A Query Evaluation Approach using Opinions of Turkish Financial Market Professionals

Bora Uğurlu [1], Esma Yenisarı [1], Bahadır Karasulu [1], Özcan Zafer Ayan [1]

[1]*Canakkale Onsekiz Mart University, Dept. of Computer Engineering, Terzioglu Kampusu 17020, Canakkale, Turkey.*

*Abstract –*People who do not have expertise in the financial area may not see the relationship between the numerical and linguistic data. In our study, a knowledge discovery approach using Turkish natural language processing is recommended in order to respond to meaningful queries and classify them with high accuracy. Query corpus consists of randomly selected unique keywords. Quantitative evaluation is done in order to measure the classification performance. Experimental results indicate that our proposed approach is sufficiently consistent with and able to make categorical classifications correctly. The approach highlights the relationship between numerical and linguistic data obtained from Turkish financial market.

Keywords – Web 2.0, knowledge discovery, financial market, Twitter, really simple syndication.

1. Introduction

Information technology is related to computer-based tools, which aim to process information used by people. Information systems are based on this technology. These systems collect, process, store and analyze information. They provide computer-based support to business intelligence systems (BISs) for making complicated and non-routine decisions. Data used in these systems should have a high level of quality. This means that the data must be accurate, complete, consistent, accessible and concise. Knowledge management process helps a system to manipulate the information. Knowledge management systems use the information technologies, and generally deal with unstructured knowledge. These kinds of systems systematize and enhance the information [1]. Decision support systems are often used as integration for BISs, thus they help decision makers in data mining. An effort aimed to discover the characteristics of high volume data is made through data discovery stage. In this way, a statistical approach is performed. The process of automatically seeking patterns considered as knowledge from data is known as knowledge discovery [2]. Data mining can be considered as an automatic or semi-automatic discovery of knowledge in high volume data. In order to explore meaningful patterns, data must be analyzed. To obtaina deep understanding of data characteristics, a study is performed through either

knowledge or data discovery stage. Each one of these stages mainly depends on a different statistical approach. In the data acquisition stage, information is obtained from document sources and experts, and then converted to a format that a computer can process. The knowledge received at this stage is stored on a knowledge base together with the rules or knowledge presentation models [1].

Web sites based on Web 2.0 enhance users' experiences through features provided by information technologies and enable possible social interactions. Twitter, which is one of these technologies, draws information retrieval researchers' attention on account of its huge data volume. Really Simple Syndication (RSS) technology enables a user to get necessary information without browsing numerous web pages when the user wants any information. Usually, RSS obtains its content from news resources and blogs [3]. Business environment performs its own transactions by considering the social, legal, economic and physical factors. Any changes in these factors put a high pressure on the markets. On the Web 2.0 based resources, professionals in the financial markets share their opinions about this pressure. In this context, it can be seen that there is a strong relationship between economic trends in a specific date range and data obtained from social media and web sites.

The main goal of our study is to develop a user-friendly financial knowledge discovery approachregarded as a query evaluationprocess. This approach's aim is to explain to the user the relationship between linguistic data from resources using Web 2.0 technologies and numerical data obtained by trend analysis of stock markets. The user is regarded as a non-expert in financial markets. In this relationship, the most important factor is the date interval in which the volatility suddenly occurs in the stock markets. Some searches based on Web 2.0 technology cause some situations to emerge. Therefore, one of the aims of the study is to generate an appropriate database. The proposed approach of knowledge discovery is capable of making queries on this database. In addition, it analyzes the query results. In the study, our main aim is to build a useful and consistent approach for knowledge discovery

based on opinions of Turkish financial market professionals.

A short review of studies in the literature which have motivated us isprovided in the related work section. In the results and evaluation section, findings which are extracted from resources based on Twitter and RSS by using our query corpus are presented. These findings depend on four financial professionals' opinions. Assessments based on these findings are given in the results and discussion section. In the last section, a general conclusion about advantages and disadvantages of our approach is highlighted.

2. Related work

Interactions on Web 2.0 web sites form what the community perceives and interprets from various social affairs. Social media networks are used as a base for event extraction and temporal analysis for data mining studied in the literature. In their study, Asur and Huberman [4] showed how social media content may be used to highlight the facts of the real-world. In this frame, they considered the interactions on Twitter about recently released movies in the USA during the three months in 2010. They created a model to make predictions for Hollywood Stock Exchange. In the study, the movie captions, time stamps of relevant tweets, authors and text bodies are used for time series analyses. Bollen et al. [5] used text content obtained from daily Twitter posts in order to measure mood state changes caused by individuals' behaviors in behavioral economy field. They made a comparison through time in series between mood states and Dow Jones Industrial Average (DJIA) values established a correlation between mood states and DJIA. Therefore, the outcomes are used to predict future values of DJIA, and experimental results show that they have high accuracy in forecasting [5].

Lau et al. [6] developed a weakly supervised cybercriminal network mining method for their novel design. This method is based on a statistical model. They applied it to put forward hidden implicit and explicit relations which belong to cybercrime messages in social media data. Drury [7] proposed an approach that uses knowledge. It is able to infer when priceson the stock market will rise and drop. After he created a database composed of several news from certain resources, he extracted useful information from these news items by using natural language processing (NLP)over RSS feeds. To identify specific public affairs, Kallus [8] used public data which were taken from social media and blogs in seven different languages. Especially, he suggested a prediction mechanism which is able to use Twitter data to predict future crowd actions.

Experimental results were used to predict the public turmoil in Egypt (i.e., Coup D'état).

Yılmaz and Yılmaz [9] have defined a novel metric to perform a statistical similarity analysis for scripts of software specifications and requirements, and this metric is also compared with other metrics in the literature. Due to automatic analysis software, the data set is analyzed with developed and existing metrics in the information retrieval research area. Abovementioned studies are related to our study on the scope of Web 2.0 technology and similarity of knowledge discovery methodology.

3. Materials and Methods

Compared to previous works, our study is based on the Turkish language and its content. We implemented our study in five steps. Data are first obtained from the web and then assessedfor the membership degree of related category. In our experiments, our software tool was randomly fed with a pool of 200 tweets and 200 RSSs. Our software tool can implement NLP indexing, scoring and ranking processes. Application programming interface (API) of Twitter is capable of querying 200 tweets on a single transaction. RSS feeds are less frequent than tweets.The main reason is that financial professionals write a column article when a financial affair happens in weekdays.

The search results are obtained as a reference to the match between the keywords in the corpus and classification results. Two independent human experts marked the entries on the corpus that are grouped into four main categories (i.e., ground truth annotation). In the study, four main categories are determined. They are given as "döviz" (i.e., foreign currency), "altın" (i.e., gold), "borsa" (i.e., stock market), and "petrol" (i.e., oil).When a datum does not match with any category above, this datum is classified into "ekonomi" category (i.e., economy).

Frequency changes in RSS feeds belonging to four financial professionals are weekly periods from November 5, 2012 to March 5, 2014. However, due to instantaneous and frequent updates in Twitter resources, time interval of our Twitter data is from January 13, 2014 to March 10, 2014. A data set is formed by browsing the Twitter accounts and websites which publish the opinions of four different professionals. These professionals were randomly chosen from the financial market field.

We gathered the same number of Twitter and RSS feeds from those professionals and extracted information from them for testing. 85 out of the 200 tweets and RSSs were chosen by considering their releases on Internet. While acquiring the data, we applied the steps as shown in Figure 1.
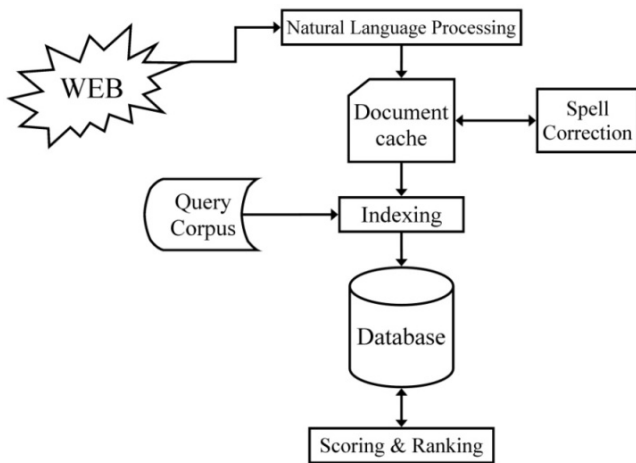
Figure 1. The data flow chart of proposed approach

According to the steps in Figure 1, the generated inputs from the data set is parsed using NLP into wordy phrases through application program interface (API) called NZemberek [10].

Spell correction process is performed by the system for the items deemed necessary. Thus, the relevant data set is passed between a cache of documents and spell correction sub-stages. Data set is stored as a "document cache". This corpus consists of some keywords which are distributed in the four categories.

On the basis of this "corpus", query-process is done, and the results of the spell-corrected data and searches are passed on to an indexing step. In this step, an indexing process is done on the query results and the data in the "document cache" is separated according to the date and categorical identifications (IDs). Then, the indexing results and their dataset are stored to the database. The results are passed from database to the scoring and ranking stage, and vice versa. In this step, the most suitable match for related categories are obtained, and also, the related results are reflected to the database. At the end, a re-ordered database is obtained by using the proposed approach. In Table 1, 182 words located in the query corpus are shown.

Table 1. 182 Turkish keywords related to the categories in query corpus

Ekonomi (i.e., economy)
Ayı, Boğa, para, cari, ticaret, faiz, ihracat, ithalat, fiyat, piyasa, maliyet, kredi, milyon, milyar, satış, alış, yatırım, sat, al, TL, lira, likitide, oynaklık, TEFE, TÜFE, portföy, resesyon, büyüme, varlık, plan, hedge, yatırımcı, liberizasyon, fiyatlama, pricing, finansman, temettü, trend, bekle ve gör, borç tavanı, kırılgan, toparlanma, trilyon, günlük, artış, global, yerli, konjonktürel, vadeli, stok, kriz, hedef, yatırım aracı, kaldıraç, kırılma, rezerv, veri, zam, istikrar, enflasyon, yıllık, senelik, arttırım, sanayi, istihdam.

Borsa (i.e., stock market)
BIST100, BIST, BIST30, borsa, IMKB, bin, IMKB100, endeks, tahvil, FED, ECB, banka, bono, risk, sermaye, finans, açığa, pozisyon, broker, trade, forward, short, long, VOP, VIOP, kontrat, future, spot, analiz, türev, fon, bant, bandı, kurum, kırdı, kırıldı, kontratı, esnek.

Döviz (i.e., foreign currency)
Döviz, Dolar, Euro, Sterlin, Kron, Frang, USD, EU, GBP, NOK, CHF, YEN, kur, TCMB, hazine, borç, deflasyon, banknot, çapraz, devalüasyon, dolarizasyon, efektif, ihale, müdahale, geçirgenlik, repo, arbitraj.

Altın (i.e., gold)
Altın, ayar, ons, gram, gr, külçe, ziynet, cumhuriyet, çeyrek, ata, yarım, tam, darphane, kuyumcu, hurda, ankese, bilezik, emtia, gümüş, gremse, filo, savaş, asker, ordu, yaptırım, darboğaz, bakır.

Petrol (i.e., oil)
LPG, doğalgaz, doğalgaza, yakıt, benzin, dizel, motorin, CNG, NYMX, otogaz, petrol, varil, brent, ham petrol, akaryakıt, petkim, petrokimya, mazot, enerji, jeopolitik, geopolitics, bölgesel, yenilenebilir, enerji arzı, enerji sektörü.

4. Results and Discussion

There are various binary classifier approaches in the literature to measure the performance of search results, in the context of information retrieval, to conduct an evaluation of the retrieval sets, Precision, Recall, F-measure and Accuracy [11] metrics are frequently used. In this respect, we used well-known accuracy metric (i.e., Balanced accuracy) in our quantitative evaluation as an indicator. The harmonic mean of Precision and Recall is calculated as F-measure.

The ground truth annotation is based on the marking made by two independent human experts. Performance measures reflect the quality of query results by considering the range between 0 and 1 value. In this range, the 0 value shows the worst case, while the 1 value represents the best case for the performance. The query corpus consists of selected keywords from [12]. The results obtained by using the query corpus are shown in Table 2. In Table 3 and Table 4, quantitative evaluations of the performance measurements of the proposed approach are presented via four selected categories.

As it can be seen from these tables, the proposed approach is evaluated on the basis of search results. The total averages are calculated from performance metrics for each category. The measurements are based on the search results of Twitter and RSS.

Table 2.Statistical assessment of query result (N=85)

Categories	RSS	Percentage	Twitter	Percentage
Borsa	35	41.17%	20	23.52%
Döviz	22	25.88%	10	11.76%
Altın	17	20.00%	9	10.58%
Petrol	0	0.0%	4	4.70%
Ekonomi	11	12.94%	42	49.41%

Table 3.Quantitative evaluation based on Precision and Recall metrics

Metrics	Precision		Mean	Recall		Mean
Category	RSS	Twitter		RSS	Twitter	
Borsa	0.76	0.28	0.52	1.00	1.00	1.00
Döviz	0.81	0.90	0.85	0.69	0.75	0.72
Altın	0.5	1.00	0.75	0.5	0.62	0.56
Petrol	0.0	0.75	0.37	0.0	0.75	0.37
Average	0.62			0.66		

Table 4.Quantitative evaluation based on F-measure and Balanced accuracy metrics.

Metrics	F-measure		Mean	Balanced accuracy		Mean
Category	RSS	Twitter		RSS	Twitter	
Borsa	0.86	0.44	0.65	0.79	0.94	0.86
Döviz	0.75	0.81	0.78	0.68	0.81	0.74
Altın	0.5	0.76	0.63	0.63	0.75	0.69
Petrol	0.0	0.75	0.37	0.45	0.73	0.59
Average	0.60			0.72		

According to the comparison based on the metrics, the best values of Precision and F-measure where obtained in "döviz" category (i.e., its mean Precision value is 0.85, and its mean F-measure value is 0.78), the best values of Recall and Balanced accuracy where obtained in "borsa" category (i.e., its mean recall value is 1.00, and its mean balanced accuracy value is 0.86), respectively. Related categories are generated by taking financial professionals' weekly columns of RSS feeds. In the scope of our study, there is not any match for "petrol" category on the query corpus which is generated by considering 16 months period of RSS feeds.

Consequently, the proposed approach implemented via software tool cannot identify the financial situation of "petrol" category. The main reason is that the underlying algorithm is not able to classify the "petrol" category from fetched RSS feeds, as well. When same query corpus and professionals are considered in our experiments, the algorithm is able to classify other categories more accurately due to more frequent matches.

In the light of the performance results, one can probably conclude that there are differences among the Web 2.0 technologies (i.e., RSS and Twitter) used to gather data based on the opinions of financial market professionals. In addition, the interaction

between these technologies and performance result of each category is also important. The F-measure is a measure of a test's accuracy, which can be interpreted as a weighted average of Precision and Recall. Therefore, this metric can be used as a good indicator of above mentioned interactions in our study. As a statistical viewpoint on the results, by using two-way analysis of variance, the main effects and interactions of the factors (i.e., RSS or Twitter) are shown in Figure 2 and 3.

When the lines denoting group (i.e., RSS or Twitter) in a two factor interaction plot means do not move parallel to one another, it is likely that there is an interaction between given factors in the plot. Otherwise, it suggests that there is fixed difference between given groups at all levels of the other factor, which means that there is no interaction between these factors. In Figure 2, mean of values show the total order of overall results based on each category's F-measure value. The trends of each category based on relevant Web 2.0 technology (i.e., RSS or Twitter as a factor) and an interaction between these two factors are clearly seen in Figure 3.
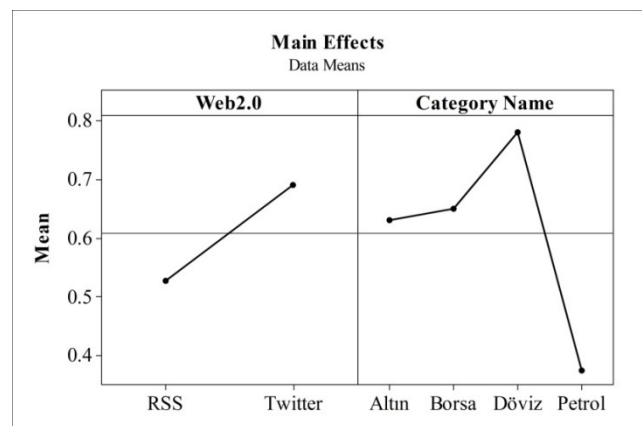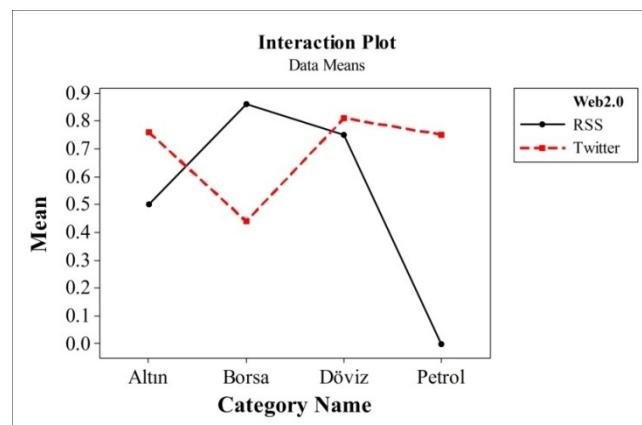

Figure 2.The main effects.

Figure 3.The interaction of metrics.

5. Conclusion

The main contribution of this study is to develop a knowledge discovery approach for querying Turkish financial market professionals' opinions. The overall number of watched phrases encountered in RSS feeds is more than that of tweets. The main reason for this is that the content of the RSS resource is more formal than Twitter resource. There is not any match in the query corpus which is generated by considering 16 months period of RSS feeds for "petrol" category in weekly columns of professionals.

Consequently, the proposed approach implemented via our software tool cannot identify the financial situation of "petrol" category. The main reason for this is that the underlying algorithm is not able to categorize the phrases in the RSS feeds as "petrol". When the same query corpus and professionals are considered, our algorithm can classify phrases more accurately for other categories.

According to the comparison based on the performance metrics, best performance results are obtained for "döviz" and "borsa" categories. While the number of financial professionals and of tweets and RSS feeds is not sufficient for a large number of categories, it is large enough for limited number of categories to build an expository model for knowledge discovery. In the future, a case-base can be created for case-based reasoning by using our proposed approach. Therefore, the approach will be used by the researchers in different sectors for similar situations. In our later studies, we plan to add new techniques and more raw data for further knowledge discovery.

References

[1]. Rainer, R.K., &Cegielski, C.G. (2013). Introduction to Information Systems. International Student Version, Fourth Ed., John Wiley & Sons, Singapore,ISBN: 978-1-118-09230-9.

[2]. Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques, 703p., ISBN:978-0-12-381479-1.

[3]. Hammerley, B. (2003)Contents Syndication with RSS,O'Reilly Media Publisher,224p., ISBN:978-0-596-00383-8.

[4]. Asur, S., &Huberman, B.A. (2010).Predicting the Future with Social Media, 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

[5]. Bollen, J., Mao H., &Zeng, X. (2010).Twitter Mood Predicts the Stock Market, Journal of Computational Science, 2(1),1-8.

[6]. Lau R.Y.K., Xia, Y., & Ye, Y. (2014). A Probabilistic Generative Model for Mining Cybercriminal Networks from Online Social Media, IEEE Computational Intelligence Magazine, .31-43.

[7]. Drury, B. (2013). A Text Mining System for Evaluating the Stock Market's Response to News, PhD. Dissertation, University of Porto, Faculty of Science, Doctorate Dissertation, 162p..

[8]. Kallus, N. (2014).Predicting Crowd Behaviour with Big Public Data, International World Wide Web Conference Committee (IW3C2), April 7-11, Seoul, Korea.

[9]. Yılmaz, İ.B., &Yılmaz, A.E. (2010). Similarity Analysis Method for Structured Statements and Sentences in Turkish. SIU 2010-IEEE 18. SinyalİşlemeveİletişimUygulamalarıKurultayı, Diyarbakır, Turkey.

[10].NZemberek Turkish Natural Language Processing. Net API. (2014). Retrieved March 13, 2014 from https://code.google.com/p/nzemberek/.

[11].Manning, C.D., Raghavan, P., &Schütze, H. (2009). Introduction to Information Retrieval, Cambridge University Press, New York,NY, 2009, Draft Online Copy (2009.04.01). Retrieved March 13, 2014 from http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf.

[12].Ugurlu, B., Yenisari, E., Karasulu B., Ayan, O.Z. (2014). An Evaluation about the Queries for Opinions of Financial Market Professionals using Resources based on the Web 2.0 Technology, Program & Abstract Booklet CD-ROM(USB) of IMCW2014 5th International Symposium on Information Management in a Changing World, November 24-26, Antalya, Turkey.