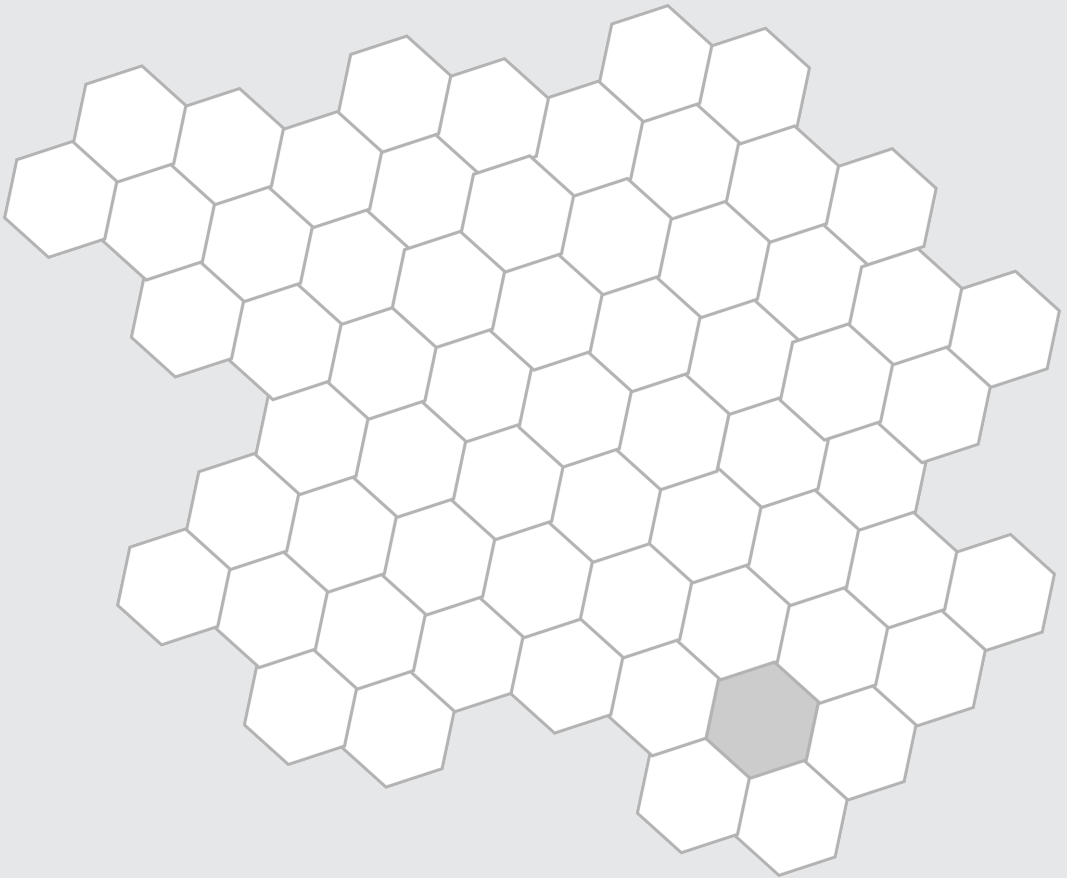


INTERNATIONAL JOURNAL ON **SOCIAL MEDIA**

MMM: **M**ONITORING,
MEASUREMENT, AND
MINING



I, 2010, 1
ISSN 1804-5251

...OFFPRINT...

INTERNATIONAL JOURNAL ON SOCIAL MEDIA

MMM: MONITORING, MEASUREMENT, AND MINING

I, 2010, 1

Editor-in-Chief: Jan Žižka

Publisher's website: www.konvoj.cz

E-mail: konvoj@konvoj.cz, SoNet.RC@gmail.com

ISSN 1804-5251

No part of this publication may be reproduced, stored or transmitted in any material form or by any means (including electronic, mechanical, photocopying, recording or otherwise) without the prior written permission of the publisher, except in accordance with the Czech legal provisions.

Opinion Analysis of Publications on Economics with a Limited Vocabulary of Sentiments

ANGELS CATENA

*Department of French and Romance Philology, Faculty of Philosophy and Arts, Autonomous
University of Barcelona, 08193 Bellaterra, Spain
e-mail: angels.catena@uab.cat*

MIKHAIL ALEXANDROV

*Department of French and Romance Philology, Faculty of Philosophy and Arts, Autonomous
University of Barcelona, 08193 Bellaterra, Spain
e-mail: malexandrov@mail.ru*

NATALIA PONOMAREVA

*Statistical Cybermetrics Research Group, School of Computing and IT, University of
Wolverhampton, Stafford Str, WV1 1SB Wolverhampton, UK
e-mail: nata.ponomareva@wlv.ac.uk*

Abstract

Opinion Analysis (OA) is a part of so-called Sentiment/Subjectivity Analysis, which aims to evaluate the author's personal characteristics and his/her attitude to objects and events. The existing well-known OA-systems use large vocabularies of classified sentiments (thousands of words) to give a positive or negative answer. In the paper we consider another case when the sentiment vocabulary is very limited (one-two hundreds of words) and the answer list includes an additional neutral category. We study OA-accuracy of Spanish documents related to economic crisis. Decision-making is implemented on regression model trained on examples. We show the dependency of OA-quality on a) granularity of sentiments and opinions b) rules used in regression model. We also compare the results with those obtained in Bo Pang and Maite Taboada research groups. In case of binary classification of sentiments and opinions the results prove to be similar.

Key words opinion analysis; sentiment vocabulary; regression model

Biographical note

Angels Catena is a member of the fLexSem Research group of the Autonomous University of Barcelona (UAB) in Spain. She works as assistant professor for the Department of French and Romance Philology (UAB). She is a linguist and an author of several publications related to lexicography, translation studies and pedagogical lexicography. Her current topics of research are paraphrase recognition and social networks.

Mikhail Alexandrov is a member of the fLexSem Research group at the Autonomous University of Barcelona in Spain. He is a professor of the Academy of National Economy under Russian Government. He is an applied mathematician and author of numerous publications related to mathematical modelling and natural language processing. His current topics of research are machine learning (inductive modelling, clustering) and internet technologies (social networking).

Introduction

Paper terminology

In this paper we use the following terminology:

Sentiments are words having a positive or negative sense in Opinion Analysis (OA). Sentiments are presented in the form of 4 vocabularies: nouns, verbs, adjectives and adverbs

Sentiment classification is a list of sentiment categories. We use two classifications: a rough two-level classification (positive and negative) and a detailed classification of 4 levels (very positive, positive, very negative and negative). Individual sentiment contribution in the first case is equal to 1 and in the second case to 1 and 0.5 respectively.

Opinion classification is a list of opinion categories. We use two classifications: a rough two-level classification (positive and negative) and a detailed classification of 4 levels (very positive, positive, very negative and negative). When the neutral category is used, the rough classification includes 3 categories (positive, negative and neutral) and the detailed classification includes 5 categories (very positive, positive, neutral, negative and very negative). According to these categories an expert evaluates each document using scales $(-1, 1)$ and $(-1, 0, 1)$ for the rough classification, and $(-1, -0.5, 0.5, 1)$ and $(-1, -0.5, 0, 0.5, 1)$ for the detailed classification.

Models of OA are all combinations of sentiment classifications with opinion classifications. It is easy to see that we have 4 such combinations: the rough categories of sentiments and the rough categories of opinions, the detailed categories of sentiments and the rough categories of opinions, etc.

The **regression model** for OA is lineal equation, the value of which is transformed into one of the categories from the opinion classifications. Arguments of the regression are so-called linguistic variables. Such a model is trained on examples prepared by experts, and then it is used on new texts. By 'lineal' we mean: a) linearity with respect to coefficients; b) linearity with respect to linguistic variables.

Linguistic variables can reflect the contribution of all positive sentiments, all negative sentiments, or their total contribution (the sum). When we have separate linguistic variables for positive and negative sentiments we deal with two-parameter regression. When the linguistic variable is a composition of positive and negative sentiments (i.e. the sum) we deal with one-parameter regression.

The **models for decision-making** on regression are the set of regression values and correspondent categories of opinions. One can change these rules and obtain different results.

Related works and problem settings

The general approach to OA that we follow in this paper is Machine Learning. Such an approach was proposed and developed by PANG ET AL. (2002) and PANG

AND LEE (2004, 2008). PANG AND LEE (2002) considered the domain of movie reviews. Their data included positive, negative and neutral reviews, but the authors concentrated only on positive and negative ones (700 and 700). They experimented with three standard methods: Naive Bayes classifier, maximum entropy classifier, and support vector machines with different sets of sentiments (2,600–32,300).

Maite Taboada's semantic orientation calculator SO-Calc is a well-known OA-system (TABOADA ET AL., 2006; BROOKE ET AL., 2009). SO-Calc uses 4 open vocabularies (nouns, adjectives, adverbs and verbs, about 5,000 sentiments in total). All sentiments are ranged on a 10-point scale. SO-Calc uses a regression model for binary classification. The authors experimented with a set of positive and negative reviews (200+200) covering 8 topics: books, cars, movies, etc.

Our tools are similar to those of Maite Taboada's group. We use 4 vocabularies, a program for calculation of sentiment contributions to a given document, and a regression model for decision-making. The difference consists in the size and granularity of vocabularies, granularity of opinions, and in flexible rules for decision-making on the regression equation.

The subject under consideration is documents related to the economic crisis: interviews, surveys, analytical papers, etc. We consider the following problems:

- 1) Sensibility of results to sentiment classification
- 2) Sensibility of results to opinion classification
- 3) Sensibility of results to rules of decision-making on regression

The paper consists of 5 sections. Section 2 describes data under consideration and sentiment vocabularies. Section 3 shows how the regression model is constructed and evaluated. Section 4 presents the results of all experiments. Section 5 contains the discussion of results and proposals for future work.

Parameterisation

Documents

The initial material consists of 50 papers in Spanish and Catalan. The papers vary greatly in length: from one to several pages. All papers were evaluated by two experts using a 4-point scale. Table 1 contains the titles (in English) and points of these documents. Table 2 shows the distribution of papers on categories. It is easy to see that the neutral category is small enough in comparison with polar categories in the rough opinion classification.

With the rough opinion classification all points 0.5 and -0.5 are transformed into 1 and -1 respectively.

Table 1: List of documents with their points (part of full list)

No	Title	Points
1	BBVA thinks that Spanish economy has already passed the point of recession	0.5
2	Spanish economy could enter to a long phase of stagnation according IESE	-1
3	The great problem of Spanish economy	-0.5
4	French economy grows in 3rd semester	0

In the paper we consider all possible models of OA. They are described in Table 3.

Vocabularies

The linguistic resources for OA are presented in the form of 4 vocabularies: nouns, verbs, adjectives and adverbs. All sentiments were ranged on a 4-point scale. Table 4 contains vocabulary descriptions. Table 5 presents the vocabulary of adverbs without English translation.

With the rough sentiment classification all points 0.5 and -0.5 are transformed into 1 and -1 respectively.

Parameterised documents

Document parameterisation consists of two steps:

- Evaluation of the sentiment contribution to a given document;
- Formation of the value for a linguistic variable(s).

In order to complete the first step we developed a program on Python. The input data of this program are the 4 vocabularies described above. The output data are numbers of positive and negative sentiments and their summary contribution

Table 2: Distribution of papers on categories

No	Category	Number	% with neutral categ.	% without neutral categ.
1	Very positive	3	6	7
2	Positive	17	34	39
3	Neutral	7	14	
4	Negative	15	30	35
5	Very negative	8	16	19

Table 3: Models of OA considered in the paper

No	Sentiment classification	Opinion classification
1	rough (2 categories)	rough (2 or 3 categories)
2	detailed (4 categories)	rough (2 or 3 categories)
3	rough (2 categories)	detailed (4 or 5 categories)
4	detailed (4 categories)	detailed (4 or 5 categories)

Table 4: Components of vocabularies

No.	Vocabulary	Size	Very positive	Positive	Very negative	Negative
1	Nouns	58	9	16	16	17
2	Verbs	50	8	15	13	14
3	Adjectives	47	9	9	15	14
4	Adverbs	33	8	9	7	9
	Total	188	34	49	51	54

to a given document. The program calculates the contribution simultaneously for the rough and the detailed sentiment classifications. Obviously, in the case of the rough classification the number of positive and negative sentiments coincides with their contribution (in absolute value) because each sentiment has a weight equal to 1. Table 6 shows a part of the full table prepared by the Python program. The contributions are located in the last four columns.

In order to complete the second step let us have a look at the distribution of positive and negative contributions within the framework of the detailed classification. Figure 1 shows this distribution. One sees a large spread. Since the regression value changes in a limited interval $[-1,1]$ such a large spread of contributions can lead to inconsistent regression.

To exclude this effect we normalise all contributions on the total number of sentiments. Therefore we have:

P_L = Positive contribution / Total number of sentiments

N_L = Negative contribution / Total number of sentiments

where P_L and N_L stand for positive and negative linguistic variables respectively.

Both linguistic variables now are located within the interval $[-1,1]$. Figure 2 shows their joint distribution.

We calculated the coefficient of correlation between P_L and N_L . It proved to be 0.91. With such a correlation we construct one linguistic variable instead of two:
 $L = P_L + N_L$

The set of linguistic variables is the final result of document parameterisation.

The regression model

Rules for decision-making on regression

Our goal is to construct and to test regression equations for all 8 models of OA reflected in Table 3. The regression equation is presented in the form:

$$R = a + bL$$

Here: R is the value of regression equation, L is the linguistic variable and a and b are unknown coefficients. Models for decision-making on regression are presented in Tables 7–10.

Table 5: Vocabulary of adverbs (in Spanish)

Positivo (0.5)	Muy positivo (1)	Negativo (-0.5)	Muy negativo (-1)
delante	positivamente	debajo	negativamente
suavemente	estupendamente	demasiado	duramente
favorablemente	brillantemente	inevitablemente	gravemente
tímidamente	excelentemente	difícilmente	cruelmente
moderadamente	felizmente	insuficientemente	alarmante
suficientemente	bien	tardíamente	desgraciadamente
oportunamente	gracias (a)	seriamente	dramáticamente
adecuadamente	acertadamente	lentamente	
convenientemente		precipitadamente	

Table 6: The results of Python program (part of full table)

Points	Text	Number of pos. sentim.	Number of neg. sentim.	Detailed classif., Positive	Detailed classif., Negative	Rough classif., Positive	Rough classif., Negative
0.5	T1.txt	19	11	9.5	-6.5	19	-11
-1	T2.txt	15	39	7.5	-27.5	15	-39
-0.5	T3.txt	2	6	1	-4	2	-6
-1	T4.txt	8	11	5	-8	8	-11
0.5	T5.txt	11	8	6	-4.5	11	-8
-0.5	T6.txt	22	14	11.5	-8.5	22	-14
-0.5	T7.txt	28	42	16	-32	28	-42

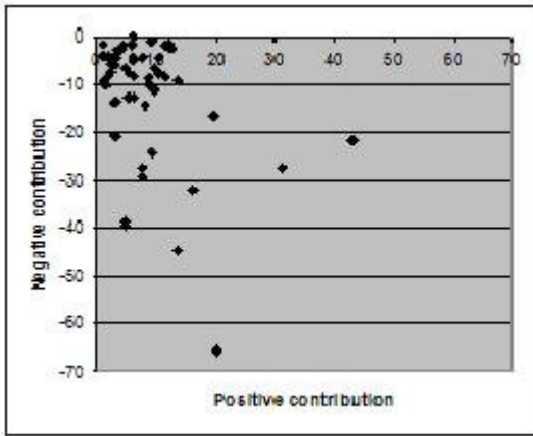


Figure 1:
Distribution before normalisation

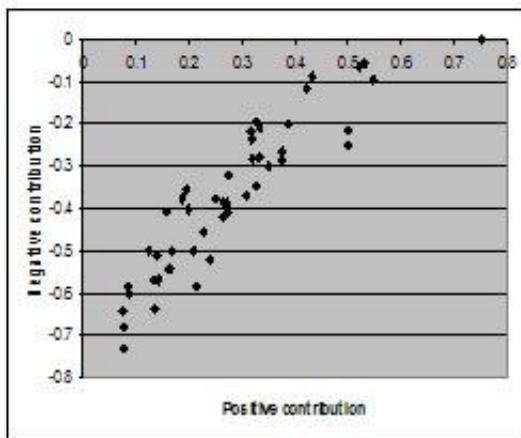


Figure 2:
Distribution after normalisation

Table 7: Rules for decision-making with 2 categories of opinions (the neutral category is absent)

Regression value	Opinion
$R < 0$	Negative
$R \geq 0$	Positive

Table 8: Rules for decision-making with 3 categories of opinions (the neutral category is present)

Regression value	Opinion
$R < -0.5$	Negative
$-0.5 \leq R \leq 0.5$	Neutral
$R \geq 0.5$	Positive

Table 9: Rules for decision-making with 4 categories of opinions (neutral category is absent)

Regression value	Opinion
$R < -0.75$	Very negative
$-0.75 \leq R < 0$	Negative
$0 \leq R \leq 0.75$	Positive
$R > 0.75$	Very positive

Table 10: Rules for decision-making with 5 categories of opinions (the neutral category is present)

Regression value	Opinion
$R < -0.75$	Very negative
$-0.75 \leq R < -0.25$	Negative
$-0.25 \leq R \leq 0.25$	Neutral
$0.25 < R \leq 0.75$	Positive
$R > 0.75$	Very positive

These intervals in Tables 7–10 should be assigned according to prior information about the distribution of opinion categories on axis R, but initially the interval stated above seems to be the most natural.

Evaluation of model quality

To evaluate the model quality we should select an index or indexes reflecting this quality. It can be the accuracy for all categories or for each category, the so-called F-measure for all categories or for each category, etc. A good survey of indexes for problems of classification is presented by PINTO (2008). In this paper we use the total accuracy, which is measured by the simplest formula:

$$\text{accuracy} = N_c / N_e$$

Here: N_c is the number of coincidences between expert opinions and model replies, N_e is the total number of experiments. $N_e=50$ when we use the neutral category and $N_e=43$ when we do not use it. One should say that the accuracy cannot be considered as the final quality index. It is necessary to take into account a so-called *Baseline*. It is the lowest value of accuracy which can be obtained on a given data set. Usually the Baseline is equal to the probability of the most frequent category. For this reason we introduce a so-called *adjusted accuracy*:

$$\text{adjusted accuracy} = \text{accuracy} - \text{Baseline}.$$

We can calculate the Baseline for all OA models using Table 3. The results are presented in Table 11.

To evaluate the *accuracy* we use the standard procedure of cross-validation. In this procedure all data are divided into a training and a control set. Then the model constructed on the training set is tested on the control one. Such an experiment is repeated on several partitions and the average error is calculated. In this paper we use leave-one-out cross validation when the training set contains N_e-1 data and the control set contains one data. Cross-validation is performed with a well-known package Weka (WEKA, 2009).

Experiments

Evaluation of Opinion Analysis without the neutral category

In this series of experiments we studied the accuracy of decision-making on a set of 43 documents. Table 12 contains the results of the calculation. These results are presented in graphic form in Figure 3.

Table 11: Baselines for different OA models

Opinion classification	Number of categories	Baseline
Rough classification without neutral class	2	0.53
Rough classification with neutral class	3	0.46
Detailed classification without neutral class	4	0.40
Detailed classification with neutral class	5	0.34

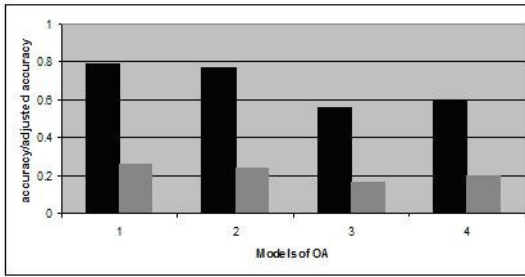


Figure 3:
Graphical illustration for Table 2

It is easy to see that in all cases the accuracy and adjusted accuracy of the rough opinion classification exceeds these values for the detailed opinion classification. In the case of the rough opinion classification the granularity of sentiments has no essential effect.

Evaluation of Opinion Analysis with the neutral category

In this series of experiments we studied the accuracy of decision-making on a set of all 50 documents. Table 13 contains the results of the calculation, also presented in graphic form in Figure 4.

The results show that the neutral category essentially decreases both accuracy and adjusted accuracy in comparison with the cases when the neutral category is absent. One of the principal reasons for such a situation is the relative small number of papers belonging to this category.

Evaluation of different rules for decision making on regression

In our previous experiments we used the rules for decision-making on regression presented in Tables 7–10. In this series of experiments we study OA with other rules.

Table 12: Values of accuracy and adjusted accuracy (the neutral category is absent)

Models of OA	Sentiment classification	Opinion classification	accuracy	Baseline	adjusted accuracy
1	rough (2 categories)	rough (2 categories)	0.79	0.53	0.26
2	detailed (4 categories)	rough (2 categories)	0.77	0.53	0.24
3	rough (2 categories)	detailed (4 categories)	0.56	0.40	0.16
4	detailed (4 categories)	detailed (4 categories)	0.60	0.40	0.20

Table 13: Values of accuracy and adjusted accuracy (the neutral category is present)

Models of OA	Sentiment classification	Opinion classification	accuracy	Baseline	adjusted accuracy
1	rough (2 categories)	rough (3 categories)	0.54	0.46	0.08
2	detailed (4 categories)	rough (3 categories)	0.52	0.46	0.06
3	rough (2 categories)	detailed (5 categories)	0.30	0.34	-0.04
4	detailed (4 categories)	detailed (5 categories)	0.42	0.34	0.08

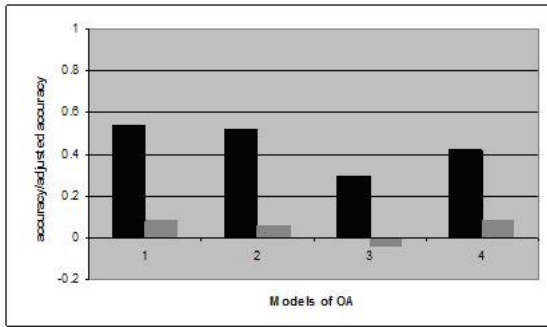


Figure 4:
Graphical illustration for Table 13

In the experiments we use all 50 documents, i.e. we consider the neutral category.

Tables 14 and 15 describes the rules with preference to the neutral category. Table 16 contains the results of the calculations and shows that there is no sense in preferences to the neutral category. It means that the linguistic variable for the neutral category is concentrated in a narrow interval near zero. Tables 17 and 18 describe the rules with preference to the polar categories adjacent to the neutral category. Table 19 contains the results of calculations.

The results show that the rules with preference to the polar categories and the detailed sentiment classification allow to obtain the best adjusted accuracy when

Table 14: Rules for decision-making with 3 categories of opinion (preference to the neutral category)

Regression value	Opinion
$R < -0.75$	Negative
$-0.75 \leq R \leq 0.75$	Neutral
$R \geq 0.75$	Positive

Table 15: Rules for decision-making with 5 categories of opinion (preference to the neutral category)

Regression value	Opinion
$R < -0.75$	Very negative
$-0.75 \leq R < -0.5$	Negative
$-0.5 \leq R \leq 0.5$	Neutral
$0.5 < R \leq 0.75$	Positive
$R > 0.75$	Very positive

Table 16: Values of accuracy and adjusted accuracy (preference to the neutral category)

Models of OA	Sentiment classification	Opinion classification	accuracy	Baseline	adjusted accuracy
1	rough (2 categories)	rough (3 categories)	0.32	0.46	-0.14
2	detailed (4 categories)	rough (3 categories)	0.34	0.46	-0.08
3	rough (2 categories)	detailed (5 categories)	0.28	0.34	-0.06
4	detailed (4 categories)	detailed (5 categories)	0.3	0.34	-0.04

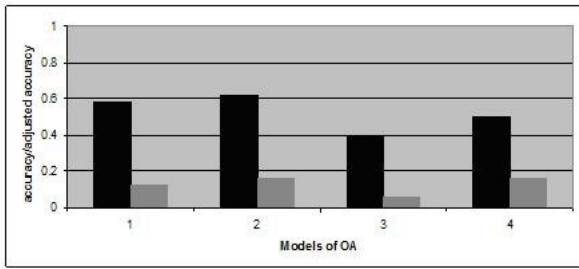


Figure 5:
Graphical illustration for Table 19

we deal with the neutral category. It concerns both the rough opinion classification and the detailed opinion classification.

Conclusion

Discussion

We completed OA of publications related to a given specific domain using very limited sentiment vocabularies. It was shown that in case of binary opinion classification regardless of the sentiment granularity the accuracy proved to be close

Table 17: Rules for decision-making with 3 categories of opinions (preference to the polar categories)

Regression value	Opinion
$R < -0.25$	Negative
$-0.25 \leq R \leq 0.25$	Neutral
$R \geq 0.25$	Positive

Table 18: Rules for decision-making with 5 categories of opinions (preference to the polar categories)

Regression value	Opinion
$R < -0.85$	Very negative
$-0.85 \leq R < -0.15$	Negative
$-0.15 \leq R \leq 0.15$	Neutral
$0.15 < R \leq 0.85$	Positive
$R > 0.85$	Very positive

Table 19: Values of accuracy and adjusted accuracy (preference to the polar categories)

Models of OA	Sentiment classification	Opinion classification	accuracy	Baseline	adjusted accuracy
1	rough (2 categories)	rough (3 categories)	0.58	0.46	0.12
2	detailed (4 categories)	rough (3 categories)	0.62	0.46	0.16
3	rough (2 categories)	detailed (5 categories)	0.40	0.34	0.06
4	detailed (4 categories)	detailed (5 categories)	0.50	0.34	0.16

to that obtained with large sentiment vocabularies and a very detailed sentiment classification ~ 0.8 (PANG AND LEE, 2002; BROOKE ET AL., 2009).

We studied the sensibility of OA to the sentiment classification and to the opinion classification. We showed that the adjusted rules of decision-making on regression equation allow to obtain satisfactory results when we deal with the neutral category.

Future work

In the framework of existing model of decision-making we suppose:

- To test the sensibility of OA to size of sentiment vocabularies;
- To consider publications on the same topic (economic crisis) written in Russian and in English;
- To consider publications on other topics (culture, politics) in Spanish.

We suppose to study OA:

- with mixed categories, such as neutral-positive and neutral-negative;
- with so-called ‘undefined’ category when it is better to say ‘I do not know’ than to give a certain answer.

In the paper we used the simplest lineal regression model with respect to linguistic variable. We intend to construct more complex models using the technique of Inductive Modelling (ALEXANDROV ET AL, 2009).

References

- ALEXANDROV, M., BLANCO, X., CATENA, A., PONOMAREVA, N. (2009): Inductive Modeling in Subjectivity/Sentiment Analysis (case study: dialog processing). In: *Proc. of 3rd Intern. Workshop on Inductive Modeling (IWIM-09)*. Poland, pp. 40-43.
- BROOKE, J., TOFILOSKI, M., TABOADA, M. (2009): Cross-Linguistic Sentiment Analysis: From English to Spanish. In: *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing*. Borovets (Bulgaria), pp. 50-54.
- PANG, B., LEE, L. (2004): A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: *Proceedings of the 42nd annual meeting of the Association for Computational Linguistics (ACL)*. Barcelona (Spain), pp. 271-278.
- PANG, B., LEE, L. (2008): Opinion mining and sentiment analysis. *Foundation and Trends in Information Retrieval*, Vol. 2, No. 1-2, pp. 1-135.
- PANG, B., LEE, L., VAITHYANATHAN, S. (2002): Thumbs up? Sentiment classification using Machine Learning techniques. In: *Proceedings of Conference on Empirical Methods in NLP*, pp. 79-86.
- PINTO, D. (2008): *On Clustering and Evaluation of Narrow Domain Short-Text Corpora*. (PhD thesis.) Valencia (Spain): Polytechnic Univ. of Valencia.
- TABOADA, M., ANTHONY, C., VOLL K. (2006): Creating semantic orientation dictionaries. In: *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)*. Genoa (Italy), pp. 427-432.
- Weka [on-line]. (2009): [cit. 2010]. Available at: <http://prdownloads.sourceforge.net/weka/weka-3-7-0jre.exe>.

Call for Papers

Papers to be included in the next issue should be preferably focused on topics related to social-networks in one or more of the following subjects (the list is indicative rather than exhaustive):

Sentiment/Opinion Analysis in Natural-Language Text Documents

Algorithms, Methods, and Technologies for Building and Analysing Social Networks

Applications in the Area of Social Activities

Knowledge Mining and Discovery in Natural Languages Used in Social Networks

Medical, Economic, and Environmental Applications in Social Networks

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Each of the submitted research papers should not exceed 26 pages. All papers are refereed through a peer review process.

Submissions should be send in the PDF form via email to the following address: SoNet.RC@gmail.com

Accepted papers are to be prepared according to the instructions available at <http://www.konvoj.cz/journals/mmm/>.