

Literal and Metaphorical Senses in Compositional
Distributional Semantic Models
(Doslovné a metaforické významy v CDMS)
Počítače a sémantika 9. 5. 2017

Helena Geržová

Metafora

- přenesení významu na základě vnější podobnosti
- nedílná součást přirozeného jazyka
- konceptuální teorie (CMT)
- výzva pro počítačové zpracování

Kompozicionální přístup

- model s dvěma maticemi zvlášť pro doslovné významy a zvlášť pro metafory
- učení na anotovaných datech
- každá fráze i je reprezentována vztahem $p_i = A_{\text{LIT}(a)} n_i$, kde $A_{\text{LIT}(a)}$ pro matici doslovných významů adjektiv a
- stejně tak platí vztah $p_i = A_{\text{MET}(a)} n_i$ pro metaforické významy
- metoda byla testována na 8593 anotovaných AN frázích za užití 10 složkové křížové validace
- výsledky: F-score 0,817, recall 0,793, precision 0,842 a accuracy 0,809

Modelování metafor jako lineárních transformací

- bez anotovaných dat
- předpoklad: CM (conceptual metaphor) mohou vytvářet nekonečně mnoho LM (linguistic metaphor)
- cíl: nalezení správné reprezentace mapování CM na doslovný adjektivní význam
- možnost získat metaforické významy bez učení na anotovaných příkladech
- výsledky: F-score 0,793, precision 0,716, recall 0,819, accuracy 0,804
- na stejném setu frází byla aplikována i předchozí metoda s výsledkem 0,838 F-score.

Data

- korpus o velikosti 4,58 bil. tokenů,
- složen z anglické wikipedie, BNC, english gigaword corpus
- tokenizovaný, lemmatizovaný a označovaný
- využití NLTK toolkit
- anotace metafor, dataset volně dostupný z:
<http://bit.ly/1TQ5czN>

Závěr

- modelování metafor v rámci CDSM může zlepšit výsledný vektor reprezentace
- systematicčnost metafor může být využita k učení lineárních transformací
- oba přístupy mohou být využity také ke zlepšení automatických parafrází z AN frází

Děkuji za pozornost

- Zdroj: <https://www.aclweb.org/anthology/P/P16/P16-1018.pdf>