

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/325923927>

Sentiment analysis in the food sector: the Greek case

Conference Paper · June 2018

CITATIONS

2

READS

236

2 authors:



[Theodore Tsiligiridis](#)

Agricultural University of Athens

77 PUBLICATIONS 968 CITATIONS

[SEE PROFILE](#)



[Anastasios Liapakis](#)

Agricultural University of Athens

8 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Fruit Fly Net - European Project [View project](#)



Neural Networks Applications [View project](#)

Sentiment Analysis in the Food Sector: The Greek Case

Anastasios Liapakis

Agricultural University of Athens
Dept. of Agricultural Economics and Rural Development
75, Iera Odos, Athens

Theodore Tsiligiridis

Agricultural University of Athens
Dept. of Agricultural Economics and Rural Development
75, Iera Odos, Athens

Abstract

Sentiment Analysis is a computational method aiming to extract opinions/evaluations of individuals for an entity, such as a product, a service, etc. It can be used to draw users' opinions from all social media networks and other web sources. Existing research has focused mainly on the recognition of English characters, while to our knowledge, limited research papers have been published so far concerning the Greek language, concentrating mainly on the banking and financial sector, neglecting contributions on food industry. Sentiment analysis in social media and other online sources, as for example is the case food websites, is able to identify all possible words (pre-processing stage) that can be used to express the feelings of a user for a specific entity and convert words giving them positive or negative significance by considering the characteristics of these media, such as time sensitivity, text size limitation and unstructured expressions. For the analysis, a set of linguistic, statistical and machine learning techniques are used to structure the information contained in text sources. The main purpose of this paper is to analyze the customers' reviews of ten leader companies in the Greek food industry as they are uploaded in their facebook and tweeter social media accounts. The most common words that used in previous reviews could help us to construct a mini vocabulary in Greek language which can be accomplished an appropriate software to transfer them in a five-dimensional vector. In this point, we should take into consideration that the vast majority of online text-based communications ignore the rules of spelling and grammar especially in Greek language (Greeklish), and therefore, this study aims to improve the calculation of a sentiment score in Greeklish idioms. As it appears, the findings will be practical and influential in the examined companies for further analysis.

KEYWORDS

Sentiment analysis, Greek language, Food industry, Social Media

1. INTRODUCTION

Social Media networks are usually changed consumers' behavior. Nowadays, more and more companies use the social media marketing in order to attract more customers. This, modify consumers' attitudes and companies cannot detect these modifications due to the big volume and the diversity of the produced information. According to surveys in American consumers [2, 6], 81,0% of the Internet users have done online research on a product at least once, the vast majority of questioned are willing to pay from 20,0% to 99,0% more for a five-star-rated item than a low-star-rated item and interestingly, Hitlin and Rainie [13] report that "Individuals who have rated something online are also more skeptical of the information that is available on the Web." In the above context Social Media Networks offer different type of services and could be categorized according the following table [5]:

Table 1 Categorization of Social Media Networks

Category	Platforms
Blogging	Blogger, LiveJournal, WordPress
Micro Blogging	Twitter, Google Buzz
Opinion and Reviews	ePinions, Yelp
Question Answering	Yahoo! Answers, Baidu Zhidao
Media Sharing	Flickr, Youtube
Social Networking	Facebook, LinkedIn, MySpace

Obviously, the variety of these networks creates multimedia data such as text, audio, images and videos files that companies face difficulties to edit or sort.

The problem can be approached with the sentiment analysis method. Sentiment analysis is contextual mining of text which identifies and extracts information from online reviews, web sites and social media networks. It helps businesses to understand the social sentiment of their brand, products or services while was examining online conversations. To facilitate sentiment analysis by means of easier processing and storage in databases, an

opinion/view of a user about an entity can be analyzed as a five-dimensional vector of the form: Opinion (Entity Name, Entity Part, Polarity, User, Date).

1.1 Related Surveys

For the purpose of this paper, we examined eight articles as present in Table 2. The first column indicates the methodology adopted by each article and it is divided into two categories ([10]), namely, the Unsupervised Learning Methods (methods that perform classifications based on some fixed syntactic phrases which are likely to be used to express opinions) and the Supervised Learning Methods (methods that should be readily applied to sentiment classification such as naive Bayesian and support vector machines). The second column specifies the language that analyzed in each article. The third column determines the sources that can be applied by each of the proposed method. The sources could be online Reviews, Forecasts etc. The fourth column specifies the part of speech that can be detected, especially nouns, pronouns, adjectives, verbs, participles, and adverbs. The fifth column shows the industry used by the corresponding article and the final column specifies the examined studies. We should note that all the following articles use the binary polarity (positive and negative).

Table 2 Related articles

Methods Used	Language	Sources	Part of Speech	Examined Industry	References
Unsupervised	English	Reviews	Adjectives & adverbs	N/A	[10]
Super/Unsuper-vised	English	Forecasts	All Parts	N/A	[8]
Supervised	English	Restaurant Reviews	All Parts	Fast Food	[9]
Supervised	English	Product Reputations	All Parts	Mobile Phones	[11]
Unsupervised	English	Reviews	Adjectives & adverbs	Movies	[14]
Supervised	Greek	Reviews	All Parts	-	[7]
Super/Unsuper-vised	Greek	Reviews	All Parts	-	[1]
Supervised	Greek	Reviews	All Parts	Movies & Tech.	[4]

The main result that comes out from the above summary is that existing research has focused mainly on the recognition of English characters and that very limited research papers on the Greek language have been published so far, focusing mainly on the technology and movie sector while there was a limited work that focus on food industry and on greeklish idioms too.

2. DESIGN OF GREEKLISH SENTIMENT LEXICON

Greek is a typical high inflection language due to its complex grammatical and syntax rules. It is a particularly challenging language for Neuro-linguistic programming and specifically for sentiment analysis. For instance, according to Giatsoglou et al. [4], in English, there are only four (4) forms of the regular verb ask (ask, asks, asked, asking), while there are 93 different forms corresponding to the regular Greek verb “ρωτώ”. In order to suggest a greeklish sentiment lexicon with opinion words (words that are usually used to express sentiments) for the Greek food industry, we examined a hundred thirty-one (131) customer reviews of ten leading Greek companies, as these reviews were uploaded in companies' facebook and tweeter social accounts. For a proper training set, the reviews were manually checked. The small volume of the examined reviews comes as a result of the General Data Protection Regulation (GDPR). This regulation may cause problems in the mining of published online reviews from companies' social media accounts. We found the most used positive adjectives and we present them in the following table:

Table 3 Proposed Greek adjectives

Adjective (Greek)	Possible Similar Writing	Frequency Positive meaning	Relative Frequency Positive meaning	Frequency Negative meaning	Relative Frequency Negative meaning
Kalo (Good)	-	37,00	0,13	3,00	0,14
Oraio (Nice)	Oreo	40,00	0,15	-	-
Nostimo (Tasty)	Nosthmo, Nostymo, Nosteimo, Nostoimo	18,00	0,07	-	-
Teleio (Perfect)	Telio	4,00	0,01	-	-
Filiko (Friendly)	-	5,00	0,02	4,00	0,18
Agapimeno (Favourite)	-	6,00	0,02	-	-
Exairetika(Brilliant)	Exeretiko	26,00	0,09	-	-
Iperocho (Fabulous)	Iperoxo, Yperoxo, Yperoho	6,00	0,02	-	-
Exypiretiko	Exypiretiko	50,00	0,18	6,00	0,27
Oikonomiko (Cheap)	Ikonomiko	10,00	0,04	3,00	0,14
Omorfo (Nice)	-	40,00	0,15	2,00	0,09
Thaumasio (Terrific)	Thaymasio	5,00	0,02	-	-
Katapliktiko (Exceptional)	-	5,00	0,02	-	-
Grigoro (Fast)	Grhgoro	8,00	0,03	2,00	0,09
Evgeniko (Polite)	Ebgeniko	15,00	0,05	2,00	0,09
Total		275,00		22,00	

The most used word is the adjective Exypiretiko which means helpful. However, we should note that this adjective has also negative meaning if for example precedes the word not (δεν) as for example in the two following examined reviews:

1. *The restaurant's staff was very helpful (To prosopiko tou estiatoriou itan poly exypiretiko), Positive meaning*
2. *The restaurant's staff was not so helpful (To prosopiko tou estiatoriou den itan poly exypiretiko), Negative meaning*

In the case of Greek language, there are many adjectives whose semantic orientations depend on contexts in which they appear (fifth column). This problem can be resolved with the negation rules. The negation words or phrases usually reverses the opinion expresses in a sentence such as no, not and never [3]. The basic rule that is applied for negations is: Negation Positive transforms to negative (eg: Not helpful/ Oxi exypiretiko). The problem will be extensively analyzed in a future work.

Another obstacle with the Greek language is that the vast majority of online text-based communications ignore the rules of spelling and grammar. After the analysis of the examined reviews, are identified alternative ways to write the same adjective (column 2, table 3). For instance, the adjective tasty was presented with the following five different ways: Nostimo, Nosthmo, Nostymo, Nosteimo, Nostoimo. This is an outcome of the complexity of the Greek alphabet. Table 4, cites an equivalent standard for transcription of Greek letters into Latin letters [12] and into greeklish from the examined reviews:

Table 4 Transcription of greeklish into Latin letters

Greek Letters	Latin Letters	greeklish
Αι, αι	Ee	Ai/ai, Ee
Αυ/αυ	Av/av, Au/au	Av, Au, Ay, Af
Ββ	Vv	Bb, Vv
Γγ	Gg, Yy	Gg, Yy
Εε	Ee	Ee
Ει, ει	Ii	Ii, Ei/ei
Ευ, ευ	Ev/ev, Eu/eu	Ev, Eu, Ey, Ef
Ηη	Ii	Ii, Hh
Ιι	Ii	Ii
Κκ	Kk, Cc	Kk, Cc
Ξξ	Xx	Xx, Jj
Οο	Oo	Oo
Οι, οι	Oi, oi	Oi, oi
Ου, ου	Ou, ou	Ou, Oy
Υυ	Ii	Yy, Ii
Υι, υι	Ii, ii	Ii, Ui, Yi
Φφ	Ff	Ff, Ph/ph
Χχ	Xx	Xx, Ch/ch
Ωω	Oo	Oo, Ww

There are two ways to write the letter e (ε/e in case of greeklish and αι/ai in case of greeklish, since they are pronounced identically in Greek), four ways to write the letter i (ι/i in case of greeklish, η/h in case of greeklish, υ/y or u/u in case of greeklish and ει/ei in case of greeklish, since they are pronounced identically in Greek), two ways to write the letter o (ο/o in case of greeklish, ω/o or ω/w in case of greeklish, since they are pronounced identically in Greek) and two ways to write the letter v (β/b in case of greeklish and β/v in case of greeklish).

According to the above analysis, the transcription of Greek letters into Latin letters (greeklish) enables us to use the previous lexicon in any recommended program that has already used in sentiment analysis in English language. The following procedure should be ideal. Firstly, the greeklish lexicon should be saved in a .csv file (greeklexicon.csv). The alternative ways to write the same adjective must be registered as different (column 1 and 2, Table 3). Each row of the first column has to contain an adjective. Secondly, the data set of the examined reviews should be saved also as a .csv file (examinedreviews.csv). Each row has to contain a review that has to be written with Latin letters (Table 4). In order to avoid the double binary semantic orientation, all neutral and negative sentiments should be removed. Fourthly, from each row in the file examinedreviews.csv the frequencies of each adjective have to be computed. Finally, the reviews which contain one of the adjectives in the greeklexicon.csv should be saved in a new file as follow: (Company's Name, Adjective, Positive, User, and Date of review).

3. CONCLUSIONS

Social media marketing helps companies to improve their brands. Due to the large and continuous volume of data that produced, companies face difficulties to identify all the necessary information. The problem can be approached with the method of sentiment analysis. While a lot has been written and researched about sentiment analysis in various languages, Greek has not drawn researcher's attention. After the analysis of a hundred thirty-one customer reviews of ten leader companies in Greek food sector, a mini vocabulary of fifteen positive adjectives and a transcription of Greek into Latin letters (and greeklish idioms) were suggested. Future work will focus, primarily, on the extension of this study regarding all the parts of speech of Greek language (verbs, nouns, adverbs) and secondly, on the restrictions as presented in section 2 (negation rules and spelling mistakes). Especially, a bigger sample of online reviews will help us to predict all the alternative ways to write Greek words as presented in greeklish idioms in the case of social media networks. Finally, a greeklish lexicon with negative and neutral words will be suggested.

REFERENCES

- [1] Adam Tsakalidis, *Greek Sentiment Lexicon*. Available online: <http://socialsensor.eu/results/datasets/147-greek-sentiment-lexicon>

- [2] comScore/the Kelsey group, "Online consumer-generated reviews have significant impact on offline purchase behavior, Press Release, <http://www.comscore.com/press/release.asp?press=1928>, November 2007.
- [3] Ding, X., Liu, B., & Yu, P. S., 2008, *A holistic lexicon-based approach to opinion mining*. In Proceedings of the 2008 international conference on web search and data mining, Palo Alto, California, USA, pp. 231-240.
- [4] Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C., 2017. Sentiment analysis leveraging emotions and word embeddings, Vol. 69, *Expert Systems with Applications*, pp. 214-224.
- [5] Hu, X., & Liu, H., 2012. *Text analytics in social media*. In *Mining text data*. Springer US, Boston, MA.
- [6] J. A. Horrigan, 2008, "Online shopping," Pew Internet & American Life Project Report.
- [7] Kalamatianos, G., Mallis, D., Symeonidis, S., & Arampatzis, A. 2015, *Sentiment analysis of greek tweets and hashtags using a sentiment lexicon*. In Proceedings of the 19th Panhellenic Conference on Informatics, Athens, Greece, pp. 63-68
- [8] Kang, H., Yoo, S. J., & Han, D., 2012, *Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews*, *Expert Systems with Applications*, Vol. 39, No. 5, pp. 6000-6010.
- [9] Li, N., & Wu, D. D., 2010, *Using text mining and sentiment analysis for online forums hotspot detection and forecast*. *Decision support systems*, Vol. 48, No.2, pp. 354-368.
- [10] Liu, B., 2010, *Sentiment Analysis and Subjectivity*. Handbook of natural language processing, Florida, USA, pp. 627-666.
- [11] Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T., 2002, *Mining product reputations on the web*, In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Edmonton, Canada, pp. 341-349.
- [12] Pedersen, T. T., 2009, *Transliteration of non-Roman scripts*, Retrieved October, 16, 2015.
- [13] P. Hitlin and L. Rainie, "The use of online reputation and rating systems," Pew Internet & American Life Project Memo, October 2004.
- [14] Turney, P. D., 2002, *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*, In Proceedings of the 40th annual meeting on association for computational linguistics, Philadelphia, Pennsylvania, pp. 417-424,