

Opinion Mining – Dolovanie názorov

Peter Vojtek

Vymedzenie pojmov

- ▶ **dolovanie názorov**

- ▶ synonymá: *sentiment analysis, sentiment classification, opinion extraction*
- ▶ spája sprístupňovanie informácií (IR) a výpočtovú jazykovedu
- ▶ dôležité je zachytenie **názoru**, nie **témy**

- ▶ **názor**

- ▶ osobný, subjektívny pohľad na vec
- ▶ nie je možné objektívne ho sledovať a overiť



Príklady použitia dolovania názorov

- ▶ Aký názor vyjadruje nasledovný text?
 - ▶ „prezentácie na PeWe sú nudné a dlhé“
- ▶ Aký je vo všeobecnosti názor na súčasnú vládu?
- ▶ Ktorý zákazníci spoločnosti SONY sú nespokojní? A prečo?



Smery vývoja

- ▶ vytvorenie jazykovedných podkladov, napríklad slovníka subjektívnych výrazov
- ▶ klasifikácia textov podľa ich názoru
- ▶ dolovanie názorov z textu, určenie autora názoru



Výrazy vyjadrujúce názor

- ▶ Určenie orientácie výrazu
 - ▶ pozitívne: *dobrý, excelentný*
 - ▶ negatívne: *zlý, nesprávny*
- ▶ Určenie subjektívnosti výrazu
 - ▶ objektívne: *žltý, rovný*
 - ▶ subjektívne: *pekný*
- ▶ Určenie váhy orientácie a subjektívnosti
- ▶ Sledovanie samotného slova nestačí, dôležité je aj jeho okolie
 - ▶ „nie je príliš spoľahlivá“



Teória ohodnotení (The Appraisal Theory)

rámec popisujúci možné spôsoby vyjadrovania subjektívnych názorov

	happy	very	„very happy“	not	„not very happy“
attitude:	affect		affect		affect
orientation:	positive		positive		negative
force:	neutral	increase	high	reverse	low
focus:	neutral		neutral		neutral
polarity:	unmarked		unmarked	marked	marked



Orientácia výrazov – rozširovanie slovníka

- ▶ hypotéza: výrazy s podobnou orientáciou majú podobný popis
 - ▶ **malý** - rozmermi nevel'ký v porovnaní s inou vecou toho istého druhu
 - ▶ **drobný** - malých rozmerov, malý, nepatrný
- ▶ dve nevel'ké trénovacie množiny pozitívnych a negatívnych výrazov
- ▶ na popisoch týchto výrazov sa naučí klasifikátor, ktorý následne zatriedi ďalšie výrazy



Anotácia názorov

- ▶ MPQA korpus pozostávajúci z manuálne anotovaných viet (~500 novinových článkov, 11 000 viet)
- ▶ *“The report is full of absurdities”, Xirao-Nima said*
 - ▶ Direct subjective
 - Text anchor: said
 - Source: <writer, Xirao-Nima>
 - Intensity: high
 - Target: report
 - Attitude type: negative
 - ▶ Expressive subjective element:
 - Text anchor: full of absurdities

Celkový názor dokumentu

- ▶ názorová orientácia celého dokumentu je súčtom orientácie jeho častí
- ▶ je možné použiť niektorú štandardnú metódu učenia s učiteľom:
 - ▶ metódy: *Naive Bayes*, *SVM*
 - ▶ jednotky textu: *n-gramy*, *prídavné mená*, *POS*
- ▶ zdá sa že to funguje:
 - ▶ recenzie filmov, 90.2% accuracy [Whitelaw et al., 2005]



Zhrnutie

- ▶ nová vedná oblasť
- ▶ slovenčina – zdá sa, že zatiaľ žiadna aktivita
- ▶ čeština:
 - ▶ Smrž Pavel: Using WordNet for Opinion Mining, In: Proceedings of the Third International WordNet Conference, GWC 2006, Brno, CZ, MUNI, 2006
<http://nlpweb.kaist.ac.kr/gwc/pdf2006/42.pdf>



Zdroje

- ▶ Bing Liu: Web Data Mining. Springer, 2006.
- ▶ Andrea Esuli: Opinion Mining. Language and Intelligence Reading Group. 2006.
`medialab.di.unipi.it/web/Language+Intelligence/OpinionMining06-06.pdf`
- ▶ Whitelaw, C., Garg, N., and Argamon, S. (2005). Using appraisal taxonomies for sentiment analysis. In Proceedings of MCLC-05, the 2nd Midwest Computational Linguistic Colloquium, Columbus, US.

