Rule-Based Sentiment Analysis in Narrow Domain
Detecting Sentiment in Daily Horoscopes Using *Sentiscope*

Željko Agić and Danijela Merkler

University of Zagreb
Faculty of Humanities and Social Sciences

SAAIP 2012, Mumbai, India, 2012-12-15
Overview

- motivation

- system design and implementation
  1. collecting horoscope texts from the web on a daily basis
  2. rule-based module for polarity phrase detection designed in NooJ linguistic development environment
  3. web-based wrapper application for counting polarity phrases and assigning overall sentiment scores
  4. simple visualization module

- evaluation

- rule-based component demo and visualization demo
Document collection

- developed a simple focused crawler
- collected horoscopes from largest websites (in Croatian)
  - selected by Google search index
  - eight different newspaper portals and specialized portals
- collected from 2012-02-11 to 2012-05-10
- 7,716 articles, 484,179 tokens
Inter-annotator agreement

- development set of 333 articles manually annotated by two human annotators for overall sentiment and polarity phrases
- lineary weighted kappa: 0.593 → moderate agreement
- excluding neutral sentiment, kappa: 0.989 → very good agreement

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>-</th>
<th>x</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>94</td>
<td>0</td>
<td>26</td>
<td>120</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
<td>82</td>
<td>31</td>
<td>114</td>
</tr>
<tr>
<td>x</td>
<td>18</td>
<td>4</td>
<td>77</td>
<td>99</td>
</tr>
<tr>
<td>Σ</td>
<td>113</td>
<td>86</td>
<td>134</td>
<td>333</td>
</tr>
</tbody>
</table>
Overall article sentiment and polarity phrases

- positive phrases imply positive overall sentiment and vice versa
- also applies when both types of phrases are present
- even distribution of phrases for neutral sentiment articles
- justifies theoretical baseline that overall sentiment is assigned from the polarity group with the highest count

<table>
<thead>
<tr>
<th></th>
<th>&lt;p&gt;</th>
<th>&lt;n&gt;</th>
<th>both</th>
<th>&lt;p&gt; in both</th>
<th>&lt;n&gt; in both</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>410</td>
<td>27</td>
<td>23</td>
<td>85</td>
<td>27</td>
</tr>
<tr>
<td>-</td>
<td>19</td>
<td>321</td>
<td>15</td>
<td>19</td>
<td>53</td>
</tr>
<tr>
<td>x</td>
<td>142</td>
<td>145</td>
<td>67</td>
<td>117</td>
<td>115</td>
</tr>
</tbody>
</table>
Phrase detection

- designed in two stages — from scratch and by observing the development set
- grouped in two NooJ local grammars
  - positive and negative sentiment detection
- focus on three POS
  - adjectives, nouns and verbs
  - adverbs are homographic with adjectives in singular nominative case in neuter gender
- 170 negative and 139 positive words and phrases
- aggregate of positive and negative words which occur with a negation, which results in expressing the opposite sentiment
  - 33 negated positive and 17 negated negative words and phrases
- a total of 203 words and phrases for negative and 156 words and phrases for positive sentiment detection
Demo

Polarity phrase detection in NooJ
Evaluation

- conducted on a manually annotated held-out test set
  - initial run also on portion of development set
  - approximately 11,500 tokens in 168 articles each
- polarity phrase detection accuracy of the rule-based component

<table>
<thead>
<tr>
<th>sample</th>
<th>precision</th>
<th>recall</th>
<th>F₁-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>0.371</td>
<td>0.283</td>
<td>0.321</td>
</tr>
<tr>
<td>development</td>
<td>0.435</td>
<td>0.469</td>
<td>0.451</td>
</tr>
<tr>
<td>test</td>
<td>0.413</td>
<td>0.393</td>
<td>0.402</td>
</tr>
</tbody>
</table>
Evaluation

- System accuracy on overall sentiment detection and confusion matrix for overall sentiment assignment.
- System performance is high in discriminating between positive and negative overall sentiment.
- Accuracy steeply decreases upon inclusion of neutral sentiment.
- Positive words and phrases are more accurately detected.

<table>
<thead>
<tr>
<th></th>
<th>+*</th>
<th>-*</th>
<th>x*</th>
<th>precision</th>
<th>recall</th>
<th>F₁-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>40</td>
<td>3</td>
<td>17</td>
<td>0.677</td>
<td>0.666</td>
<td>0.671</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>25</td>
<td>17</td>
<td>0.555</td>
<td>0.568</td>
<td>0.561</td>
</tr>
<tr>
<td>x</td>
<td>17</td>
<td>17</td>
<td>30</td>
<td>0.468</td>
<td>0.468</td>
<td>0.468</td>
</tr>
</tbody>
</table>
Prototype web interface for data visualization
Conclusions and future work

- detecting sentiment in narrow domain such as daily horoscope texts is not easy to achieve
  - complex phrases and syntax
  - specific style, even for each individual author
- obtained results as baseline for further work
  - overall $F_1$-score: 0.566
  - $F_1$-score for phrase detection: 0.402
  - moderate inter-annotator agreement
- obtained data can be used for different types of linguistic analysis
- re-implementation of the link between polarity phrases and overall sentiment
  - elimination of neutral sentiment category
- model adjustment and application for sentiment annotation and visualization in other domains
  - precision and recall shown to be much higher (0.9, 0.6) using the same framework for financial texts
Thank you for your attention! 😊