (verbs, adverbs).

Acknowledgements

Work co-financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education and in part by the National Centre for Research and Development, Poland, under grant no POIR.01.01.00-04/16.