Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 190 (2021) 137–141

Procedia Computer Science

www.elsevier.com/locate/procedia

2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society

Methods for identifying an information object in social networks

Andrey Cherkasskiy[a]*, Alexey Artamonov[a,b], Marina Cherkasskaya[b], Natalia Leonova[a]

[a]*National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe hwy, 31, Moscow, 115409, Russia*

[b]*Plekhanov Russian University of Economics, Stremyanny lane, 36 , Moscow, 117997, Russia*

Abstract

Social networks are a unique phenomenon in which a large amount of unstructured information about various users is collected. The collected data can be used to identify different groups of users for the purpose of delivering targeted information to them. The article discusses the issues of building models of thematic groups of users based on multi-criteria assessment and using agent technologies of information collection and processing. The implementation of this method expands the possibilities of social research and the formation of thematic user groups for monitoring and analyzing situations in various areas of human activity. The proposed concept has shown its effectiveness on the training and control sample of objects, which makes it possible to predict the effectiveness of the use of agent technologies for scanning information resources of social media.

Keywords: multi-agent systems; agent; social network; wave scanning; statistical analysis of training sample

1. Introduction

Over the past few years, the volume of information generated increases every year not at times, but by an order of magnitude. To solve the problems of processing large and extra-large amounts of data, fundamentally new algorithms, called Big Data, are being developed. The complexity of processing is not only in scope, but also in the diversity of data [1, 2].

* Corresponding author. Tel.: +7-963-651-83-16.
E-mail address: aicherkasskij@mephi.ru

1877-0509
10.1016/j.procs.2021.06.017

Currently, Big Data technologies are actively used, including for marketing purposes, to classify and identify the target audience. One of the promising tasks is the classification of users of social networks for the delivery of specialized content to them.

The complexity of the task is due to the need to process all types of data - numerical characteristics, text, audio, video images, relationships between objects, since at the moment the actual text content gives way to audio and video content. With the advent of these types of data, the task is to develop new approaches to processing and analyzing information objects [3].

The amount of data is extremely large - there are more than 400 different social networks in the world, in which over four billion users are registered [4]. The most popular foreign social networks are Facebook [5] (USA), Twitter (USA), Tencent Qzone (PRC), Sina Weibo (PRC).

The peculiarity of social networks in the Internet environment is that the objects are registered users (people) and groups of people united by a particular thematic community of interests. This feature of social networks also determines the characteristics of the requirements for agent (and any other) technology used in social networks.

In thematically oriented systems, it is necessary to use special route methods for scanning information resources, since continuous scanning of tens and hundreds of millions of information sources causes the Big Data problem. If the objective conditions of the study are long-term observation of the behavior of social objects and, at the same time, it is necessary to record many values of their characteristics in various scales of their measurements, then the labor intensity of information and analytical work increases sharply to a physically unrealizable level. It is for solving such a class of information and analytical problems that it is advisable to use agent technologies [6, 7].

Agent technology refers to the regular targeted automatic search of thematic information in a given cluster of Internet sites. To achieve the goals, agents can interact with each other and with the passive medium they form, forming multi-agent systems. Each agent of such a system has its own ideas about the outside world, the tasks and logic that determine its behavior. In the process of work, agents communicate with each other [8].

2. Methodology for monitoring and analysis of social networks

The purpose of the work is to develop methods for training intelligent search agents to independently solve the problems of recognizing target objects when scanning a social network.

The solution to this problem enhances the possibilities of social research as well as the possibility of forming thematic groups of users for monitoring and analysis of situations in various areas of human activity.

The solution search consists of two stages:

- Generation of training sample of target objects
- Determination of the boundary values of the target state of the object

At the initial stage, a list of profiles is compiled that corresponds to the target objects - these can be already known objects or objects that have some unique, explicitly formalized characteristic. As an initial condition, it is necessary to refer to the actual characteristics of the objects adopted in this social network, since it is not known in advance, what goals the user-researcher will set for themselves. The methodology of "training agents by examples" proposed by the researcher user as a training sample (inductive method) is selected [9]. The task of this stage is to identify implicit characteristics of the object that can attribute it to the target group.

Each object in the training set is described by dozens of standard characteristics adopted in the social network. For example, the Vkontakte social network uses more than 30 characteristics from a variety of categories (main, contacts, activities, interests, life position). The characteristics of objects in social networks themselves are described by almost all conceivable types of data: real numbers (for example, date of birth), scores with various scales for assessing qualitative characteristics (for example, beauty, complexity, cultural level, etc.), symbols, then there are some images that have a certain meaning (for example, state emblems, flags, religious symbols, etc.), texts with a certain thematic content, defined using thesauri with weighted concepts, video and audio objects. Moreover, the user is not required to indicate the values of all their characteristics when registering in the network. Many of these characteristics are irrelevant to the targeted search, so experts are tasked with excluding them from further consideration. As a result of this action, only objects with target characteristics remain in the training sample.

However, the significance of each characteristic for successful target recognition is different. This distinction must be presented in quantitative form. The characteristics should be weighted according to their importance in the target description. This problem is solved by the method of collective expert ranking of characteristics with subsequent conversion of collective rank estimates into normalized from 0 to 1 weight coefficients. That is, characteristics that have one or another physical dimension are converted into relative dimensionless quantities. The conversion method depends on the type of data used to measure the characteristic on the social network.

In general, based on the results of the first stage, a table is obtained with a list of characteristics and their maximum values.

Table 1. Weighted components of the target search vector

Item No	Criterion	The weight
1	Criterion1	W_1
2	Criterion 2	W_2
3	Criterion 3	W_3
…	…	…
N	Criterion N	W_N
		$\sum_{i=1}^{N} W_i = 1$

The second stage involves the determination of the boundary values of the target state of the object.

The results of solving the previous problems allow calculating the value of the integral criterion of the target correspondence of each object in the training sample. Suppose that *m* experts are involved in the formation of the training sample and as a result of their work a L_0 training sample is created:

$$L_0 = \sum_{i=1}^{m} L_i \tag{1}$$

Where, L_i is the number of target objects represented by the *i*-th expert in the training set.

The integral criterion of the target correspondence of an object is defined by the following expression:

$$E_i = \sum_{j=1}^{N} W_j * v_{ij} \tag{2}$$

Where, W_j is the weight (relative significance) values of each *j*-th characteristic;

v_{ij} is relative importance of the *j*-th characteristic of the target search vector;

N is number of criteria.

Since it is known that all the objects of the training sample are target, therefore, if the values of the integral criterion of the new object lie in the range from E_{min} to E_{max} calculated from the training sample, then the object is target. E_{min}, which delimits target and non-target objects, will be called the target object marker.

Thus, the rule of agent recognition of target objects is obtained: if the value of the integral criterion of the target correspondence of the new object is greater than E_{min}, then it is marked as a target object and included in the agent collection of target objects.

In agent processing of social network object flow, target objects are marked with a conditional code and received in the database of the agent system for further processing of agent collections of target objects and solving various application tasks on the generated set of target objects.

To implement the above concept, the development of two agents is necessary: search and diagnostic. To minimize the time of an agent-based search in a social network, a search agent “downloads” network information without a detailed analysis of its relevance. The filtering of noise in the collection of data brought by the search agent is carried out by the diagnostician agent, which works only in the internal environment of the multi-agent information-analytical system.

After the formation of the starting content of the agent-based search system, they proceed to the computational experiment. The essence of the experiment consists in presenting objects from the control sample to the trained intelligent search agent. During the experiment, the correct and erroneous decisions of the agent on the belonging of

the object to the target or non-target group are recorded. The completeness and accuracy of agent collections obtained as a result of the experiment determines the quality of the search problem solution by a trained intelligent agent.

To determine the target interval of the target object marker values, the training sample of 48 profiles was loaded into the system, and the minimum marker value was fixed – 0.1429, the maximum recorded value – 0.588. After tagging objects, the obtained data were manually analyzed.

Of the 48 users of the training sample, 42 users entered the target interval indicated above, and the weights of 6 users were below the minimum value. To test the system, in addition to the initial training sample of 48 people, a control sample was created from the profiles of users of the VKontakte social network. The control sample included 20 profiles of knowingly target objects and 38 profiles of randomly selected knowingly non-target objects. A total of 58 users were included in the control sample. Control sample data were uploaded to the system and all users included in the sample were tagged. After marking the objects, the obtained data were uploaded from the system in the form of an Excel spreadsheet and analyzed.

When evaluating the accuracy of marking of agent search objects, the authors revealed that the accuracy of the system according to the training sample was 87.5%. The accuracy of the system for the control sample of target objects was 76.92%, and the accuracy of the system for the control sample of non-target users was 84.2%.

3. Conclusions

The proposed methodology showed effectiveness in the control sample, but further research is needed in several thematic areas and real data in order to determine the operability of the methodology for analyzing incoming objects in streaming mode.

The effectiveness of agent technologies can be expected when solving a wide class of information and analytical tasks for processing data streams in social networks; therefore, the authors plan to introduce and pilot-test the developed methods for solving various applied problems.

References

[1] Kulik, Sergey. (2015) “Neural network model of artificial intelligence for handwriting recognition.” *Journal of Theoretical and Applied Information Technology*. Volume **73**, (**2**): 202 – 211.

[2] Kulik, Sergey. (2015) “Model for evaluating the effectiveness of search operations.” *Journal of ICT Research and Applications (ITB Journal of Information and Communication Technology)*. Volume **9**, (**2**):177 – 196.

[3] Galkin, Timofei, Popov, Dmitry, Pilyugin, Victor and Grigorieva, Maria. (2019) “The visualization method pipeline for the application to dynamic data analysis.” Conference Paper 27th Symposium on Nuclear Electronics and Computing, NEC 2019; Budva; Montenegro. Volume **2507**: 295 – 299.

[4] Sergeeva, Yulia. (2021) “All internet statistics for 2020 - numbers and trends in the world and in Russia”. Retrieved February 03, 2021, from https://www.web-canape.ru/business/internet-2020-globalnaya-statistika-i-trendy/

[5] Perrin, Andrew, and Anderson, Monica. (2019) “Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018”. *Pew Research Center*. Retrieved from: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

[6] Artamonov, Alexey, Ionkina, Christina, Kirichenko, Alexander, Lopatina, Catherine, Tretyakov, Evgeniy and Cherkasskiy, Andrey. (2018) “Agent-based search in social networks.” *International Journal of Civil Engineering and Technology*. Volume **9**, (**13**): 28 – 35.

[7] Onykiy, Boris, Antonov, Evgeniy, Artamonov, Alexey and Tretyakov, Evgeniy. (2020) ‘Information Analysis Support for Decision-Making in Scientific and Technological Development.” *International Journal of Technology.* Volume **11**, (**6**): 1125 – 1135.

[8] Artamonov, Alexey, Leonov, Dmitry, Nikolaev, Vadim, Onykiy, Boris, Pronicheva, Larisa, Sokolina, Ksenia, Ushmarov, Ilya. (2014) “Visualization of semantic relations in multi-agent systems”. *Scientific visualization*. Volume **6**, (**3**): 68 – 76.

[9] “Inductive learning algorithm”. (2019). Retrieved February 18, 2021, from: http://espressocode.top/inductive-learning-algorithm/