Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname: Bc. Vladimír Macko

Study programme: Computer Science (Single degree study, master II. deg., full time form)

Field of Study: Computer Science, Informatics

Type of Thesis: Diploma Thesis

Language of Thesis: English

Secondary language: Slovak

Title: Improving LSA word weights for document classification

Annotation: NLP is currently dominated by prediction-based approaches (e.g. Word2Vec) [1]. Here we revisit a famous co-occurrence-based approach, the Latent Semantic Analysis (LSA) [2]. LSA finds the most representative rather than the most discriminative features of documents and hence may perform poorly in document classification. Inspired by the work on supervised term weights [3], the LSA can be enhanced to learn task specific weights using gradient descent.

Aim: 1. Propose a novel approach to learning task-specific word weights using gradient descent.
2. Explore the performance of this new approach using experimental approach.
3. Compare the new approach with other (commonly used) weighting schemes.

Literature: [1] Baroni, M., Dinu, G. and Kruszewski, G., 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Vol. 1, pp. 238-247).
[2] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R., 1990. Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), p.391.
[3] Deng, Z.H., Luo, K.H. and Yu, H.L., 2014. A study of supervised term weighting scheme for sentiment analysis. Expert Systems with Applications, 41(7), pp.3506-3513.

Keywords: natural language processing, document classification, LSA, gradient descent

Supervisor: RNDr. Kristína Malinovská, PhD.

Department: FMFI.KAI - Department of Applied Informatics

Head of department: prof. Ing. Igor Farkaš, Dr.

Assigned: 09.04.2018

Approved: 11.04.2018   prof. RNDr. Rastislav Kráľovič, PhD.
Guarantor of Study Programme

....
Student   Supervisor