

Sentence sentiment classification using fuzzy word matching combined with fuzzy sentiment classifier

Marcin PIETRAS



Max de Raad, Manon Schriever, Wout Kooijman

Problem definition

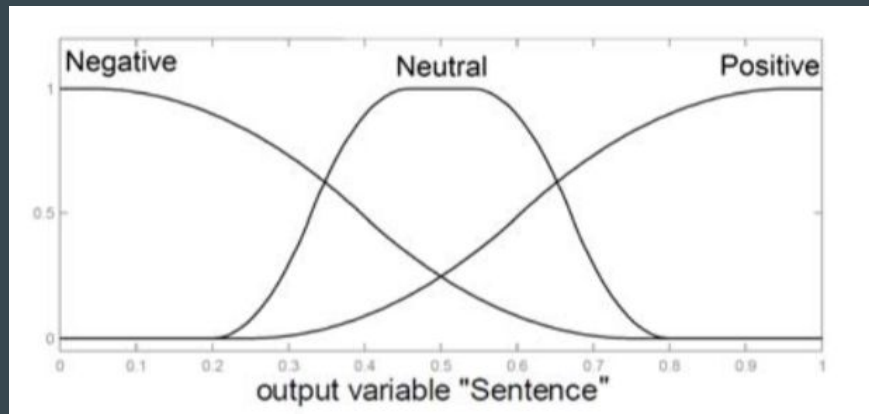
- Automated sentiment analysis provides useful information
- Useful in the field of behavioural targeting and contextual advertising
- Words can have different senses depending on the expression

Proposed solution

- Analyse whole sentences with a fuzzy system
- Using a database of (Polish) words with assigned sentimental value
- Modified Levenshtein algorithm to map words with different suffixes/prefixes
- Different input variables
 - **Neutral** (Neutral / Not Neutral) determines whether a given word carries a sentimental value at all.
 - **PreciseNegPos** (Very Negative/ Negative/ Neutral / Positive / Very Positive) specifies the type of sentimental value of a word.
 - **GeneralNegPos** (Negative/ Neutral/ Positive) determines the degree of sentimental value expression.
 - **SoPMI** (Negative/ Neutral/ Positive) Semantic orientation, defines the semantic orientation of a word with relation to the environment.

Our approach

- Our idea: combining different sentiment lexicons
- Different input, similar output
- Levenshtein distance
- Lookup table
- Mean of words



Evaluation of the approach in the paper

- Suffix Mitigated Levenshtein Distance

- Fuzzy Input →

$$(1) \text{ FuzzyIN} = \frac{\left(\sum_{i=1}^{NoSw} WsP(i) \cdot SML(i) \right)}{NoSw}$$

- Ruleset consisting of 15 rules →

10. (Neutral==NotNeutral) & (GeneralNegPos==Positive) & (PreciseNegPos==VeryPositive) & (SoPML==Positive) => (Sentence=Positive) (1)

Results

Table 2. Summary of the word matching and related to it the sentimental value

Estimator	rodzina	ciepły	życzliwy	troskliwy	luksusowy	Average
WordSimilarity	0.667	1.000	0.778	1.000	0.889	0.8667
Neutral	0.167	1.000	0.889	1.000	0.944	0.8000
GeneralNegPos	0.500	1.000	0.778	1.000	0.944	0.8667
PreciseNegPos	0.500	0.500	0.889	0.750	0.500	0.6278
SO-PMI	0.720	0.850	0.889	0.300	0.588	0.6696

Result of fuzzy sentiment classification: **Positive 0.7626**

In our family home it was warm, kindly, lovingly, although without the luxuries.

I've enough of infantile, egocentric, devoid of empathy, ugly guys who aren't able to grow up

Table 3. Summary of the word matching and related to it the sentimental value

Estimator	infantyl ny	egocentr yczny	pozbawić	brzydki	facet	dorosnąć	Avera ge
WordSimilar.	0.833	0.867	0.667	0.778	0.714	1.000	0.810
Neutral	0.917	0.933	0.167	0.889	0.143	1.000	0.675
GenNegPos	0.083	0.067	0.500	0.111	0.143	0.000	0.151
PrecNegPos	0.083	0.067	0.167	0.111	0.143	0.000	0.095
SO-PMI	0.416	0.327	0.313	0.788	0.614	0.140	0.433

Result of fuzzy sentiment classification: **Negative 0.2337**

Criticism

- One Dataset
- Connected input parameters (GeneralNegPos VS PreciseNegPos)
- Modified Levenshtein algorithm
- No mixed sentence results