

Abstract

Název práce: Quantifying Determiners from the Distributional Semantics View

Autor: Maria Ximena Gutierrez Vasques

Katedra: Ústav formální a aplikované lingvistiky

Vedoucí diplomové práce: doc. RNDr. Markéta Lopatková, Ph.D.

Abstrakt: Distribuční sémantika představuje moderní přístup k zachycení sémantiky přirozeného jazyka. Jedním z témat, kterým zatím v rámci tohoto přístupu nebyla věnována dostatečná pozornost, je možnost automatické detekce logických relací jako vyplývání. Tato diplomová práce navazuje na práci autorů Baroni, Bernardi, Do and Shan (2012), kteří se zabývají relací vyplývání mezi kvantifikujícími výrazy. Citovaná práce využívá detekce pomocí SVN klasifikátorů natrénovaných na sémantických vektorech reprezentujících relaci vyplývání. Popisované experimenty se nezaměřovaly na nastavení parametrů SVN klasifikátoru, proto se v této práci vracíme k původním experimentům popisujícím relaci vyplývání mezi kvantifikovanými jmennými konstrukcemi, navrhuje nové konfigurace klasifikátoru a optimalizujeme nastavení parametrů. Dosaženou přesnost predikce porovnáváme s původními výsledky a ukazujeme, že SVM klasifikátor s kvadratickým polynomiálním jádrem dosahuje lepších výsledků. Analyzujeme úspěšnost výsledků a navrhuje vysvětlení, proč jsou některé kombinace nastavení úspěšné a co tato úspěšnost odhaluje o zpracovávaných datech. Závěrečné experimenty potvrzují hypotézu, podle které každá sémantická doména vyžaduje vlastní model relace vyplývání.

Klíčová slova: Distribuční sémantika, SVN (support vector machines) klasifikátor, vyplývání, kvantifikující výrazy.

Title: Quantifying Determiners from the Distributional Semantics View

Author: Maria Ximena Gutierrez Vasques

Department: Institute of Formal and Applied Linguistics

Supervisor: doc. RNDr. Markéta Lopatková, Ph.D., Charles University in Prague

Abstract: Distributional semantic models are an approach to natural language semantics that has become popular due to its capacity to successfully capture semantic relations in a relative easy way. One aspect that has been explored little is the use of distributional semantic representations to automatically detect logical relations such as entailment between words or phrases. This thesis starts from the work of Baroni, Bernardi, Do and Shan (2012) that addresses for the first time the entailment between quantifying determiners using distributional semantic models. In the mentioned work, the entailment detection is done by means of an SVM classifier trained with semantic vectors representing pairs of quantifying phrases that are in an entailment relation. The original experiments paid little attention to the parameters involved in an SVM classifier. We repeat the experiments to detect entailment between quantifier-noun constructions, proposing new configurations of the SVM classifier and performing parameter optimization. We compare the prediction accuracy and show that an SVM classifier with a quadratic polynomial kernel is more suitable for the task. We analyze why some combinations of parameters are more successful and what does this reveal about the dataset and the entailment relations. Finally, we run an experiment to reinforce the evidence, from the distributional semantics perspective, that each semantic domain has its own entailment relation.

Keywords: Distributional semantics, support vector machines, entailment, quantifiers.