

Semantic Analysis and Helpfulness Prediction of Text for Online Product Reviews

Yinfei Yang, Yaowei Yan, Minghui Qiu, Forrest Sheng Bao

Motivy

- hodnocení výrobků může být rozhodujícím faktorem
- občas těžké najít kvalitní recenzi (emoce, počet recenzí)
- dosud výzkum zaměřen i na mimotextové prvky

Hypotézy:

- přívětivost textů základním rysem
- využít sémantických rysů
- testována je převoditelnost

Testovací sada

Category	Total number of reviews	Number of reviews with at least 5 votes, selected for experiments
Books	391,666	81,014 (20.7%)
Home	116,194	13,331 (11.5%)
Outdoors	52,838	6,158 (11.7%)
Electronics	135,998	15,377 (11.3%)
Overall	696,696	115,880 (16.6%)

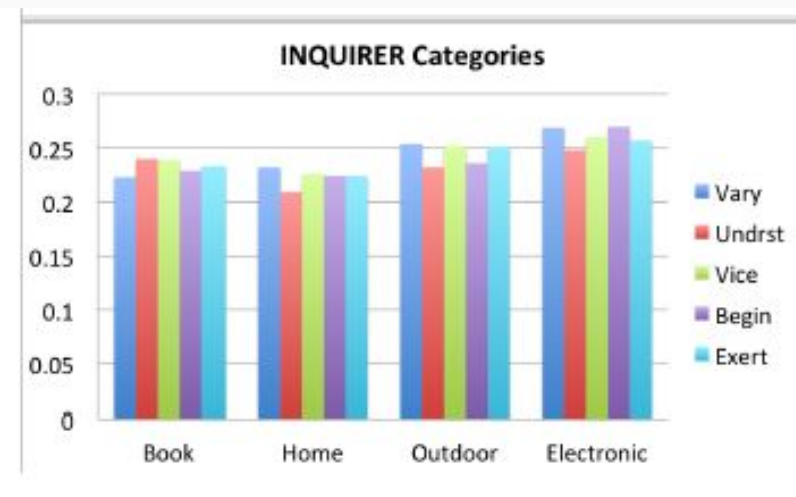
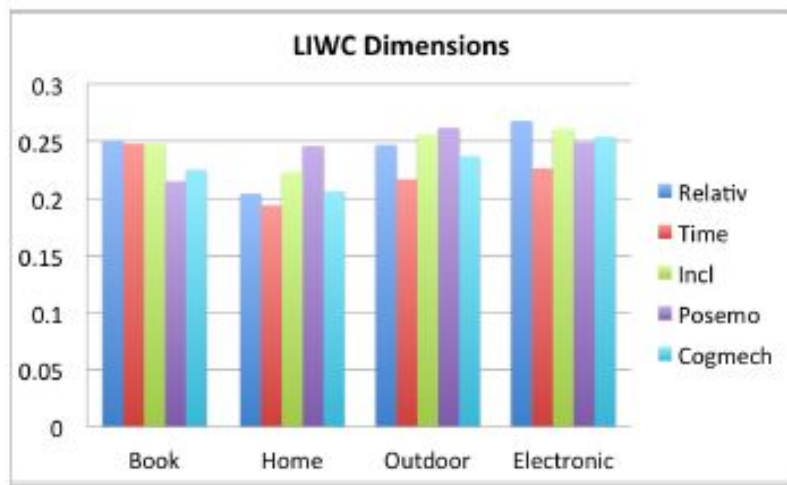
Použité funkce

- STR (Structure)
- UGR (Unigram)
- GALC (Geneva Affect Label Coder)
- LIWC (Linguistic Inquiry and Word Count)
- INQUIRER

Výsledky automatického značení

	Books	Home	Outdoors	Electronics
STR	0.500	0.280	0.333	0.351
UGR	0.507	0.467	0.458	0.471
GALC	0.239	0.216	0.255	0.274
LIWC	0.742	0.439	0.424	0.475
INQUIRER	0.720	0.487	0.455	0.498
Fusion _{Semantic}	0.744	0.490	0.467	0.527
Fusion _{All}	0.682	0.525	0.535	0.539

Co dělá recenzi užitečnou?



Závěr: užitečné recenze obsahují slova, které odrážejí více rozumovou stránku než emoční

Manuální anotace

	Books	Home	Outdoors	Electronics
STR	0.539*	0.522*	0.471*	0.635*
UGR	0.607*	0.560*	0.579*	0.626*
GALC	0.214	0.405*	0.156	0.418*
LIWC	0.524*	0.553*	0.517*	0.702*
INQUIRER	0.620*	0.662*	0.620*	0.676*
Fusion _{Semantic}	0.556*	0.680*	0.569*	0.603*
Fusion _{All}	0.610*	0.801*	0.698*	0.768*

Závěry

- Zformulovali nový problém (zaměření pouze na textovou část)
- Vstřícnost je skutečně základní vlastnost textu
- Užitečná recenze obsahuje více slov, které odrážejí spíše kritické myšlení než emocionální výrazy