



ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Vladimír Macko
Študijný program: informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Improving LSA word weights for document classification
Adaptácia váh slov v LSA pre zlepšenie klasifikácie dokumentov

Anotácia: V NLP momentálne dominujú prístupy založené na predikcii (napr. Word2Vec) [1]. V práci sa vraciame k známemu prístupu na báze konkurencií slov, k latentnej sémantickej analýze (LSA) [2]. Keďže LSA reprezentuje dokumenty cez najreprezentatívnejšie črty a nie cez najdiskriminatívnejšie, môže zlyhávať pri klasifikačných úlohách. Podobne ako v slovných váhach s učiteľom [3], aj LSA možno adaptovať aby sa učilo špecifické váhy slov pre konkrétne klasifikačné úlohy pomocou gradientového učenia.

Cieľ:

1. Navrhnuť nový spôsob učenia váh slov s učiteľom pre konkrétny problém.
2. Vyhodnotiť úspešnosť navrhovaného prístupu pomocou experimentov.
3. Porovnať navrhovaný prístup s inými (bežnými) váhovými schémami.

Literatúra:

[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 52nd Ann. Meeting of the ACL, 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.

Kľúčové slová: spracovanie prirodzeného jazyka, klasifikácia dokumentov, LSA, gradient descent

Vedúci: RNDr. Kristína Malinovská, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: prof. Ing. Igor Farkaš, Dr.
Dátum zadania: 09.04.2018

Dátum schválenia: 11.04.2018

prof. RNDr. Rastislav Kráľovič, PhD.
garant študijného programu