Link between Sentiment and Human Activity Represented by Footsteps
Experiment Exploiting IoT Devices and Social Networks

Jaromir Salamon and Roman Moucek
*Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia,
Universitni 8, Pilsen, Czech Republic*

Keywords: Sentiment Extraction, Sentiment Analysis, Body Sensors, Machine Learning, Experiment.

Abstract: The Internet of Things world brings to our lives many opportunities to monitor our daily activities by collecting data from various devices. Complementary to it, the data expressing opinions, suggestions, interpretations, contradictions, and uncertainties are more accessible within variety of online resources. This paper deals with collection and analysis of hard data representing the number of steps and soft data representing the sentiment of participants who underwent a pilot experiment. The paper defines outlines of the problem and presents possible sources of reliable data, sentiment evaluation, sentiment extraction using machine learning methods, and links between the data collected from IoT devices and sentiment expressed by the participant in a textual form. Then the results provided by using inferential statistics are presented. The paper is concluded by discussion and summarization of results and future work proposals.

1 INTRODUCTION

Analysis of human sentiment is currently a very popular research task. Sentiment, identified in and extracted from the text and/or speech, represents subjective emotion of the writer/speaker associated with some event, topic or situation. Sentiment analysis is then usually considered as the use of natural language/text processing methods and computational linguistics to identify this kind of information in the source text/speech.

However, emotions and feelings do not necessarily have to be identified using only the methods mentioned above. In addition to spoken and textual expressions there exist physiological symptoms and human activities that can be used as data sources for sentiment analysis. Since the aim of this article is not to provide a detailed analysis of how to measure emotions, feelings and identify sentiment, let us mention just a few examples. In (Chanel, 2006) there was investigated arousal dimension of human emotions from two different physiological sources: peripheral signals and electroencephalographic (EEG) signals from the brain. Emotions were recognized from brain activity, measured by the EEG signal, also in (Horlings, 2008).

Presence and quality of emotions, feelings, and sentiment are often associated with overall human life activities that include both physical activities (e.g. total daily movement, physical exercises, and sports training) and mental activities (e.g. learning, mental load, and stress resistance). Then numeric values obtained from various body sensors during common human activities that can affect sentiment could be considered as quantitative (depending on the number of sensors and their capturing rate) and qualitative (depending on the suitability of body sensors to the task) descriptions of emotional reactions of human body to a topic, event or situation.

As a support of this abstract statement let us imagine a situation when two people are arguing. During this activity we can expect that both subjects show up higher levels of blood pressure together with higher levels of heart rate. Also the content of their arguing written into a text form and translated to the sentiment should correspond to elevated levels of blood pressure and heart rate.

It is reasonable to expect that the quantity, quality, and availability of various body sensors will rapidly increase in near future. This will be accompanied by overall availability of various sensors within the IoT (Internet of Things) world.

Salamon, J. and Moucek, R.
Link between Sentiment and Human Activity Represented by Footsteps - Experiment Exploiting IoT Devices and Social Networks.
DOI: 10.5220/0005818204500457
In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 5: HEALTHINF*, pages 450-457
ISBN: 978-989-758-170-0

Then appropriate selection and use of suitable sensors will play an important role in case of context-aware applications.

Still, this article does want to present abstract design principles for building such a complex architecture, it focuses on design and implementation of a pilot study that considers two data sources for sentiment analysis: “classic” textual source representing subjective emotions of people participating in the experiment and raw data collected from body sensors representing human activities that could be associated with their emotional state. We have also taken into account that both data sources contain a lot of information that is present, but irrelevant to our task.

2 DATA SOURCES

Before we start looking for a link between the sentiment expressed by text/speech and data collected from body sensors, we need to find appropriate source for both.

Thus we started to look for such data sources. Both these types of data could be collected either simultaneously or serially (in any order) but have to be associated with a common emotional basis. So, we defined several data source categories that should be considered:

- The first category includes un-supervised data sources, typically data that are available on the Internet with widely defined source and time periods (e.g. a measurable activity implies some textual discussion, more specifically: the number of calories burned during sport activities and provided with a sport tracker implies discussion about the effectiveness of various physical exercises).
- The second category includes semi- or un-supervised data sources that are available on the Internet and/or as outputs from research projects with strictly defined source and time periods (e.g. blood pressure measurements giving results for specific countries, population, gender, or age accompanied by discussions about blood pressure).
- The third category includes supervised and tightly defined data sources. It could be, for example, an experiment on subjects that provide data for sentiment analysis by writing textual/spoken comments and collecting raw data from body sensors in a specified time period. The second example could be research that provides blood pressure sensory values and questionnaire results from patients having blood pressure difficulties in a specific time period.

The presented categories differ in requirements of what to be done during data collection. The third category seems to be the easiest one for final data interpretation. Since we were not successful in finding appropriate and free data sources (we found many sources satisfying required data source types, but the sources did not have common emotional basis), we decided to design and perform an experiment that would provide data satisfying the requirements for the third category of data sources.

3 EXPERIMENT

One subject (the first author of the paper) participated in a pilot experiment, which lasted 14 days. His task was to measure the number of footsteps using a professional pedometer and write short texts expressing his sentiment that were evenly distributed throughout the day. The original goal was to walk at least 10.000 steps a day and provide at least 20 short text messages a day. The final numbers of steps and text messages achieved during the pilot experiment are shown in Table 1.

Table 1: Daily measures for the experiment.

Date	Steps per day	Texts per day
08/18/2014	1816	8
08/19/2014	8284	23
08/20/2014	8903	19
08/21/2014	3775	25
08/22/2014	11633	28
08/23/2014	4335	21
08/24/2014	5852	22
08/25/2014	10426	17
08/26/2014	5503	12
08/27/2014	5078	19
08/28/2014	11945	21
08/29/2014	3833	13
08/30/2014	3273	22
08/31/2014	3279	15
09/01/2014	6656	15
Average	6625	20
Total	94591	280

The experiment started 08/18/2014 at 7:40 PM and ended 09/01/2014 at 12:40 PM. The average numbers of steps and text messages were calculated only for 13 days, because the experiment was carried out only for a limited number of hours during the first and last days. It is obvious from Table 1 that originally defined goals were not met, at least not on daily basis.

During the experiment the subject usually walked only when he really needed to go somewhere. When the subject wanted to reach the daily goal, he had to walk without any purpose. The distribution of steps is uneven during the day (cumulative numbers of steps and their distribution during the hours of the day are depicted in Figure 1). It was much easier for the subject to meet the target number of text messages. When he was behind with the daily goal, he shortened the time period he wrote them.

3.1 Sentiment Expression

The sentiment derived from the subject's data is a subjective measure. This leads to the question how to express and quantify it.

The recording phase of the experiment requires expression of the sentiment from the subject during a specific time frame. This is done through description of activities and related mental and/or physical processes that are accompanied by positive and negative feelings.

More formally, if D_t is an entire set of documents in a specific time frame $\{d_1, d_2, \ldots, d_n\}_t$ and $\{a_1, a_2, \ldots, a_m\}_t$ is a set of all activities in the same specific time frame, then the document d_j representing sentiment is created based on the specific activity a_i:

$$\{a_i\}_{(t_n)} \rightarrow \{d_j\}_{(t_n)} \quad (1)$$

Then the sentiment quantification leads to the classification task (if a text fragment is positively or negatively polarized), see (Cambria, 2013) or (Ikonomakis, 2005). This classification task is based on the occurrence of positive and negative words and their usage in a sentence or text fragment.

More formally, for an entire set of documents D_t in a specific time frame we have a classification set$\{c_1, c_2, \ldots, c_n\}_t$. Then one classification category is assigned to each document.

$$\{c_i\}_{(t_n)} \rightarrow \{d_j\}_{(t_n)} \quad (2)$$

3.2 Sentiment Collection

The requirement to describe subject's activities and mood during the whole day and throughout 14 days implied also requirements for the recording system. After considering available online systems as cloud services for text document recording, online task/planning systems, and discussion forums we decided to use the social networking service Twitter that has convenient characteristics. There is also extensive experience with sentiment classification on Twitter (Go, 2009).

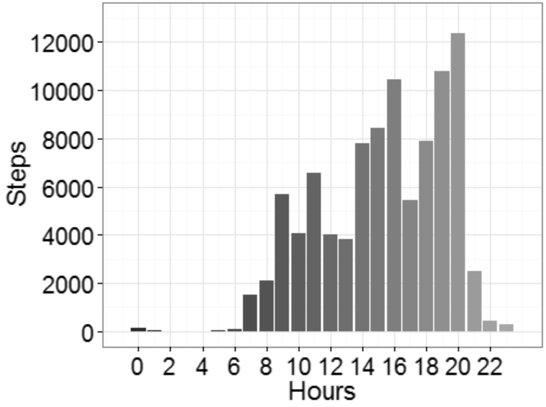

Figure 1: Cumulative numbers of steps and their distribution during the hours of the day.

We were afraid of one possible negative feature of Twitter when the text is limited to 140 characters. However, having experience from the pilot experiment we finally started to consider it as a practical feature.

As the result, descriptions of subject's activities and mood were recorded as up to 140 characters long texts with timestamps providing information about the order of each text in timeline. The limit of 140 characters finally seems to be advantageous from two points of view. At first, participants do not need to think about message content too much, they just express the current situation and their mood. Secondly, short messages support their willingness to express themselves often or at all.

3.3 Sentiment Extraction

The crucial part of this work was sentiment extraction and evaluation. Overall results and their precision are dependent on the sentiment polarity which was evaluated by human and also extracted from the experimental corpus.

3.3.1 Human Evaluation

After collecting a complete corpus of 280 documents we let a small group of 7 people p to evaluate all documents and assign one of three classification c values (1 for positive, -1 for negative and 0 for neutral) expressing sentiment to each document.

Using equations (3) and (4) we get an evaluated sentiment class c_i by summing up all numeric representations of classification per person p for a specific document d_i and normalizing it according to classes represented by original numeric values:

$$c_i = \frac{\sum_p c_{ip}}{\left|\sum_p c_{ip}\right|} \quad (3)$$

$$c_i = 0\ for \sum_p c_{ip} = 0 \quad (4)$$

Table 2 shows the results of human sentiment evaluation of the document corpus.

Table 2: Summary of human evaluated sentiment classes.

Class name	Classes count
negative (-1)	52
neutral (0)	21
positive (1)	207
Total	280

For every individual participant of human sentiment evaluation there was a calculated percentage consensus.

Table 3: Percentage consensus for human evaluated classes.

	A	B	C	D	E	F	G
A		49	7	46	40	49	51
B	49		6	50	47	52	52
C	7	6		6	6	6	7
D	46	50	6		47	49	53
E	40	47	6	47		45	48
F	49	52	6	49	45		83
G	51	52	7	53	48	83	

3.3.2 Machine Learning Sentiment Extraction

For the subsequent comparison to the human evaluation described above and for the future work over larger data we decided to use supervised machine learning methods. From a set of possibly usable methods we have chosen the most suitable one and compared the results of human evaluated sentiment to the sentiment extracted by machine learning during testing of the final hypothesis.

Corpus selection

Because our corpus containing only 280 documents (see Table 1) is really small we considered another approach than usual. We used the whole experimental corpus and its human evaluation for testing (see Table 1 and Table 2) and following external corpora for training:

- Movie review corpus (Cornell University) with 10662 documents annotated as positive / negative (Pang, 2004) and (Pang, 2008). It contains movie reviews that were classified with machine learning methods.

Table 4: Summary of Movie review corpus classes.

Class name	Classes count
negative (-1)	5331
positive (1)	5331
Total	10662

Twitter sentiment corpus (Technische Universität Berlin) with 10786 documents annotated as positive, neutral, negative, and n/a (Narr, 2012). It contains tweets that were human-annotated with sentiment labels by three Mechanical Turk workers.

Table 5: Summary of Twitter sentiment corpus classes.

Class name	Classes count
negative (-1)	1758
neutral (0)	5647
positive (1)	2873
n/a	508
Total	10786

Because the Movie Review corpus contains only positive and negative classes, we needed to eliminate all documents with the neutral class from our test (human evaluated) corpus. For full comparison we needed to keep consistency about classes in all corpora, thus we eliminated neutral and n/a classes from the Twitter sentiment corpus.

Machine Learning Methods

Considering the fact that sentiment was included into text messages through the Twitter social network for which Naïve Bayes and Support Vector Machine learning methods (Annett, 2008) or (Yang, 1999) are mostly used, we decided to use them and compare with more methods for both training corpora and choose the most suitable of them from the following list:

- Decision Trees (DT)
- Random Forrest (RF)
- Naïve Bayes (NB)
- Maximum Entropy (ME)
- Support Vector Machines (SVM)

R language RTextTools (text classification via supervised learning) and text mining packages were used as tools. We used a supervised method and calculated the following measures: precision, recall, F-measure, and accuracy. The results are shown in Table 6 and Table 7.

The most exact methods for Movie review corpus are Random Forest and Support Vector Machine method based on precision, recall, and accuracy values.

Table 6: Metrics for Movie reviews corpus.

[%]	DT	RF	NB	ME	SVM
Precision	10.0	55.5	24.2	54.5	56.5
Recall	50.0	59.0	55.8	56.5	59.5
F-measure	16.5	52.5	33.7	49.0	49.0
Accuracy	20.1	58.7	56.0	52.9	51.7
Order	5	1	4	3	2

Table 7: Metrics for Twitter sentiment corpus.

[%]	DT	RF	NB	ME	SVM
Precision	40.0	62.0	22.9	55.0	58.0
Recall	50.0	62.5	57.7	56.0	59.5
F-measure	44.5	62.0	32.8	54.5	58.0
Accuracy	20.1	74.9	52.5	65.6	70.3
Order	5	1	4	3	2

The most exact methods for Twitter sentiment corpus are also Random Forest and Support Vector Machine method based on precision, recall, and accuracy values.

From previous comparison are chosen Random Forest and Support Vector Machine prediction used with Twitter sentiment corpus. And for better decision about the final method and its prediction results used in further analysis we have compared cross tables in Table 8 and Table 9.

Table 8: SVM cross table.

		predicted		
		negative (-1)	positive (1)	
evaluated	negative (-1)	22	30	52
	positive (1)	47	160	207
		69	190	259

Table 9: Random Forest cross table.

		predicted		
		negative (-1)	positive (1)	
evaluated	negative (-1)	22	30	52
	positive (1)	35	172	207
		57	202	259

Thanks to its much better measures and slightly better cross table results is chosen predicted sentiment with Random Forest method.

4 ANALYSIS

In the final analysis we were interested in trends in link between the number of steps walked by the participant and his sentiment expressed by tweets. We did not need to care about measurement uncertainty and errors, because it should not influence trends during the time period in which the experiment was done.

The previous steps were necessary to get basis for the future analysis and research.

Data used in the further analysis are coming from human evaluated sentiment and from machine learning extracted sentiment, both in combination with measured steps.

4.1 Hypothesis Definition

According to (Sibold, 2009) pilot study and other research (Anderson, 2010) we defined the following null and alternative hypothesis:

- H_0: Movement does not affect mood.

$$H_0: \mu_{negative} = \mu_{positive} \tag{5}$$

- H_A: Movement results in a positive mood.

$$H_A: \mu_{negative} < \mu_{positive} \tag{6}$$

Then we were looking for a pattern that could be described as: there are more steps with positive sentiment than with negative sentiment.

4.2 Data Overview

The variables of interest for the analysis are:

- Sentiment: categorical ordinal and explanatory variable
- Steps: continuous numeric and response variable

4.2.1 Data Pre-processing

Steps are obtained with higher granularity with approximately one minute. But the sentiment expressed via tweets is recorded only several times a day. This needs to be aligned together through aggregation of steps.

According to the alternative hypothesis “Movement results in a positive mood” which could be also interpreted as: “Expression of sentiment is a result of movement”, all preceding steps were aggregated to the nearest sentiment (see Figure 2).

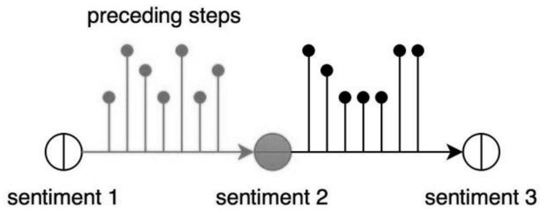

Figure 2: The whole preceding time window of steps is aggregated to following sentiment.

4.3 Hypothesis Testing

4.3.1 Exploratory Data Analysis

An interesting output of exploratory data analysis is the mean value for both sentiment classes and density plots (Table 10, Figure 3, Table 11, and Figure 4).

Table 10: Statistical results for human sentiment evaluation.

	Sentiment classes	
	negative (-1)	positive (1)
N	52	207
mean (μ)	304.13	380.56
sd	727.43	630.99

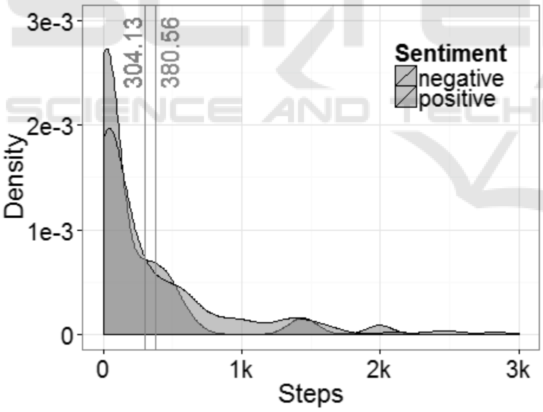

Figure 3: Statistical results for human sentiment evaluation.

Table 11: Statistical results for machine learning sentiment extraction.

	Sentiment classes	
	negative (-1)	positive (1)
N	57	202
mean (μ)	292.79	385.65
sd	364.01	710.42

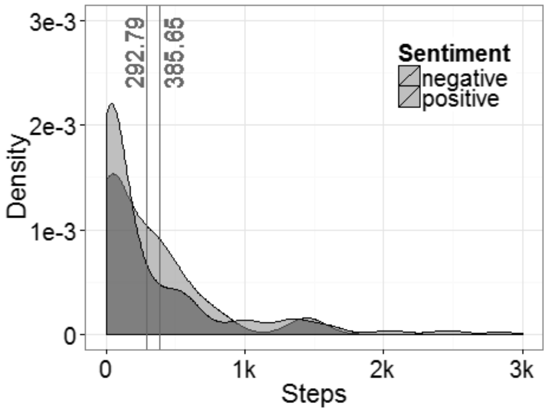

Figure 4: Statistical results for machine learning sentiment extraction.

The results from exploratory data analysis shown that alternative hypothesis (6) is valid for both human sentiment evaluation and machine learning sentiment extraction. This observed assumption needs to be confirmed through the following statistical inference. We will look if the significance of difference between means of steps assigned to positive and negative sentiment is big enough to confirm alternative hypothesis (6).

4.3.2 Statistical Inference

When testing a hypothesis with a categorical explanatory variable with two levels and a quantitative response variable, the paired t-test is used for statistical inference. As a value of statistical significance we have taken α = 0.05. The results are presented in Table 12.

Table 12: t-test for equality of means.

		Sentiment eval. methods	
		Human	Machine learning
t		-0.70	-1.34
df		71.47	182.37
mean difference		-76.43	-92.86
se difference		96.44	-346.41
95% C.I.	lower	-295.73	-229.89
	upper	142.88	44.16
p		0.4894	0.1828

Looking at the results it can be seen that true difference in means is not equal to 0, but probability p is higher than statistical significance. Thus we need to accept the null hypothesis H_0 and reject the alternative hypothesis H_A.

5 CONCLUSIONS

5.1 Bias

We are fully aware that the experiment has several biases. However, we considered them (and also the options which we had at the time) when we had designed the pilot experiment.

At first, walking is related to need to move somewhere. When you do not plan to have an active day, then it is hard to produce steps. Distribution of activity during the day is not even.

Secondly, expression of sentiment is subjective as well as its evaluation. The questions how to express sentiment and extract it are discussed in many publications cited in this paper. The subject mostly described his activity and current feelings as hate or pleasure.

Lastly, but not less importantly, we need to consider conditions for tweet writing. If the aim was to write 20 tweets a day, we had effectively 16 hours of active day (assuming 8 hours of sleep). Thus the subject was required to produce a tweet expressing sentiment every 48 minutes. This was little bit pushy and also affected content and even willingness to express mood.

5.2 Corpora

Selection of a right corpus plays a significant role in future analysis. As we can see in Section 3.3.2, there was an observational difference in measures using machine learning methods for two selected corpora.

Better results for the Twitter sentiment corpus might be given by the fact that we also used tweets in our experiment. The limit of 140 characters can lead to short expressions of what subject wants to say. As the result, the words in the Twitter corpus could be more similar than the words in the Movie review corpus.

Another fact is that whilst the Movie review corpus contains both positive and negative values equally, the Twitter sentiment corpus has 38% tweets with negative sentiment and 62% tweets with positive sentiment. This proportion is closer to the distribution of sentiment classes in our experimental corpus which is 20% tweets for negative sentiment and 80% tweets for positive sentiment.

5.3 Hypothesis Testing Results

Summarizing all the results from Table 12 we can see that in both cases we can accept the null hypothesis H_0: “Movement does not affect mood” and reject the alternative hypothesis H_A: “Movement results in a positive mood”. We did not found a link between the quantified value (the number of steps) and the sentiment evaluated or extracted from the text in this pilot experiment.

5.4 Result and Improvements

5.4.1 Result Summary

It is obvious that the result we got it is not corresponding to what we expected, namely to confirm the alternative hypothesis. There could be many reasons for that as it is listed in previous conclusion subchapters including the typical issue with a relative small population for statistical analysis.

Nevertheless, we still see the potential to explore the issues like data source suitability, machine learning method evaluation, and exploratory data analysis. Also with improved experiment design and larger population sample we are more likely to reach the goal.

5.4.2 Research Proposal Follow-ups

During the past 15 months when the pilot experiment has been conducted, IoT devices suitable for this kind of experiment increased their capabilities and offer not only steps as a quantified measure but also a continuously measured heart rate with nearly ECG precision and more precise algorithms for steps measurement.

We believe that the heart rate measured independently of current human activity, but together with steps and in the same timeline with sentiment can give us better answers.

ACKNOWLEDGEMENTS

This work was supported by the UWB grant SGS-2013-039 Methods and Applications of Bio- and Medical Informatics and by the project LO1506 of the Czech Ministry of Education, Youth and Sports.

REFERENCES

Balahur, A., Mihalcea, R., Montoyo, A., 2014, *Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications*, Computer Speech and Language, Volume 28, Issue 1, 1-6.

Cambria, E., Schuller, B., Xia, Y., Havasi, C., 2013. *New Avenues in Opinion Mining and Sentiment Analysis*, IEEE Intelligent Systems, Volume 28, no. 2, 15 – 21.

Go, A., Bhayani, R., Huang, L., 2009. *Twitter sentiment classification using distant supervision*, Processing.

Chanel, G., Kronegg, J., Grandjean D., Pun,T., 2006. *Emotion assessment: arousal evaluation using EEG's and peripheral physiological signals*, Proceedings of the 2006 international conference on Multimedia Content Representation, Classification and Security, September 11-13, 2006, Istanbul, Turkey.

Pang, B., Lee, L., 2004, *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*, In Proceedings of the ACL.

Pang, B., Lee, L., 2008, *Opinion Mining and Sentiment Analysis*. Now Publishers Inc.

Ikonomakis, M., Kotsiantis, S., Tampakas, V., 2005, *Text Classification Using Machine Learning Techniques*, WSEAS Transactions on Computers, Vol. 4, Issue 8.

Sibold, J., Berg, K., 2009, *Mood enhancement persists for up to 12 hours following aerobic exercise: a pilot study*, American College of Sports Medicine.

Agarwal, B., Mittal, N., Bansal, P., Garg, S., 2015, *Sentiment Analysis Using Common-Sense and Context Information*, Hindawi Publishing Corporation, Computational Intelligence and Neuroscience, Article ID 715730.

Kim, H. D., Ganesan, K. A., Sondhi, P., Zhai, Ch., 2011, *Comprehensive Review of Opinion Summarization.*

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M., 2011, *Lexicon-Based Methods for Sentiment Analysis*, Journal Computational Linguistics archive Volume 37 Issue 2.

Annett, M., Kondrak, G., 2008, *A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs*, 21st Conference of the Canadian Society for Computational Studies of Intelligence, pp 25-35.

Collomb, A., Costea, C., Joyeux, D., Hasan, O., Brunie, L., 2014, *A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation*.

Yang, Y., Liu, X., 1999, *A re-examining text categorisation methods.*Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp 42-49.

Narr, S., Hülfenhaus M., Albayrak S., 2012, *Language-Independent Twitter Sentiment Analysis*, In Knowledge Discovery and Machine Learning (KDML), LWA.

Anderson, R., J., Brice, S., 2010, *The mood-enhancing benefits of exercise: Memory biases augment the effect*, Psychology of Sport and Exercise.

Horlings, R., Datcu, D., Rothkrantz, L. J. M., 2008, *Emotion recognition using brain activity*, Gabrovo, ACM.