RESOURCE APPROACH TO DATA MINING BASED ON NETWORK TRAFFIC

Vochkov D.B., Bakanov A.S., Tashev T.D.

***Abstract**: In the presented paper is considered resource approach to intellectual data analyses based on network traffic. A model for calculation of loading on different segments of the network in the distributed systems for organization control is designed. This allows more effective execution of data searching and analysis procedure and also significant reduction of the time for data processing.*

***Keywords**: Intellectual data analyses, network traffic, resource approach.*

***ACM Classification Keywords**: I.2.1 Applications and Expert Systems*

Introduction

In the presented paper is considered resource approach to intellectual data analyses based on network traffic. The term resource in our context includes information as well as telecommunication component. A data mining and information evaluating approach is proposed based on preliminary quantity evaluation of the resource using combined criteria with followed by multicriterial analyses.

Resource approach to intellectual data analyses

Intellectual data analysis problems in the organization control systems are considered in details in the work of [Баканова Н.Б., 2007], [Bakanova N., Atanasova T., 2008], [Петровский А.Б., 2004], and also other authors. In the presented paper intellectual data analysis is applied to the resource. In term resource we will understand information as well as telecommunication component. One of the most important problem while creating distributed information systems for organization control is information interaction with far apart departments and sub-administration authorities. In accordance with the specifics of the application problems of the organization control operative data exchange between the distributed network nodes is required and stable interactive mode execution for control activity support is demanded. In the organization control components of the information flows are documents that can be estimated using a row of criteria in accordance of a chosen scale. As criteria we will consider the following: date of the creation of the document/date of the document changing, number of readings (editings), number of document authors, document size etc. so it is possible to organize combined resources using a row of criteria in accordance of one or another scale. The existence of a systematic structure capable of analyzing data using multicriteria will allow more effective execution of the procedure of data searching and analyzing and also to significantly reduce the time needed for data processing.

Quantified estimation of the network traffic

For characteristics determination is proposed to use real data which are collected in subsystems for monitoring implemented in the life-supporting complex of application systems for organization control. The size of the data required depends on the object type and the request conditions which means that data size will be calculated individually for every single request. The size of the information exchange will depend on the stage of document processing (data about workers, execution, control, additional images etc.). The data size concerning the results

of request searches and communiqués is determined by the quantity of the chosen records and the set of the returned fields according to the criteria chosen. The data size, when a request is directed to an element of a reference book, depends on the request context (the set of meanings available for the user to choose, set of fields dependent on the object from which the reference book is called out etc.) Generally object data are stored in different fields of several tables. Let object data are stored in M-tables. In every j-th table object data are distributed in Kj-attributes. Then object data size (Qj) situated in j-th table and object data size d can be determined using the equations in [Волчков Д.В. и др., 2012]. The so obtained size of the requested data is recorded in monitoring Data Base at the moment when the object is called out.

The rate of channel utilization is determined by the part of the general time during which data transfer is provided through the channel. To determine if the channel's data transfer ability for traffic transmission using the current level of documentation load we have as a first step to calculate the channel loading rate in the time interval (t_0; T), when the number of call outs is maximum. Source data are information about access time to objects of Information system and the size of requested data. This information is available from the monitoring Data Base.

Let for the observation time interval (t_0; T) are fixed call outs to N-objects of Information system.

From the monitoring Data Base are known access time to objects $t_1, t_2, \dots, t_N$ and size of the transmitted data $d_1, d_2, \dots, d_N$. In order to simplify the calculations we will accept that object data will be chosen from queue with length w and transmitted bit by bit. Current loading rate can be determined using the equations in [Волчков Д.В. и др., 2012]. If we accept the required loading rate ρ, using the dichotomy method we can calculate the magnitude of the recommended throughout ability of the channel utilizing data about documentation loading

Based on the so described approach a program packet for calculation of the recommended network characteristics is developed.

Conclusion

As a result of this work a model for calculation of loading on different segments of the network in the distributed systems for organization control is designed. The new element of the proposed approach is that in its foundation lays analysis of monitoring data about the current documentation loading which are collected in a constantly working system. This allows more effective execution of data searching and analysis procedure and also significant reduction of the time for data processing. Based on this approach a prototype of a program packet is designed which collects data about documentation loading, analyses the current loading rate of the network segments and calculates the throughout ability of the channel required for stable functioning of the information system of the organization.

References

[Bakanova N., Atanasova T., 2008] Bakanova N., Atanasova T. Development of the combined method for dataflow system. // IJ "Information technologies & knowledge". 2008, Volume 2, № 3, p.262-266

[Атанасова Т.В. и др., 2010] Атанасова Т.В., Савченко Т.Н., Головина Г.М., Баканов А.С. Интеллектуальная информационная среда обитания и субъективное восприятие качества жизни // Методы исследования психологических структур и их динамики. Труды ИП РАН. М., 2010.

[Баканов А.С., 2009] Баканов А.С. Особенности психологического подхода к моделированию человеко-компьютерного взаимодействия // Вестник ГУУ. 2009. №6. С. 15–18.

[Баканова Н.Б. и др., 2012] Баканова Н.Б., Волчков Д.В. Программный комплекс расчета нагрузки на основе данных мониторинга // Электросвязь. 2012. № 3, С. 54-57.

[Баканова Н.Б., 2007] Баканова Н.Б. Использование программно-технических комплексов для повышения эффективности контроля в системах документооборота, «Электросвязь», 2007 г., №6, стр. 51-53

[Волчков Д.В. и др., 2012] Волчков Д.В., Шарануца В.А. Анализ сетевой нагрузки в распределенной системе организационного управления. Информационные технологии и системы (ИТиС'12): сборник трудов конференции.– М.: ИППИ РАН, 2012. – с. 276-281.

[Журавлев А.Л., 2004] Журавлев А.Л. Психология управленческого взаимодействия. М.: Изд-во «Институт психологии РАН», 2004.

[Ларичев О.И., 1987] Ларичев О.И., Петровский А.Б. Системы поддержки принятия решений. Современное состояние и перспективы развития. // Итоги науки и техники. Серия Техническая кибернетика. М. ВИНИТИ, 1987. т.21, с.131-164.

[Петровский А.Б., 2004] Петровский А.Б. Многокритериальное принятие решений по противоречивыми данным: подход теории мультимножеств. // Информационные технологии и вычислительные системы, 2004, №2, 56-66.

Authors' Information

Dmitry Vladimirovich Volchkov *– Institute for Information Transmission Problems, Bolshoy Karetny per. 19, Moscow, 127994, Russia; e-mail: vol@iitp.ru*

Major Fields of Scientific Research: System Analysis, Software technologies

Arsenii Sergeevich Bakanov *– Ph. D., Institute of Psychology of Russian Academy of Sciences, ul. Yaroslavskaya., Moscow, 129366, Russia; e-mail: arsb2000@pochta.ru*

Major Fields of Scientific Research: Intelligent Systems, Software technologies

Tasho Dimitrov Tashev *– Institute of Information and Communication Technologies - B.A.S., ul. Acad.G.Bonchev bl.2, Sofia 1113, Bulgaria; e-mail: ttashev@iit.bas.bg*

Major Fields of Scientific Research: Distributed Information Systems Design, Networks models