UNIBA: Super Sense Tagging at EVALITA 2011

Pierpaolo Basile

basilepp@di.uniba.it
Department of Computer Science
University of Bari "A. Moro" (ITALY)

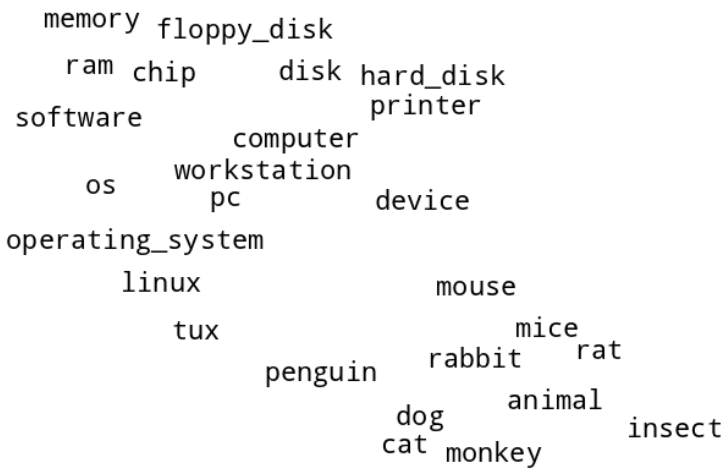
EVALITA 2011, Rome, 24-25 January 2012

Motivation

- Super Sense Tagging as sequence labelling problem [1]
- Supervised approach
  - lexical/linguistic features
  - distributional features
- **Main motivation:** tackle the data sparseness problem using word similarity in a **WordSpace**

WordSpace

- **You shall know a word by the company it keeps!** [5]
- Words are represented as points in a geometric space
- Words are related if they are close in that space



WordSpace: Random Indexing

Random Indexing [4]

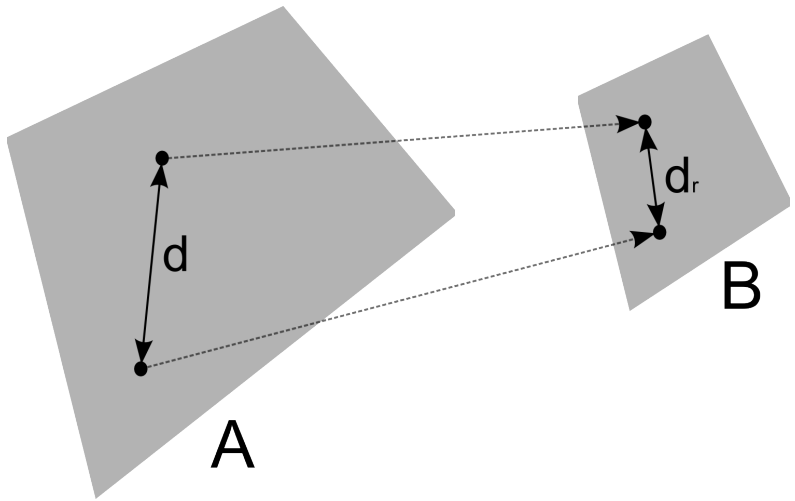
- builds WordSpace using document as context
- no matrix factorization required
- word-vectors are inferred using an incremental strategy
  1. a random vector is assigned to each context
     - sparse, high-dimensional and ternary ({-1, 0, 1})
     - a small number of randomly distributed non-zero elements
  2. random vectors are accumulated incrementally by analyzing contexts in which terms occur
     - word-vector assigned to each word is the sum of the random vectors of the contexts in which the term occur

WordSpace: Random Indexing

Formally Random Indexing is based on Random Projection [2]

$$A^{n,m} \cdot R^{m,k} = B^{n,k} \quad k < m \tag{1}$$

where $A^{n,m}$ is, for example, a term-doc matrix



After projection the distance between points is preserved: $d = c \cdot d_r$

WordSpace: context

Two WordSpaces using a different definition of context

- *Wikipedia*$_p$: a random vector is assigned to each **Wikipedia page**
- *Wikipedia*$_c$: a random vector is assigned to each **Wikipedia category**
  - categories can identify more general concepts in the same way of super-senses

Table: WordSpaces info

WordSpace	*C*	*D*
Wikipedia$_p$	1,617,449	4,000
Wikipedia$_c$	98,881	1,000

- C=number of contexts
- D=vector dimension

Methodology

- Learning method: LIBLINEAR (SVM) [3]
- Features
  1. word, lemma, PoS-tag, the first letter of the PoS-tag
  2. the super-sense assigned to the most frequent sense of the word computed according to sense frequency in MultiSemCor
  3. word starts with an upper-case character
  4. grammatical conjugation (e.g. -are, -ere and -ire for Italian verbs)
  5. distributional features: word-vector in the *WordSpace*

Evaluation

Table: Results of the evaluation

System	A	P	R	F
close	0.8696	0.7485	0.7583	0.7534
no_distr_feat	0.8822	**0.7728**	0.7818	0.7773
$Wikipedia_c$	**0.8877**	0.7719	**0.8020**	**0.7866**
$Wikipedia_p$	0.8864	0.7700	0.7998	0.7846

Final Remarks

- Main motivation: distributional features tackle data sparseness problem in SST task
  - increment in recall proves our idea
- Further work: try a different supervised approach more suitable for sequence labelling task
  - in this first attempt we are not interested in the learning method performance itself

That's all folks!

For Further Reading I


Ciaramita, M., Altun, Y.: Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. pp. 594–602. Association for Computational Linguistics (2006)


Dasgupta, S., Gupta, A.: An elementary proof of a theorem of Johnson and Lindenstrauss. Random Structures & Algorithms 22(1), 60–65 (2003)


Fan, R., Chang, K., Hsieh, C., Wang, X., Lin, C.: Liblinear: A library for large linear classification. The Journal of Machine Learning Research 9, 1871–1874 (2008)

For Further Reading II


Sahlgren, M.: An introduction to random indexing. In: Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE. vol. 5 (2005)


Sahlgren, M.: The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. Ph.D. thesis, Stockholm: Stockholm University, Faculty of Humanities, Department of Linguistics (2006)