Recognition of Hyponymy and Meronymy Relations in Word Embeddings for Polish

Gabriela Czachor, Maciej Piasecki, Arkadiusz Janz

G4.19 Research Group, Department of Computational Intelligence
Wrocław University of Science and Technology, Wrocław, Poland

& CLARIN-PL `clarin-pl.eu`

Politechnika
Wrocławska

Agenda

- Need, contradiction, goal
- Classification for hyponymy recognition
- Experiments
  - Corpora and training-testing dataset
  - Scheme
  - Tests
- Results
- Conclusions

Need, contradiction, goal

- Need
  - Even in a very large wordnet some relation instances can be omitted
  - Corpus-based *Measures of Semantic Relatedness* express many different lexico-semantic relations
- Contradiction
  - In several works, classifiers were applied to MSR to recognise hypernymy instances, e.g. (Fu et al., 2014)
  - (Levy et al., 2015) convincingly claim that this is not possible (sic!)
- Goal
- to check these contradictory points of the view on large corpora and comprehensive wordnet for Polish
- to expand this research with *meronymy* – a more difficult relation

Classification for hypernymy recognition

- Wordnet or thesaurus as source of hyponymy instances: $\langle x, y \rangle$
  - x – a hyponym, and y – a hypernym,
  - are lemmas belonging to two separate synsets
- $\langle x, y \rangle$ represented by $\mathbf{x} - \mathbf{y}$
  - $\mathbf{x}$, $\mathbf{y}$ – word embedding vectors
- Hypernymy projection (Fu et al., 2014): a linear projection of the vector $\mathbf{x}$, i.e. $\Phi\mathbf{x}$, on a vector $\mathbf{y}'$
- Automatically clustering $\mathbf{x} - \mathbf{y}$ into n groups by *k-means*
- Separated classifier was trained by the linear regression method for each cluster

Experiments

Corpora and training-testing dataset (1)

- Limitations of the semantic representation based on word embeddings
  1. the whole model can be biased by the particular selection of texts
     $\rightarrow$ **large corpus needed**
  2. senses of polysemous words are merged together
     $\rightarrow$ **separate experiments for monosemous and polysemous lemmas**
  3. and the representation of less frequent words and senses can be blurred by the statistical noise
     $\rightarrow$ **words with more than 1,000 occurrences**

Experiments

Corpora and training-testing dataset (2)

- Gold standard: plWordNet – a very large wordnet of Polish
- Corpus: plWordNet Corpus 10
  1. more than 4 billion words: several corpora supplemented with text acquired from the Web, only text in Polish, automated elimination of duplicates
- Vectors
  - *word2vec* (Mikolov et al., 2013), *Gensim* implementation (Řehůřek and Sojka, 2010)
  - all words with the minimal frequency $\geqslant 8$ (min_count=8)
  - vector size: 300
- Number of clusters
  - dataset divided into: *training*, *testing* and *development* in the ratio 6:2:2
  - automated optimisation on a *development* subset

Experiments

Scheme (1)

- Two types of data sets for the experiments
  1. *random* division into subsets,
  2. *lexical train/test splits* rule proposed by (Levy et al., 2015)
     - positive cases: direct hypernyms & cannot include hypernyms from the training set
     - negative cases: excluding indirect hypernyms &

$$T_x^+ = \{x \mid (x, y) \in T^+\} \tag{1}$$

$$T_y^+ = \{y \mid (x, y) \in T^+\} \tag{2}$$

$$S = (T_x^+ \times T_y^+) \setminus T^+ \tag{3}$$

where T^+ is a set of word pairs belonging to the given relation

Experiments

Tests

Hypo-Mono – hyponymy recognition, monosemous words: 6k positive pairs, 6k negative; two variants: *random* & *lexical split*; the vector size: 100.

Hypo-Poly – 20k hyponymy pairs including polysemous words; 20k negative, two variants; the vector size: 100

Hypo-Mono300 – as in Hypo-Mono but the vector size: 300, only *lexical split*

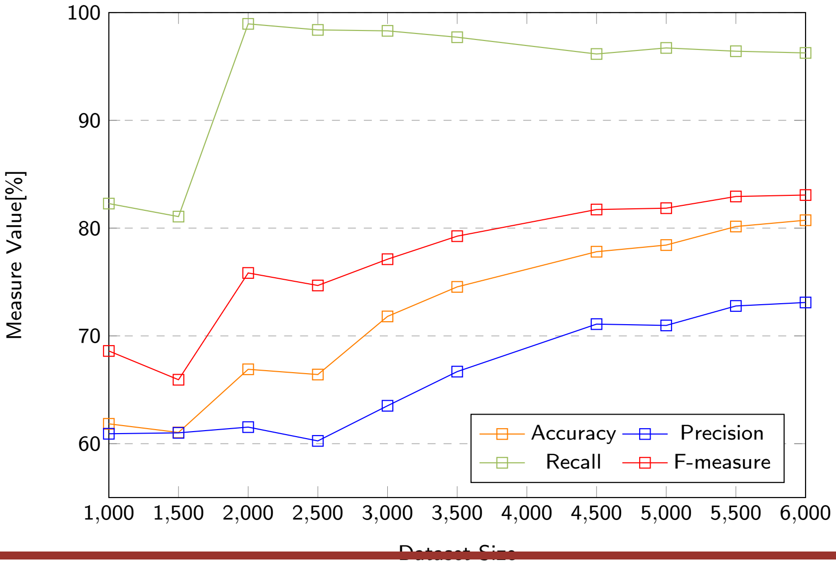
Hypo-Poly300 – as above, but 20k hyponymy pairs including polysemous words, 20k negative pairs by the *lexical split*, the vector size was: 300.

Mero-Poly – 7,900 meronymy pairs (only *part of*), 8,000 negative pairs: not connected or connected by paths longer than 3 links, the *lexical split*, the vector size: 100.

Results

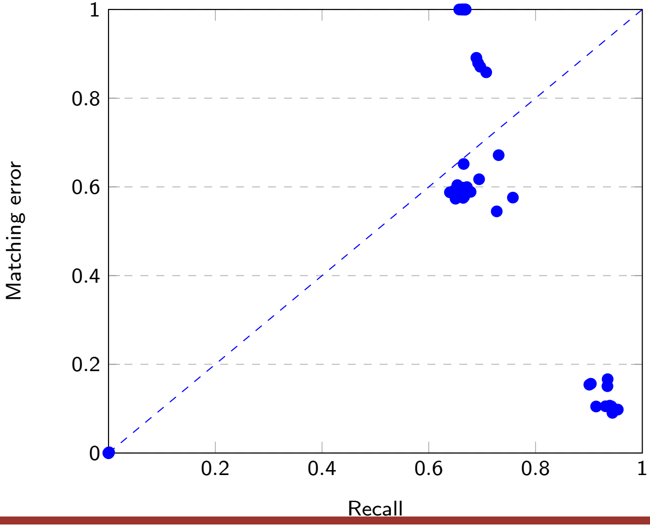
Experiment	**Acc**	**P**	**R**	**F**	**Err**	**Type**	Vec. Size
Hypo-Mono	85.22%	78.91%	96.27%	86.72%	27.91%	Rnd	100
std. dev.	0.64%	1.00%	0.65%	0.65%	1.92%	Rnd	100
Hypo-Mono	84.98%	78.90%	95.18%	86.27%	28.05%	Split	100
std. dev.	0.61%	1.59%	0.79%	0.91%	2.22%	Split	100
Hypo-Poly	78.94%	74.35%	88.35%	80.74%	31.63%	Rnd	100
std. dev.	0.65%	0.41%	1.70%	0.79%	1.78%	Rnd	100
Hypo-Poly	77.23%	73.83%	84.66%	78.85%	30.54%	Split	100
std. dev.	0.79%	1.40%	2.39%	1.04%	2.25%	Split	100
Hypo-Mono300	73.31%	65.16%	98.20%	78.32% –		Split	300
std. dev.	1.11%	1.82%	0.39%	1.31%	–	Split	300
Hypo-Poly300	82.54%	84.51%	94.72%	89.32%	–	Split	300
std. dev.	1.01%	1.11%	0.69%	0.73%	–	Split	300
Mero-Poly300	79.95%	74.66%	90.43%	81.77%	–	Split	100
std. dev.	1.05%	1.71%	1.38%	0.99%	–	Split	100

Results


100
90
80
70
60
Measure Value[%]
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
Dataset Size
Accuracy
Precision
Recall
F-measure

Results

Ration: matching error vs recall for the hyponymy recognition by SVM



Results

Examples of clusters

Hypernym	Hyponym	Cluster ID
wyziew ‘vapour’	spaliny ‘engine exhausts’	973
usługa ‘service’	przewóz ‘transport’	973
usługa ‘service’	fryzjerstwo ‘hairdressing’	973
usługa ‘service’	outsourcing ‘outsourcing’	973
usługa ‘service’	usługa powszechna ‘common service’	973
usługa ‘service’	usługa telekomunikacyjna ‘telecom. service’	973
usługa ‘service’	produkt bankowy ‘bank product’	973
nudziarz ‘bore’	sztywniak ‘staffed shirt’	973
dysputa ‘≈debate’	polemika ‘polemic’	1101
dostojnik ‘high official’	podsekretarz ‘undersecretary’	1101
dostojnik ‘high official’	wiceminister ‘vice-minister’	1101
dygnitarz ‘dignitary’	wiceminister ‘vice-minister’ 1	1101
oficjel ‘high-up’	wiceminister ‘vice-minister’	1101
dostojnik ‘high official’	wicepremier ‘deputy prime minister’	1101
dygnitarz ‘dignitary’	wicepremier ‘deputy prime minister’	1101
oficjel ‘high-up’	wicepremier ‘deputy prime minister’	1101
dezaprobata ‘disapproval’	wotum nieufności ‘vote of censure’	1101

Conclusions

- Good results in the recognition of hyponymy
  - contrary to the claim of (Levy et al., 2015) that are learning in fact that one of the words is a prototypical hypernym
  - the lexical split selection of negative samples caused the decrease of the results, but only $\approx 1-2$ %, not significant
  - classifiers are not deviating to prototype recognition
  - this substantial discrepancy of our findings can be also caused by the choice of different classification methods – clustering & linear regression vs SVM in (Levy et al., 2015)
  - we showed that in some settings SVM classifier can produce much worse results for this particular task
- Lower results for longer embedding vectors of 300 are surprising, but can be caused by insufficient number of training examples

Conclusions

- Manual selective analysis of clustering revealed that
  - Clusters do not represent different realisations of hyponymy, contrary to (Ruiji et al., 2014), but rather group difference vectors according to the more general lemmas
  - Pairs related to different top hypernyms, e.g. 'animal' and 'furniture' were linked together only in later stages of hierarchical clustering
  - Negative pairs first were merged together and only after this their subbranches were linked with other clusters
- Very good result for meronymy were achieved

Thank you very much for your attention!

http://clarin-pl.eu

http://nlp.pwr.edu.pl

http://plwordnet.pwr.edu.pl