УДК 004.8

Avdeeva T.G.

undergraduate student, group IIS-Tg11

Institute of Service and Business (branch) DSTU in Shakhty

Alekseenko Ju.V.

undergraduate student, group IIS-Tg11

Institute of Service and Business (branch) DSTU in Shakhty

Lyashov M.V.

Candidate of Engineering Sciences, Researcher

Institute of Service and Business (branch) DSTU in Shakhty

Scientific adviser: Dmitrienko N. A.

Associate Professor Institute of Service and Business

(branch) of Don state technical university in Shakhty

RESEARCH OF INTELLECTUAL ANALYSIS AND MONITOR ALGORITHMS FOR SOCIAL NETWORKS

Summary:

The article presents the rating of the most popular social networks in 2017 in Russia. Methods and algorithms of intellectual monitoring and analysis of social networks are considered. The most used text data analysis methods are presented. The Section 3 describes the application of systems for the monitoring and analysis of social networks.

Keywords: social networks, machine learning, multimedia, intellectual analysis, monitoring.

Introduction

Lately the popularity of social networks has greatly increased. One of the main social networks advantages is providing users with a convenient and fast way of exchanging information. Also, social networks reflect both the structure and dynamics of modern society and the interaction of Internet generation with technologies or other people. The importance of this actively developing research area is undoubtedly confirmed by many new technologies and applications, for

example, online services and common interest communities, multimedia communication via the Internet, multimedia search, interactive and entertainment services, health and safety applications.

Monitoring social networks will allow scientists learn public opinion which is an important indicator of the state of existed socio-economic and political systems, because it reflects the degree of social tension. Social networks have become a new source of opinions and people's perspectives and motives as many users share their opinions on a purchase or event.

1. Review of popular social networks in Russia

Before analyzing social networks it is important to review the most popular networks where users prefer to share their opinions. Social networks are mainly used as an indicator of social communication source, that is why monitoring such networks will help to see how public opinion is being formed.

According to Cossa website, social networks take the first place in terms of the number of users and posts [4].

Social Media in Russia (May, 2017)	
The number of authors every month 38 million	**The number of messages every month** 670 million
Activity in social media by source type (thousands of messages)	

Source	%	Messages
Social Networks	70,2%	470 737
Twitter	11,7%	78 372
Video	10,8%	72 443
Forums	3,6%	25 387
Feedbacks	1,8%	11 907
Blogs	0,6%	5 289
Mass Media Publications	0,6%	4 040
Mass Media Comments	0,4%	2 432

Figure 1 – Activity by Source Type

Also, all social networks can be ranked according to the number of authors and public messages (Figure 2) [5]:

- First place was estimated by means of the number of active authors as by their public messages in May, 2017. It was determined that VK.com is supposed to be one of the most popular social networks with 25 722 users who sent more than 310 795 messages that month.

- Instagram takes the second place by the number of active users – in May there were about 7 143. Otherwise, the number of messages in Instagram takes the third position (around 71 733 messages)

- Twitter is on the second place if to consider the number of messages – 78 372 in May. Otherwise, Twitter is only the fourth one in terms of active authors – 1 171.

- The third place by the number of authors takes Facebook – 1 953, as the number of messages takes the fourth place – 53 413.

Figure 2 – Authors and Messages in Social Networks

2. Social Network Intellectual Analysis Methods and Algorithms

There are two main methods of monitoring social networks: manual and automated.

Manual monitoring presents the analysis of user messages (some resources, blogs, pages, communities, user walls) in order to find the reference to a person, brand name, company, and etc. Usually such type of monitoring includes search engines where there is a useful feature called key words (but not always these words are needed). Yandex Blogs services are one of the ways to analyze social networks manually [3].

Automated monitoring is based on special services for automatic monitoring services (IQbuzz , Babkee, Wobot, and etc.). These services can help to analyze blogs, forums or social networks (Facebook, Twitter, VK, LiveJournal, LiveInternet, YouTube) to find client reviews, competitor monitoring, etc.

Automated monitoring provides the opportunity:

- to reduce labor costs by automating routine processes;
- to achieve the high level of accuracy due to the better systematization of data and use of a large number of analytical tools.

Besides authors can review a conceptual model of monitoring social networks by means of intellectual sentiment analysis as a tool of automated intellectual processing of natural language data. This model consists of several main steps [6]:

– determining the topic of monitoring, specifying the spectrum of social network agents (based on the number of subscribers) or keywords;

– automated monitoring of publications (including all related data, such as likes, reposts, retweets, and comments on these publications);

– filtering extracted data of deleting meaningless messages and excluding them from further analysis;

- analyzing messages stored in the database (after the filtering step);
- providing the analysis results.

The structural scheme of this process is shown in Figure 3. The central core of the monitoring and analyzing system is a database of messages to provide the ability to store millions of extracted messages. It is shown in the structural scheme that the monitoring element (monitor) is an intermediate element between different social networks (their APIs) and the DB of messages. The purpose of this element is to make an API survey periodically according to the set of p^{net} query parameters of each social network. The directions of arrows show the way how some data is transferred (for example, user's opinion presented as text publications). Thus, data collected from social networks is being sent to the monitoring element, and later to the database of messages. The data should not be returned back from the database.
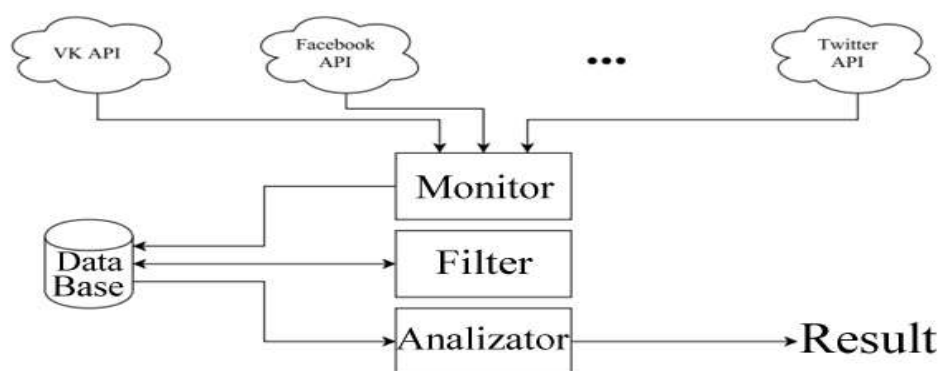

Figure 3 – The Process of Monitoring and Analysis of Text Data in Social Media

2.1 Review of methods for text data analysis

Now there are two often used approaches to solve the problem of text data analysis: the one is based on machine learning algorithms, and the other one is based on special dictionaries of tonality.
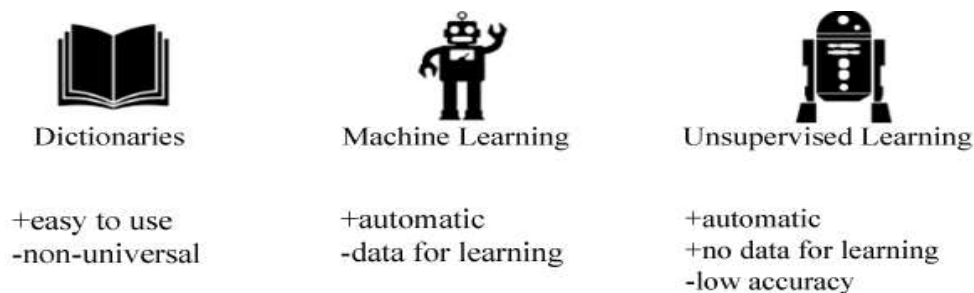

Figure 4 – Approaches for Text Data Analysis

The approach based on using such dictionaries is to analyze individual words (terms) in the text and then classify the text according to the topic. Usually experts use dictionaries, which were prepared specially by linguists, and/or some linguistic rules for the classification of text data [1].

In the approach based on machine learning algorithms, the analysis represents the task of classifying texts, which can be solved by training the classifier on a pre-labeled collection of texts.

Each of these described approaches has some advantages and disadvantages. Now there are methods based on the use of dictionaries avoiding the step of manual text labelling. Moreover, these methods do not need to compile a training function, and also, the "solutions" taken by the classifier can be easily explained. In this case, pre-labeled dictionaries are required which must consider the subject area of the text. In contrast, no dictionaries are used in any machine learning algorithms. As practice shows, classifiers demonstrate high quality of classification. Moreover, the quality of classification can be improved by selecting characteristics (features) of classification and correctly selected by means of combinations in text documents typical for the training sample. Similarly, one cannot ignore that a classifier trained on texts of one subject area is able to cope with texts from completely another subject area.

3. The application of systems for the analysis and monitoring of social networks

The idea of analyzing and monitoring social networks can find its theoretical and practical application in many areas, for example:

1. The application of analysis and monitoring of social networks can help to teach some linguistic rules constructed on a natural language, and educate people to perceive the language at a level close to the human level. Previously, the text was perceived by a machine in an abstract form, and then as the set of letters and content (meaning).

2. The quality of machine translation can significantly be improved by analyzing and monitoring text in social networks. The professional translation is considered to be the standard. To teach a machine work as a professional translator is possible only if to use the same techniques as any translator uses while translating a text., you can only take into account everything a professional translator uses when translating this or that text.

3. The system of analysis and monitoring of social networks can be used in the field of information security, for example, detecting terrorists or tracking malicious activity.

4. The analysis and monitoring of social networks is very popular in online commerce activity. Monitoring allows experts constantly monitor the information on websites, to understand the motives and trends in society, and receive feedback on a particular product.

Conclusion

The authors stated some popular approaches used for monitoring social networks. Also, this article analyses the advantages and disadvantages of each approach.

Methods and algorithms for intellectual analysis and monitoring of social networks are described, as well as the application areas of this technology.

Used sources:

1. Open source intelligence, Available at: http://ru.wikipedia.org/wiki/OSINT (Accessed: 20.03.2018).

2. Vyugin V.V. Matematicheskie osnovy teorii mashinnogo obucheniya i prognozirovaniya [Mathematical Basis of Machine Learning Theory and Forecasting]. MTsNMO Publ., 2013, 387 p..

3. Social network analysis software, Available at: http://en.wikipedia.org/wiki/ Social_network_analysis_software (Accessed: 11.04.2018).

4. E. Novruzova. Brand Analytics. Social networks in Russia, 2017: numbers and trends, Available at: http://www.cossa.ru/289/166387/ (Accessed: 16.04.2018).

5. Adamic L., Glance N. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog II Proc. of the 3rd ACM international workshop on Link discovery, 2005. - P. 36—43

6. Domingos, P. & Pazzani, M. On the optimality of the simple Bayesian classifier under zero-one loss // Machine Learning. - 1997. - № 29. – C. 103-137.