V. N. KARAZIN KHARKIV NATIONAL UNIVERSITY
School of Economics

Svobody Sq. 4, 61022, Kharkiv, Ukraine
Tel.: +38 057 706 13 96
Fax: +38 057 707 51 94

STSM Scientific Report

Cost Action CA15140 - “Improving Applicability of Nature-Inspired Optimisation by Joining Theory and Practice (ImAppNIO)”

Trust Mining for Social Applications

Visiting: Dr. Kateryna Kononova, Faculty of Economics, Karazin Kharkiv National University, Ukraine

Host: Dr. Michail Salampasis, Intelligent Systems Laboratory, Technological Educational Institute of Thessaloniki, Greece

Duration of STSM: Jan. 8-15, 2017

1 Purpose of the STSM

Trust mining (TM) is an application subarea of Sentiment Analysis, which is concerned trust evaluation methods and the estimations of trust that can be done on the information retrieved from Social Media. The aim of the STSM was to make a comparative analysis between opinion retrieval and the application area of trust mining. This aim was detailed in following tasks: (i) Investigation of opinion mining methodology, finding the bridges to the application area of trust mining, (ii) Detection the opinion mining technologies, applicable for trust mining. Discussion the tools, which allow evaluating sentiments beyond the polarity, the ways of trust distinguishing from a set of emotions, (iii) Identification nature-inspired optimization methods, applicable for trust mining, (iv) Discussion the neural networks architectures, suitable for trust recognizing.

The first task was addressed to some extent in (Sherchan 2013). As for second task, the analysis of opinion mining technologies, applicable for trust mining showed that most research has tended to focus on polarity identification in text, rather than using a finer-grained category system for classification (Diamantaras et al. 2016, Mao et al. 2013, Zheludev et al. 2012). Researches, aimed at recognizing emotions in text, indicate that evaluating sentiments beyond the polarity is not a trivial task (Bollen et al. 2010, Lachanski 2014, Zhang et al. 2010). The third task was studied within specific domains, such as statistical (Yu and Singh 2002, Josang et al. 2006, Mui et al. 2002) and machine learning techniques (Vafeiadis et al. 2015, Stalidis et al. 2015), heuristics-based (Xiong and Liu 2004, Huynh et al. 2006) and behavior-based techniques (Adali et al. 2010).

The purpose of this STSM at the Intelligent Systems Laboratory in the Technological Educational Institute of Thessaloniki was to discuss and investigate with members of the laboraratory the adaptations and extensions of methods tools of standard (polarity-based) opinion mining, which will allow to effectively mine the emotion of trust from data retrieved from social media.

V. N. KARAZIN KHARKIV NATIONAL UNIVERSITY
School of Economics

Svobody Sq. 4, 61022, Kharkiv, Ukraine
Tel.: +38 057 706 13 96
Fax: +38 057 707 51 94

2 Description of the work carried out during STSM

During this STSM I concentrated on the development of a pipeline of tools and methods for trust mining in the e-commerce domain taking into account the language specifics of the dataset. Research includes about 20`000 posts and comments in Russian and Ukrainian languages in social networks Vkontakte and Facebook concerning SUN InBev and Target Brands in the period from Jan 1, 2016 to Dec 12, 2016. A fragment of dataset is shown in the table:

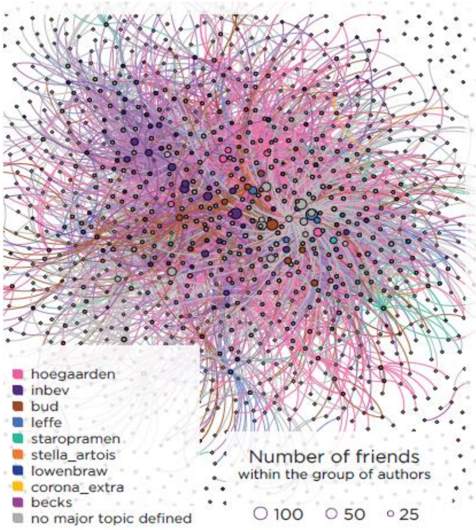
Brand	Search query [no register sensitive]
SUN InBev	inbev, suninbev, abinbev, anheuser-busch, ab inbev, сан ин бев, санинбев, санинбэв, сан ин бэв, sun in bev, сан ін бев, санінбев, сан ін бев, інбев, ін бев, инбев, ин бев, ин бэв, инбэв
Staropramen	staropramen, старопрамен
Leffe	leffe, лёфе, лефе, леффе, лёффе, льофе, льоффе
Hoegaarden	hoegaarden, хюгарден, хугарден
Löwenbräu	löwenbräu, lowenbraw, лёвенброй, ловенбра, лёвенбра, ловенброй, льовенбра
Taller	taller, талер, таллер, талєр, таллєр
Beck's	beck`s, beck's, becks, бекс, бэкс, бэк`с, бек`с,
Bud	bud , бад, budweiser, бадвайзер, Будвайзер, будвейзер, бадвейзер
Stella Artouis	stella artouis, стелла артуа, стела артуа
Corona Extra	corona extra, корона екстра, корона экстра

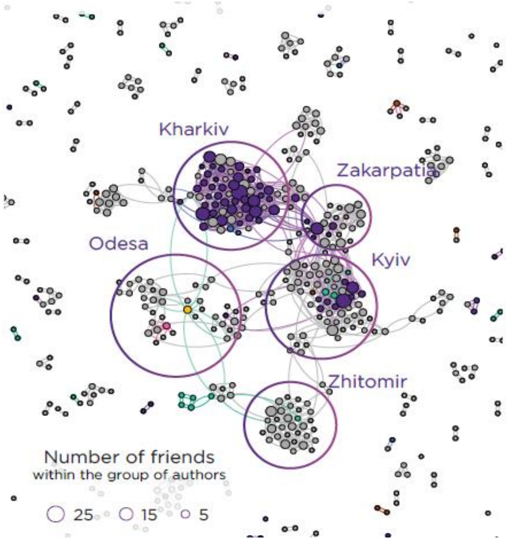
I have explored two main aspects: Profile and Posts analysis

2.1 Profile analysis

Profile analysis, among other possibilities, allows me looking through connections of target audience, and finally spot the actual clusters. This type of analysis allows evaluating the trust which is based on the connections of the social network’s users.

Comparative analysis of social graphs of FB and VK users has shown that FB users are keener to connect via friendship with each other as distinct from VK users. In FB it’s easier to identify clusters related to different brands while VK users are mostly connected within particular regions.





2.2 Posts analysis

Posts analysis allows finding associations with brands to know what exactly brand mean for customers, does it trusted or not. I have identify two types of association, which could be called “right” (which lead to trust, marked blue) and “wrong” (which lead to distrust, marked red).

lowenbraw		**bud**		**staropramen**		**leffe**		**corona_extra**	
franziskaner	0,46	whassup	0,39	хмелить	0,31	сладковатый	0,47	охлаждать	0,26
reinheitsgebot	0,52	клейдесдальский контроль	0,39	ячмень	0,33	горчинка	0,47	алонсо де хэрер	0,31
мюнхен	0,53	alcohol free	0,42	чехия	0,35	монах	0,49	grupo modelo	0,38
gruppe	0,53	конфедерация	0,44	снэк	0,35	blonde	0,49	monde selection	0,38
баварский	0,58	czechvar	0,47	kozel	0,35	brune	0,52	cerveceria	0,4
львиная	0,59	budwar	0,47	молсон корс	0,37	ароматизировать	0,56	мексика	0,42
интербир	0,6	адольфус буш	0,53	чешский	0,38	notre dame	0,62	аскорбиновый	0,43

inbev		**stella_artouis**		**becks**		**hoegaarden**		**taller**	
anheuser busch	0.44	sushibao	0.76	скатываться	0.37	оригинал	0.31	посуда	0.52
енергія	0.43	суши	0.6	robertcarlyle	0.26	штука	0.31	лаконичность	0.5
ЛТК-ИнБев-НПУ	0.43	лакшери	0.41	rumbelle	0.26	stella artois	0.3	силикон	0.48
тур	0.39	стелла	0.4	бедовый	0.26	ящик	0.27	микроволновый	0.44
зоря	0.36	подарок	0.37	воторсон	0.26	самовывоз	0.24	пластичность	0.44
ураган	0.36	вспенить	0.33	детсво	0.26	бира	0.22	экологичность	0.44
титан	0.34	закупорить	0.33	джентельмен	0.26	вайнота	0.22	нержавеющий	0.43

V. N. KARAZIN KHARKIV NATIONAL UNIVERSITY
School of Economics

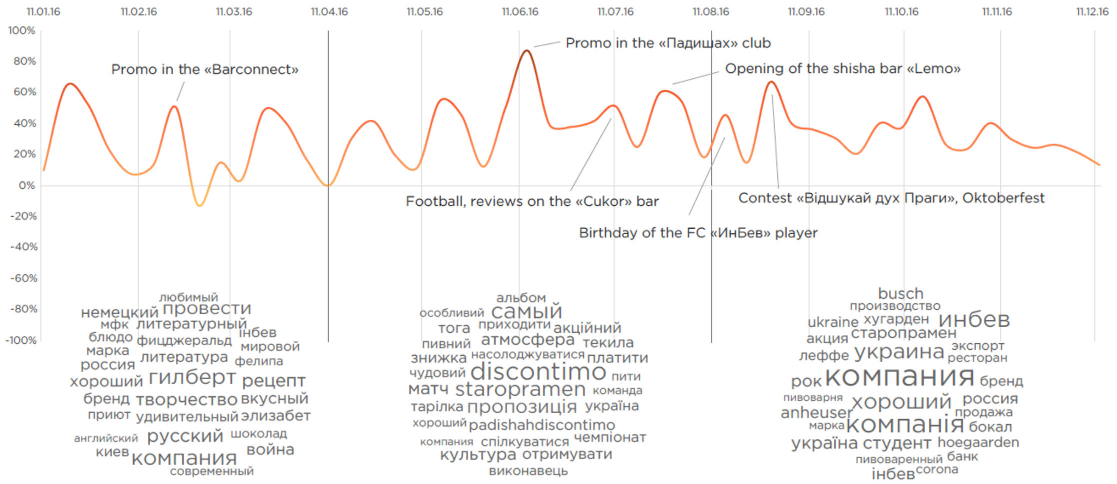
Svobody Sq. 4, 61022, Kharkiv, Ukraine
Tel.: +38 057 706 13 96
Fax: +38 057 707 51 94

Sentiments analysis by regions and target groups, over time was held on lexicon base using SentiStrength. It proved efficiency for estimates the strength of positive and negative sentiment in short English texts, however, its Russian sentiment dictionary after close look turned to be automatically translated English dictionaries and consist of 1700 terms. Furthermore, auxiliary word lists (boosters, negations, questions) were automatically translated too, which makes them nonsense at all. Regarding We've introduced modified tool which employs original SentiStrength algorithms and custom Russian and Ukrainian dictionaries. For Russian and Ukrainian language manual translation and synonyms enrichment for boosters words list were made. Some details of such There are some numbers:

	English	*Russian*	*Ukrainian*
Sentiments dictionary	2546	1320*	1320*
Boosters list	28	166**	215**
Negations list	17	9	8

* Due to huge difference in emotional lexicon (full of idioms), it's impossible to find direct translations for some terms.
** Synonyms enrichment were made.

Sentiment analysis has shown the positive attitude to company at large. Following Dr. Michail Salampasis results (Salampasis et al. 2011) some reasons of sentiment variations were detected. It was shown that HORECA[1] promotions causes positive feedback.



Unfortunately SentiStrength algorithms don't allow go beyond the polarity. Some details of adopting a machine learning approach for training beyond the polarity classification model were discussed in Dr. Kostas Diamantaras paper on Sentiment analysis leveraging emotions and word embeddings (Diamantaras et al. 2016). The development of sentiment analysis algorithm is seen in applying Plutchik Psycho-evolutionary theory of basic emotions (Plutchik 2002), in which trust is a one of the basic emotions (eight primary emotions were considered: anger, fear, sadness, disgust, surprise, anticipation, trust, and joy).

In the discussions with Dr. Salampasis general details of computational model that can go beyond the polarity and allow identifying trust and its grades in textual documents using machine-learning algorithms were identified.

[1] HoReCa is an acronym, which means Hotel, Restaurant, Cafe/Catering

Svobody Sq. 4, 61022, Kharkiv, Ukraine
Tel.: +38 057 706 13 96
Fax: +38 057 707 51 94

2.3 Other

During my STSM with Dr. Salampasis we discussed the problems of identifying negation patterns in different languages as well as detection of irony and stance in texts, we were trying to clarify distinguish between irony, stance and humor.

3 Description of the Main Results Obtained

The main results of the STSM are the following:

a. Investigation of opinion mining methodology in a case of trust mining distinguishing has been done;
b. The tools, which allow evaluating sentiments beyond the polarity, the ways of distinguishing trust from a set of emotions (as machine learning techniques applied to psychometrics datasets) are discussed;
c. Mashing learning algorithms are decided better suited for trust distinguishing in comparison with lexicon based approach;
d. The neural networks architectures, suitable for trust evaluating basing on information retrieved from Social Data are discussed (in particular, the prospects of recurrent networks for emotion recognizing).

4 Future Collaborations with the Host Institution

Together with Dr. Salampasis we plan to continue working on trust mining methodology development.

5 Projected Publications to Result from the STSM

I plan to continue collaboration with the Intelligent Systems Laboratory of Technological Educational Institute of Thessaloniki and prepare the publication on Trust mining in e-com domain, which is planned to be submitted to forthcoming HAICTA conference and then printed in Springer edition.

6 Confirmation by the Host Institute of the Successful Execution of the Mission

Please see the attachment.

7 References

Adali, S., Escriva, R., Goldberg, M. K., Hayvanovych, M., Magdon-Ismail, M., Szymanski, B. K., Wallace, W. A., and Williams, G. (2010). Measuring behavioral trust in social networks. In Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI'10), 150–152.

Bollen J., Mao H., Zeng X. (2010). Twitter mood predicts the stock market. Retrieved from: https://arxiv.org/PS_cache/arxiv/pdf/1010/1010.3003v1.pdf

V. N. KARAZIN KHARKIV NATIONAL UNIVERSITY
School of Economics

Svobody Sq. 4, 61022, Kharkiv, Ukraine
Tel.: +38 057 706 13 96
Fax: +38 057 707 51 94

Giatsoglou M., Vozalis M., Diamantaras K., Vakali A., Sarigiannidis, G. (2016). Sentiment analysis leveraging emotions and word embeddings. Expert Systems with Applications 69, 214-224

Huynh, T. D., Jennings, N. R., and Shadbolt, N. R. (2006). Certified reputation: How an agent can trust a stranger. In Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'06). New York, NY, 1217–1224.

Josang, A., Hayward, R., and Pope, S. (2006). Trust network analysis with subjective logic. In Proceedings of the 29th Australasian Computer Science Conference (ACSC'06). Australian Computer Society, Hobart, Australia, 85–94.

Lachanski M. (2014). Did Twitter "Calm"-ness Really Predict the DJIA? Retrieved from: http://files.meetup.com/7616132/DC-NLP-2015-07%20Michael%20Lachanski.pdf

Mao Y., Wei W., Wang B. (2013). Twitter Volume Spikes: Analysis and Application in Stock Trading. Retrieved from: http://nlab.engr.uconn.edu/papers/SNAKDD025.pdf

Mui, L., Mohtashemi, M., and Halberstadt, A. (2002). A computational model of trust and reputation. In Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS'02). IEEE Computer Society, Los Alamitos, CA, 188–196.

Plakandaras V., Papadimitriou, T., Gogas P., Diamantaras, K. (2015). Market sentiment and exchange rate directional forecasting. Algorithmic Finance 4 (1-2), 69-79

Plutchik, R. (2002). Emotions and Life: Perspectives from Psychology, Biology, and Evolution, Washington, DC: American Psychological Association

Salampasis, M., Paltoglou, G., Giahanou, A. (2011). Using Social Media for Continuous Mining of Consumer Behaviour, in Proceedings of the International Conference on Information and Communication Technologies for Sustainable Agri-production and Environment, 193-202

Sherchan, W., Nepal, S., Paris, C. (2013). Survey of Trust in Social Networks. *ACM Computing Surveys*, Vol. 45, No. 4, Article 47

Stalidis, P., Giatsoglou, M., Diamantaras, K., Sarigiannidis, G. (2015). Machine Learning Sentiment Prediction based on Hybrid Document Representation. arXiv preprint arXiv:1511.09107

Vafeiadis, T., Diamantaras, K., Sarigiannidis, G., Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory 55, 1-9

Xiong, L. and Liu, L. (2004). PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities. IEEE Trans. Knowl. Data Engin. 16, 7, 843–857.

Yu, B. and Singh, M. P. (2002). An evidential model of distributed reputation management. In Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'02). ACM Press, New York, 294–301

Zhang X., Fuehres H., Gloor P. (2010). Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". Retrieved from: doi:10.1016/j.sbspro.2011.10.562

Zheludev I., Smith R., Aste T. (2012). When Can Social Media Lead Financial Markets? Retrieved from: http://www.nature.com/articles/srep04213 #ref6Predicting

Kateryna Kononova

13.01.2017