SENTISCOPE: A SYSTEM FOR SENTIMENT ANALYSIS IN DAILY HOROSCOPES

DANIJELA MERKLER
ŽELJKO AGIĆ

Abstract

We present *Sentiscope* – a prototype system for collecting, sentiment annotation and visualization of daily horoscopes from news portals written in Croatian. System architecture and evaluation is provided and the system itself is made available to users on the web. The sentiment analysis component is based on rules implemented as finite state transducer cascades in NooJ development environment.

Introduction

Sentiscope is a prototype system for sentiment analysis in daily horoscopes written in Croatian. It crawls the Croatian web on a daily basis and collects horoscope texts from several specialized websites and daily news portals. The texts are processed with a manually designed rule-based module for polarity phrase detection. The texts are then assigned with overall sentiment scores which are calculated by counting polarity phrases. The results of semantic processing are stored and the texts with the respective annotations of both polarity phrases and the overall sentiments are provided to users via a graphical user interface in the form of a web application.

Implementation of *Sentiscope* draws from the work on approaches to sentiment analysis in financial texts and related work on sentiment analysis presented in e.g. (Ahmad et al. 2005, 2006a, 2006b, Almas and Ahmad 2007, Devitt and Ahmad 2008, 2009, Daly et al. 2009, Remus et al. 2009). More specifically, drawing from the experiment with rule-based sentiment analysis in financial reports written in Croatian presented in (Agić et al. 2010) – which resulted with a high precision prototype system – and the previously mentioned work on rule-based sentiment analysis in general, we attempted to approach the problem of sentiment analysis in Croatian text from a very specific, narrow and expectedly difficultly processable genre, i.e. horoscope text from the web. In the paper, we

describe the system implementation and evaluation on the tasks of detecting polarity phrases and detecting overall article sentiment. The system prototype is available on the web (URL?).

System implementation

System overview is given in Figure 1 (left side). The system is basically a web- and Linux-based application built by open source technologies and it consists of four main components:

1) the focused web crawler written in PHP that collects and stores horoscopes from a number of Croatian horoscope and daily news portals,
2) the rule-based sentiment detector that detects positive and negative polarity phrases in horoscope text and is implemented as a set of local grammars designed in the NooJ linguistic development environment (Silberztein 2004, 2005),
3) overall sentiment detector written in PHP that estimates overall article, i.e. horoscope sentiment by counting positive and negative polarity phrases and
4) the graphical user interface for assessing sentiment annotated daily horoscopes and sentiment statistics over periods of time, as illustrated by Figure 2.

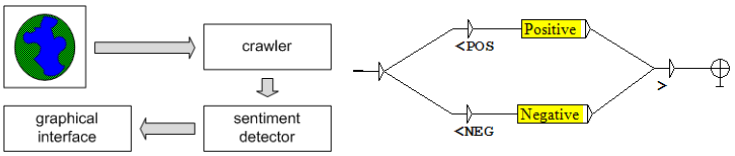

Figure 1. System overview and main polarity phrase detection grammar

All horoscopes, respective polarity phrase annotations and overall sentiment scores are stored in a MySQL database. The user interface currently provides daily horoscopes with inline annotations for all twelve zodiac signs (see Figure 2) and historical data in the form of overall sentiment diagrams. Both visualizations also conveniently and entertainingly serve as indicators of sentiment inconsistencies across zodiac signs and web sources. However, regardless of the overall purpose (or purposelessness) of such texts, it is shown here that texts from the specific horoscope genre written in Croatian are very difficult to process

with respect to sentiment annotation and thus deserving the given research focus.

2012-07-18 **ovan**

24sata.hr

Uznemirenost i loši predosjećaji činit će vas ***nemirnima*** cijeli dan. Intuitivno ćete izbjegavati blizinu osoba za koju osjećate da šire ***negativnu*** energiju. ***Napeti*** obiteljski odnosi narušavat će vaš unutarnji mir, što će vas prisiliti na neke ***odgađane*** odluke. Patite li od alergije, izbjegavajte ***štetne*** utjecaje iz okoliša.

dalmacijanews.com

Konačno ćete moći postaviti pravu dijagnozu za sve svoje neuspjehe. Bit ćete prisiljeni potisnuti neke svoje ***vrline***, za dobrobit sebe i svojih ***najdražih***. Konačno ćete shvatiti kako je dosta postavljanja ljubavi na prvo mjesto, kad vam nije uzvraćena. Poželjet ćete za sebe ***samo najbolje*** i u tome ***ćete uspjeti***. Ljubav vas vodi po labirintu. Saznat ćete ***zanimljive*** događaje o osobi ***koju volite***.

Figure 2. Screenshot of the system

As mentioned previously, overall article sentiment is estimated from the number of detected phrases denoting positive or negative sentiment. Currently, articles are tagged as positive if the number of positive phrases is greater than the number of negative phrases contained within them and vice versa. If their counts are equal, the article is tagged as neutral. Polarity phrase detection is done by using a series of rules in form of local grammars or lexical finite state transducer cascades implemented in NooJ linguistic development environment, as illustrated in Figure 3.

Rules were designed in two stages – first from scratch and then by observing a development set of horoscope texts. For development and testing, we have collected horoscopes from seven largest Croatian websites containing daily horoscopes as indicated by the Google search index. Horoscopes were collected from 2012-02-11 to 2012-05-10. 7,716 articles with 484,179 tokens were collected. 333 articles were chosen for the development set and were manually annotated for overall sentiment and polarity phrases. Observed agreement of 75.97% on overall sentiment annotation was measured between the two annotators. The stats are given in Table 1 and they indicate that the disagreement between the annotators was distributed almost exclusively within the category of neutral sentiment articles. The annotators agreed on positive sentiment in 80.69% of the

annotations, while the observed agreement was 82% on negative sentiment and 66.09% on neutral sentiment.

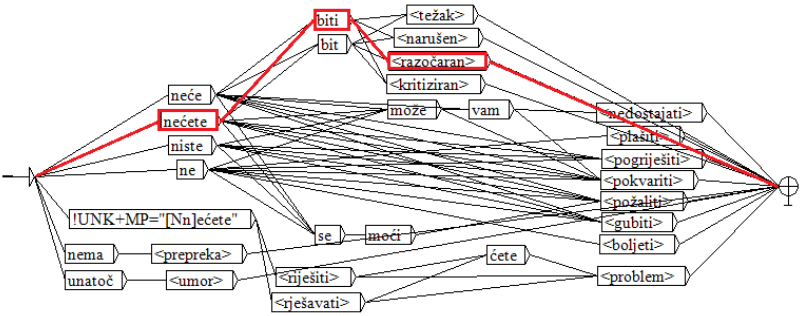

Figure 3. Example of positive polarity phrase detection – hr. *nećete biti razočarani* (en. *you will not be disappointed*)

	+	-	x	∑
+	94	0	26	120
-	1	82	31	114
x	18	4	77	99
∑	113	86	134	333

Table 1. Annotator agreement on overall sentiment

Table 2 emphasizes the relation between the polarity phrases detected in articles and the overall sentiment of the articles and as such, it is the theoretical baseline for building a system that estimates overall sentiment of text from the number and type of polarity phrases that it contains. The table shows that the positive sentiment articles tend to contain much more positive polarity phrases, as 71.80% of the positive polarity phrases was found in positive sentiment articles, as opposed to 3.33% in negative and 24.87% in neutral sentiment articles. The same was found to apply for negative polarity phrases: 65.11% of them were located in negative sentiment articles, 5.48% in positive sentiment articles and 29.41% in articles carrying neutral overall sentiment. This justified a system design in which polarity phrases are counted in articles and overall sentiment assigned from the polarity group with the highest count. In addition to this, Table 2 also shows the number of articles in which both positive and negative polarity phrases were observed (table column *both*), along with separate counts of positive and negative polarity phrases (table columns *<p> in both* and *<n> in both*) for these articles. The distribution further

supports the system design, being that positive polarity phrases are once again predominant in positive sentiment articles (75.89% positive vs. 24.11% negative) and negative polarity phrases dominate in negative sentiment articles (73.61% negative vs. 26.39% positive) while they are almost evenly spread in neutral sentiment articles (50.43% positive vs. 49.57% negative polarity phrases).

	<p>	<n>	both	<p> in both	<n> in both
+	410	27	23	85	27
-	19	321	15	19	53
x	142	145	67	117	115

Table 2. Overall article sentiment (+, -, x) vs. Polarity phrases (<p>, <n>)

Rules for polarity phrase detection are grouped in two NooJ local grammars – one for positive sentiment and one for negative sentiment detection (see Figure 1, right side). Each of theese grammars consists of lists of words and phrases for three parts-of-speech: adjectives, nouns and verbs. Words and phrases are manually derived from a number of daily horoscopes and – except for the characteristic key words and key phrases for the horoscope domain – there is a number of domain independent words and phrases (e.g. *dobro* (en. *good*), *izvrsno* (en. *great*), *odlično* (en. *excellent*), etc. for positive sentiment, and *loše* (en. *bad*), *slabo* (en. *weak*), *nedovoljno* (en. *unsatisfying*), etc. for negative sentiment). We derived 170 words and phrases for negative and 139 words and phrases for positive sentiment detection. In addition to the lists of positive and negative sentiment phrases based on their POS, there is also an aggregate of words which express positive or negative sentiment in itself, but in a cotext, they often occur with a negation, which results in expressing the opposite sentiment. In the rules, there are 33 negated positive and 17 negated negative words and phrases (an example grammar for detecting negated negative words and phrases is given in Figure 3), which adds up in a total of 203 words and phrases for negative sentiment detection and 146 words and phrases for positive sentiment detection.

Evaluation

The evaluation was conducted on a manually annotated held-out test set containing 11,500 tokens in 168 articles. The initial prototype of the polarity phrase detection module, that was designed from scratch in NooJ, was first evaluated on the test set in a form of a *dry run* test for purposes

of further development. The results are given in Table 3 joint for positive and negative polarity phrases. The results of the dry run were shown to be rather low, with an F_1-score of only 0.321. The rules were thus tuned, as previously mentioned, by observing the development set and another two tests were performed with the improved rules – one on the development set itself and the other on the test set. These results are also given in Table 3 and they show a substantial improvement over the baseline for both the development set and the test set. Being that the horoscope texts are highly complex in terms of irregularities of phrases, i.e. showing rare re-occurrences of polarity phrases among texts from varying sources, these scores were considered to be a satisfactory entry point for overall article sentiment detection.

sample	precision	recall	F_1-score
initial	0.371	0.283	0.321
development	0.435	0.469	0.451
test	0.413	0.393	0.402

Table 3. Overall system accuracy and confusion matrix

The results of system evaluation with respect to overall article sentiment are given in Table 4. The rows of the confusion matrix represent gold standard annotation while the columns present system annotation. The matrix clearly indicates that the system performance is high for the task of discriminating between positive and negative overall sentiment, while its accuracy steeply decreases upon inclusion of the neutral sentiment article category. This observation is also supported by the observed inter-annotator agreement and the data in Table 1 and 2. The correlation between the number of polarity phrases and overall sentiment given in Table 2 is clearly manifested in the evaluation results, being that the overall performance of the system is satisfactory even with the rule-based phrase detection module performance might be considered somewhat low in absolute terms, especially with respect to those obtained for e.g. well-structured financial texts (cf. Agić et al. 2010).

	+*	-*	x*	precision	recall	F_1-score
+	**40**	**3**	17	0.677	0.666	0.671
-	**2**	**25**	17	0.555	0.568	0.561
x	17	17	33	0.468	0.468	0.468

Table 4. Overall system accuracy and confusion matrix

Table 4 also shows that positive words and phrases are more accurately detected than the negative ones – the observed difference in F_1-scores of the positive and negative phrase detection is as high as 0.11 in favor of the positive phrase detection. Considering that there are substantially more negative words and phrases in the rules for detection (203 vs. 146) and that there are also considerably more negated positive phrases than vice versa (33 vs. 17), we can conclude that in this type of texts, unlike positive sentiment which is expressed more clearly and explicitly, negative sentiment is often covert and masked with various modifiers and within very complex expressions where negations occurs far from the positive word (e.g. in hr. *danas nećete imati baš dobar dan*, en. *you will not have such a good day today*), so they are very difficult to detect with the rules.

aries	x	x	+	x	+	+	x
taurus	-	+	+	+	x	x	x
gemini	+	-	+	-	x	x	x
cancer	-	+	+	x	-	-	x
leo	x	x	x	-	-	x	-
virgo	-	+	+	+	x	+	-
libra	-	-	+	-	+	+	x
scorpio	x	+	x	-	x	-	-
sagittarius	+	+	x	-	-	-	x
capricorn	x	x	+	+	x	x	x
aquarius	+	-	x	-	+	-	+
pisces	+	+	+	+	x	x	x

Table 5. Horoscope sentiment by web-source on 2012-05-18

Table 5 is an illustration of the sentiment trend information provided by the system. As mentioned previously, the texts are processed on a daily basis and both the texts and the respective annotations are stored in a database. This enables graphical display of sentiment trend across text sources (websites) and text categories (zodiac signs). The table indicates that the overall horoscope sentiment is consistently inconsistent across the seven different web sources and – perhaps even more interestingly – that the possible consistencies might be observed only within single web sources, not respecting the zodiac signs. In the specific case of sentiment analysis in the narrow domain of daily horoscope texts, this might therefore support the claim that perhaps the most reliable sentiment detection feature is the daily sentiment of the text authors.

Conclusions and future work

Detecting text sentiment in a very specific and narrow domain such as daily horoscope texts has shown not to be trivial and easy to achieve, given that such texts are characterized both by a specific and often very complex phrases and syntax and a particular, domain-dependent style, which can be specific for each individual author, too. This considered, obtained F_1-score of 0.566 for overall system accuracy and 0.402 for phrase detection accuracy, with annotator agreement of 75.97%, are here regarded as satisfactory and useful. For future work, obtained data can be used for different types of linguistic analysis, e.g. discourse analysis, sociolinguistic analysis, etc. Besides, the developed model can be easily adjusted and applied for sentiment annotation and visualization in other domains as well.

Acknowledgement

The results presented here were partially obtained from research within project CESAR (ICT-PSP, grant 271022) funded by the European Commission, and partially from projects 130-1300646-0645 and 130-1300646-1776 funded by the Ministry of Science, Education and Sports of the Republic of Croatia.

References

Agić Ž, Ljubešić N, Tadić M. (2010). Towards Sentiment Analysis of Financial Texts in Croatian. Proceedings of LREC 2010. ELRA, 2010, pp. 1164-1167.

Ahmad K, Gillam L, Cheng D. (2005). Society Grids. Proceedings of the UK e-Science All Hands Meeting 2005. Swindon, EPSRC, 2005, pp. 923-930.

Ahmad K, Gillam L, Cheng D. (2006). Sentiments on a Grid: Analysis of Streaming News and Views. Proceedings of the 5th International Conference on Language Resources and Evaluation. ELRA, 2006.

Ahmad K, Cheng D, Almas Y. (2006). Multi-lingual Sentiment Analysis of Financial News Streams. Proceedings of the First International Conference on Grids in Finance. International School for Advanced Studies, Trieste, Italy, 2006.

Almas Y, Ahmad K. (2007). A note of extracting "sentiments" in financial news in English, Arabic and Urdu. The Second Workshop on Computational Approaches to Arabic Script-based Languages. Linguistic Society of America, 2007, pp. 1-12.

Daly N, Kearney C, Ahmad K. (2009). Correlating Market Movements With Consumer Confidence and Sentiments: A Longitudinal Study. Text Mining Services. Leipzeiger Beitrage zur Informatik, 2009, pp. 169-180.

Devitt A, Ahmad K. (2007). Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. Proceedings of the Anual Meeting of the Association of Computational Linguistics (ACL 2007). Prague, Czech Republic, 2007.

Devitt A, Ahmad K. (2008). Sentiment Analysis and the Use of Extrinsic Datasets in Evaluation. Proceedings of the 6th International Conference of Language Resources and Evaluation. ELRA, 2008.

Remus R, Heyer G, Ahmad K. (2009). Sentiment in German language news and blogs, and the DAX. Text Mining Services. Leipzeiger Beitrage zur Informatik, 2009, pp. 149-158.

Silberztein M. (2004). NooJ: an Object-Oriented Approach. INTEX pour la Linguistique et le Traitement Automatique des Langues, C. Muller, J. Royauté M. Silberztein (eds), Cahiers de la MSH Ledoux. Presses Universitaires de Franche-Comté, pp. 359-369.

Silberztein M. (2005). NooJ's Dictionaries. Proceedings of the 2nd Language and Technology Conference, Poznan University, 2005.

Vučković K, Tadić M, Bekavac B. (2010). Croatian Language Resources for NooJ. CIT – Journal of computing and information technology, 18 (2010), pp. 295-301.

Vučković K, Tadić M, Bekavac B. (2010). Croatian Language Resources for NooJ. Proceedings of the 32nd International Conference on Information Technology Interfaces. Zagreb, SRCE University Computer Centre, University of Zagreb, 2010, pp. 121-126.