



Větná polarita v češtině

Kateřina Veselovská
Žďárek – Hořovice, 27. 11. 2009

- Polarita - úvod do problematiky
- Větná polarita:
 - a) Cíl a motivace
 - b) Charakteristika
 - c) Možnosti výzkumu
- Větná polarita a vyhledávání
- Závěr

Polarita

- podoblast tzv. Opinion Miningu

klasifikace textu v rámci Opinion Miningu:

faktorem míra subjektivity – **fakta** vs. **názory**

*Český prezident je volen oběma komorami
Parlamentu vs.*

Český prezident je ve světě velmi oblíben.

- určení polarity = zhodnocení názoru vyjádřeného v textu v rámci

opozice **pozitivní** vs. **negativní**

Mám ráda rýži vs. Nesnáším brambory

- Polaritu určujeme na základě **polárních výrazů** (např. *dobrý*, *špatný* atd.)
- polárními výrazy nemusejí být vždy frekventovaná slova
- polární výrazy jsou závislé na kontextu

Př. vyjádření členské negace:

*[Ne **hodný**] byl ten chlapec, ale vypočítavý a mstivý.*

- Polarita - úvod do problematiky
- **Větná polarita:**
 - a) **Cíl a motivace**
 - b) Charakteristika
 - c) Možnosti výzkumu
- Větná polarita a vyhledávání
- Závěr

Cíl

- Rozhodnout, zda daná věta vyjadřuje pozitivní či negativní názor.
- Všechny věty, jejichž polaritu budeme určovat, musejí být vyhodnoceny v rámci Opinion Miningu jako subjektivní názory.

Motivace

Proč se zabývat polaritou **na úrovni věty**?

- přesnější rozlišení než na delších úsecích textu –
užitečné v oblastech NLP jako je
 - Question Answering
 - Sumarizace textu
- polární výrazy obvykle nebývají v textu rozmístěny rovnoměrně (vliv na celkové hodnocení)

TEXT

... **dobře**

.....

... **špatný**

.....

..... **skvělý** ...

.....

... **rád**

.....

... **úžasné**

.....

..... **pěkně**

.....

... **bezvadný**

.....

P OLARITA = **POZITIVNÍ**

Polarita na úrovni textu

- Určující je četnost
- Není nutná přesná interpretace každého slova
- Není nutné slovnědruhové zařazení

..... **dobře**
.....
POLARITA = POZITIVNÍ

Polarita na úrovni věty

- Nutná přesná interpretace jednotlivých výrazů
- Ve větě obsaženo méně informací
- Nutno určit slovní druh

Přínos

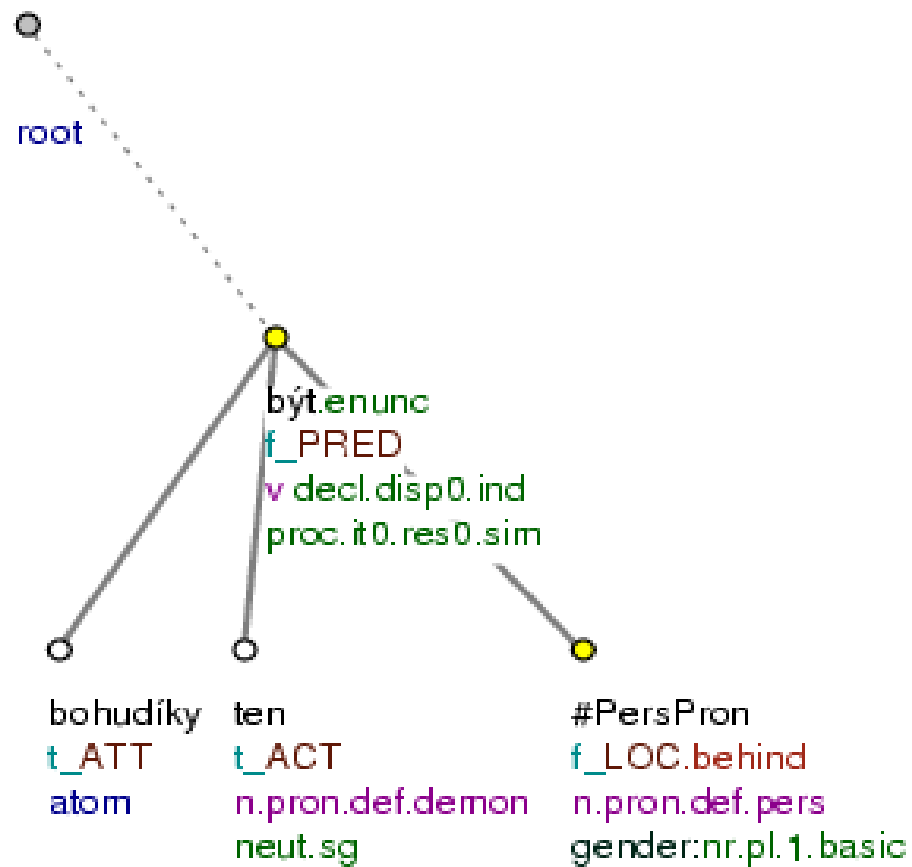
- Zkoumání jazykových rysů v souvislosti se strukturou věty
- Využití kombinace jazykových rysů + polarity ve strojovém učení (supervised machine learning)

- Polarita - úvod do problematiky
- Větná polarita:
 - a) Cíl a motivace
 - b) Charakteristika**
 - c) Možnosti výzkumu
- Větná polarita a vyhledávání
- Závěr

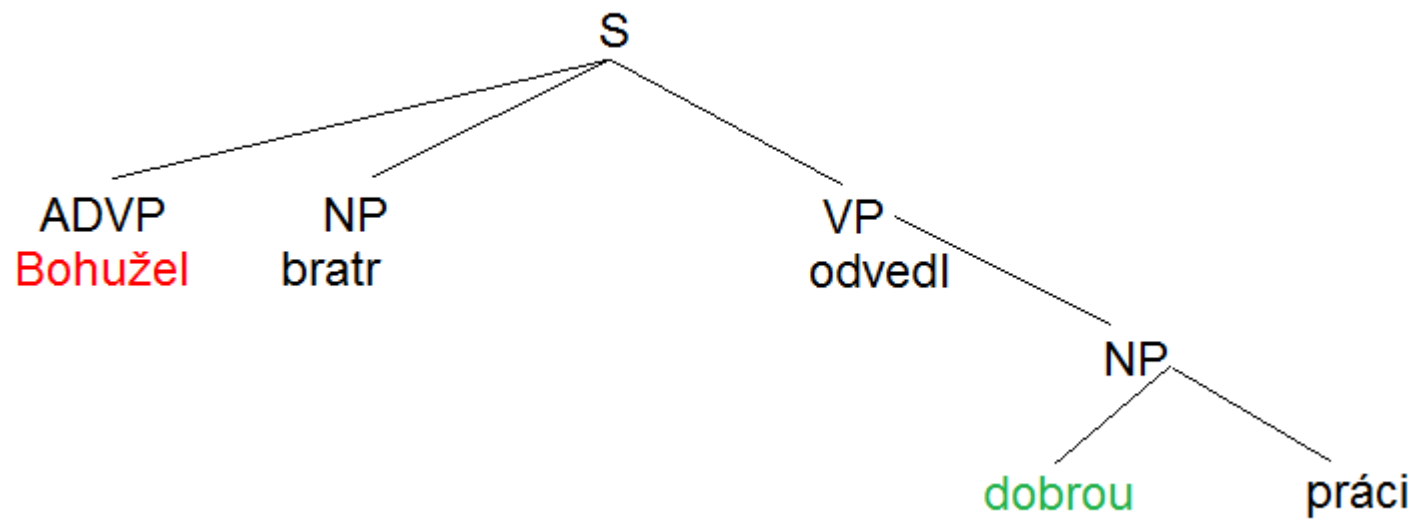
Polarita na úrovni slova

Slovo

- neobsahuje samo o sobě žádnou jasnou informaci o polaritě
- obsahuje (obecnou) morfologickou charakteristiku (x tvarová homonymie)
- neobsahuje informace o hloubce uzlu ve stromě



Je to bohudíky za námi.



Další důležité rysy

- důležitá informace o pozici slova ve větě (je větným členem?)
 - dosah částic
 - určující vliv predikátu
 - druh klauze
- „síla“ spojky či předložky

Polarita na úrovni věty

- kontext – snazší orientace +/-
- stejně jako v textu rozhodující počet +/- výrazů, jejich větněčlenská platnost a dosah
 - počet polárních výrazů v hlavní větě
 - celkový počet „silných“ polárních výrazů
 - počet znegovaných pozitivních výrazů

(Př. *Ne hezky, nemám rád atp.*)

- Polarita - úvod do problematiky
- Větná polarita:
 - a) Cíl a motivace
 - b) Charakteristika
 - c) Možnosti výzkumu
- Větná polarita a vyhledávání
- Závěr

Možnosti výzkumu

- vytvořit „**subjectivity lexicon**“ pro češtinu – set frekventovaných polárních výrazů s přisouzenou platností + / -
- implementace hodnot těchto výrazů do vět v korpusu
- trénovací data – porovnání úspěšnosti
(s jakou přesností odpovídá výsledná polarita na označkových datech)

- Polarita - úvod do problematiky
- Větná polarita:
 - a) Cíl a motivace
 - b) Charakteristika
 - c) Možnosti výzkumu
- **Větná polarita a vyhledávání**
- Závěr

Větná polarita a vyhledávání (QA)

- Cílem je najít v korpusu dat věty, které mají nějakou polaritu vzhledem k danému dotazu
- Zadání:
Vyhledat pozitivní názory spojené se jménem Alexandra Sticha.
- Výsledek:
K největším **zásluhám všestranného** bohemisty Alexandra Sticha patří to, že na počátku 90. let minulého století inicioval a zaštil studium jazyka a literatury doby baroka.

Předpoklady úspěšného vyhledávání

- správně označované polární výrazy
- **shoda** vyhledávaného jména/pojmu a polárního výrazu
- blízkost vyhledávaného jména/pojmu a polárního výrazu

- Polarita - úvod do problematiky
- Větná polarita:
 - a) Cíl a motivace
 - b) Charakteristika
 - c) Možnosti výzkumu
- Větná polarita a vyhledávání
- Závěr

Závěr

- Větná polarita:
 - jednotlivá slova sama o sobě neposkytují relevantní informaci
 - předpoklad relevantních výsledků:
větší množství slov + informace o jejich polaritě
+ další charakteristika
- Větná polarita vztažená k tématu:
 - velký podíl jazykových rysů
 - důležitá informace o vzdálenosti daných výrazů

Souhrn faktorů ovlivňujících větnou polaritu

- Morfologické vlastnosti
- Hloubka polárního výrazu ve stromě
- Polarita predikátu
- Druh klauze
- „Síla“ polárního výrazu
- Dosah negace (*ne zrovna milý*)
- Dosah intenzifikátoru (*velmi milý*)
- Vliv modality
- Modifikace pomocí pozitivního nebo negativního polárního výrazu
(*hrozně milý, pěkný hlupák*)

Ideální věta

Polární výraz:

- Morfologie: sloveso
- Větný člen: predikát
- Hloubka ve stromě: nízká
- Typ věty: hlavní

+ modifikace polárním výrazem se **stejnou platností**

- pozitivní pozitivním a naopak – *strašně zlý*

Literatura:

- ***Annotating Expressions of Opinions and Emotions in Language.*** J. Wiebe, T. Wilson, C. Cardie. *Language Resources and Evaluation, 2003.*
- ***Making Large-Scale SVM Learning Practical. T. Joachims.*** In B. Schölkopf, C Burges, and A. Smola eds. *Advances in Kernel Methods- Support Vector Learning. MIT Press, 1997.*
- ***Examining the Role of Linguistic Knowledge Sources in the Identification and Classification of Reviews.*** V. Ng, S. Dasgupta, and S. M. Niaz Arifin. *Proceedings of the COLING/ACL, 2006.*

Děkuji za pozornost