Ontology Model for Telecom Operator Big Data Representation

Globa L.S., Novogrudska R.L.
The National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine
lgloba@its.kpi.ua
rinan@ukr.net

***Abstract*—This paper presents an approach to telecom operator Big Data description using ontology model. Elements of ontology model suggested the description of different data attribute, as well as set semantic correlation of such data. Using them it is possible to organize searching engine and perform logic inference. Special parameters for telecom Big Data are described. The estimation of telecom operator quality based on fuzzy logic is proposed and correlation between it and big data description based on ontology is proposed.**

***Keywords*—telecom operator, Big Data, parameters, ontology, fuzzy knowledge base.**

I. Introduction

At the present development stage of information and communication technologies, the term Big Data means a number of approaches, tools and methods of processing of structured and unstructured vast amounts of data and considerable diversity.

Big Data is a term used to identify data sets. It can't be coped with such data sets using existing methodologies and software because of their large size, rate of arrival, analysis and complexity. Such researchers as M. Hilbernt, S. Strinivasa and others developed methods and software tools for data transmission and information granules mining from Big Data (objects collections formation, that are usually formed for numeric attributes. They are placed side by side because of their similarity, functional or physical commonality), but the appearance of new data formats requires constant expansion and improvement of existing methods and data analysis tools [1].

With that, accumulating data from network nodes rapidly increase every year. This causes necessity in powerful constantly increasing computing resources to increase processing speed and data access. In this regard, necessity to improve large amounts of data processing algorithms becomes more relevance.

According to research were conducted by a number of leading companies in the world [2] telecom operators are faced with the urgent need the complex account of different characteristics of provided services (technical, economic, experience of using services by end-user) due to the rapidly growing range of services and the transition to digital space. They want a clear understanding and process management, that occurring between the operator and its subscribers. This whole range of parameters is too large and complex data for the collecting, processing and analysis with the use of current computing infrastructure and is characterized by:

– significant amount of data (from terabytes to petabytes);

– necessity for high processing speed in real time to reduce the volume of storage;

– heterogeneity (can be structured, unstructured, semi-structured);

– necessity to fulfil a validity requirement (may be disrupted due to the variety of data sources and processing methods, violations of safety requirements);

– importance (using of forecasted methods and analysis methods allows to predict the direction of company's development).

At the same time for assigned tasks data analysis is often performed based on data, which is obtained as a result of the economic operator's activity, or based on sociological interrogations, or based on technical parameters of the operator's infrastructure functioning (for example, the decision to invest in one or another part of the system does not consider the influence and analysis of all possible factors and consequences). For example, if we consider the problem of users' satisfaction degree by services, which are provided by telecom operator, it is quite obvious that the frequent technical failures affect satisfaction degree, price operator's policy and the service performance jointly impact on the final services performance evaluation by subscriber.

Modern facilities of Big Data analysis require a transition from unstructured to structured data, thus forming, relatively speaking, "volume data compression volume to their meaning" and generating data processing strategy for Big Data as "data - information - knowledge - prediction" (Fig. 1.), in this case the entered processing steps are understood as [2]:

– data – streams of raw facts such as business transactions;

– information – clusters of facts that are meaningful to human beings such as making decisions;

– knowledge – data/information organized to convey understanding, experience, accumulated learning, and expertise.

In Fig. 1. it is conditionally shown, that huge amount of data accumulates in time. The posibility of making prediction based on previosly gatherd data (thgougth its trnsformation into information and knowledge) is urgent task nowdays.

Telecommunication companies are investing a lot of money in analytical tools and services development. With such data

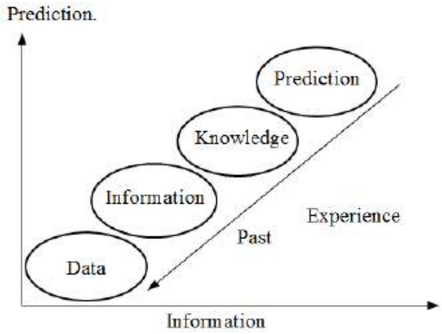

Figure 1. Scheme of the transition from unstructured data to information and to knowledge

telecom operators aim at: increase sales; assure revenue (detect and prevent revenue leakage); reduce churn and fraud; improve risk management; decrease operational cost; improve visibility into core operations, internal processes and market conditions; discern trends and establish forecasts; cross-sell/up-sell products and service plans.

Telecom systems deals with different types of Big Data (*Call Detail Data*, Subscriber (customer) data, Network (operations) data, Others sources data). Another challenge is data scale. As well Big Data influense the proces of telecom analitics management (Fig. 2)

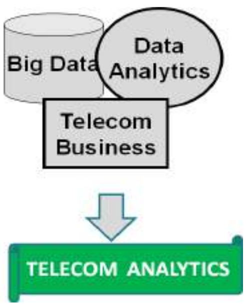

Figure 2. Proces of telecom analitics taking into the account Big Data

Telecom like any other business operates with: customers, services, processes. But, there are unique aspects, such as: network infrastructures, being used as a pipe by other business, regulations. All this shoud be tooked into the account while developing systems for telecom analitics.

Thus, urgent task is to include poverful methods and model of telecom Big Data analises and procesing into general telecom analics system. Main tasks that should be performed are Big Data systematization and stucturing, selection of data groups and identification of semantic connections between them. Such tasks ecxecution allows to gaine more effective data analises and proccesing.

II. Relevant Mathematicals Methods for Big Data

To implement the Big Data collection processes, storage, processing, analysis and forecasting different methods are used, such as: classification [3, 4]; clustering [5, 6]; machine learning [7, 8, 9]; neural networks; support vector machines; decision tree learning; etc.

In [10], it was proposed to consider a fuzzy logic approach that helps to reduce computational complexity in the process of classifying large amounts of telecom data, the efficiency is considered on the example of obtaining a general estimation of quality index of providing services by telecom operator. It is proposed to use decision-making methods based on fuzzy logic. Fuzzy expert rules are formed, which are the basis of the expert system.

Fuzzy rules in such a system can be periodically adjusted to the current status of the technical infrastructure by means of their reformation (refinement) in the process of providing services by telecom operator.

Fuzzy knowledge base (FKB) formation is considered on the example of service quality estimation. Formation algorithm in the generalized form can be presented as follows:

Initial data: measurement tables provided by telecom operators. The measurement table is a set of parameters, which we denote by $X_1 ... X_n$. It should be noted that $X_1 ... X_n$ is set of parameters by which telecom operator estimates general state of the system, but none of them characterizes the system quality cumulatively (Y).

To build FKB it is necessary to split measurement table into 3 samples:

1) Training sample with M_1 rows, where $M_1 = \{1, k\}$, which is needed to form fuzzy logical rules of knowledge base. To form FKB rules it is necessary to determine integral quality index, which is obtained using desirability function, we denote it as Y_D;
2) Test sample with M_2 rows, where $M_2 = \{k + 1, n\}$, which is needed to check fuzzy logical rules quality of knowledge base;
3) Examination sample with M_3 rows, where $M_3 = \{n + 1, m\}$, which is required for the final verification of the correct operation of the obtained FKB.

Generic algorithm for determining an unknown value of Y:

1) Y_D calculation using the desirability function based on M_1 data.
2) FKB formation using M_2 and obtained values of Y_{FKB}. FKB is formed by a set of rules:

$$\text{IF } X_1,\ X_2,\ X_3,\ \ldots,\ \text{THEN } Y\ .$$

Such kind of approach is useful for analyzing measurement data for a number of parameters to circulate in the operator's system but there are a lot of very important factors which have no digital meaning and are able to significantly influence the final result. Taking into account this the paper presents the approach to ontology model of telecom operator Big Data development aiming to optimize the process of such data classification and systematization for its further analytics.

III. Source Data

Source data provide us with different parameters of telecom Big Data. Such parameters were divided on 3 groups according to their meaning and purpose of usage.

First group represents basic notions (objects) that are involved in the communication process:

Provider – deals with information about the service provider.

Service – describes information about the services provided by the provider, services description. It possible to distinguish three types of services: Internet, Television, and Telephony services.

Tariff plan – contains information about the tariff plans of the provider.

Account – describes personal user account information.

User – is directly information about a person who has an account (his / her name, address, etc.).

Bill – describes information about the cash bills associated with this user account. General account deals with three types of actions: get receipt, look through balance, and get invoices.

Second group includes parameters that influence an estimation of quality index of providing services:

– Connection Success Rate – the success of data connections 2G/3G;

– Connection Block Rate – percentage of locks due to overloads 2G/3G;

– Connection Drop Rate – percentage of data connection interruptions 2G/3G;

– PS Attach Success Rate – percentage of successfully completed procedures Attach 2G/3G;

– PDP Context Activation Success Rate – percentage of successful activation procedures PDP Context 2G/3G;

– Speed DL – average speed of HSDPA data transmission to the subscriber;

– Iub Congestion – 3G base station share with high overloads per Iub interface;

– Backhaul Accessibility – availability of zonal transport network;

– DNS Success Rate – successful of DNS resolution;

– DNS Response Latency – DNS resolution time.

Third group represents parameters of the monitoring system are proposed based on the analyses of measurement tables provided by telecom operators:

– COUNT_DAYS_OVER_1MB – the amount of days during which consumed traffic is over 1MB;

– COUNT_DAYS_OVER_5MB – the amount of days during which consumed traffic is over 5MB;

– DUAL_SIM_PROBABILITY – probability of dual sim usage;

– SIM_PRIORITY – the priority of sim card;

– OBLAST – district of sim card usage;

– CITY – city of sim card usage;

– N_SERVICES – amount of services used by permanent subscriber;

– etc.

The consideration of the presented data characteristics leads to the conclusion about their different nature and complexity for the operators to obtain the estimates of their activities. In suggested approach described above data are used for the elements of telecom Big Data ontology definition.

IV. Ontology Model for Telecom Data

The ontology model is developed to realize representation of telecom Big Data. Ontology allows to describe all concepts of problem domain and sets groups of classes, relations of different types, class elements, etc.

Ontology model includes such elements as [11, 12]:

– classes,

– classes attributes,

– relations set on classes,

– types of attributes values,

– class exemplars.

There are two types of relations: the relations between classes and the relations between data types. You can also classify the relations by the following types:

– Internal relations – these are the relations that are inextricably linked with the object;

– External relations are those relations that describe the connection of objects with external objects;

– Relations between instances of this class;

– The relations between instances of different classes from different parts of the hierarchy.

Telecom Big Data ontology classes and their attributes were defined using basic notions that interacts in communication process and source data parameters. Ontology includes next classes: Provider, Service, Tariff Plan, Account, User, Bill, Telephony, Internet, Television, Receipt, Balance, Invoice, Connection Success Rate, Connection Block Rate, Connection Drop Rate, PS Attach Success Rate, PDP Context Activation Success Rate, Speed DL, Iub Congestion, Backhaul Accessibility, DNS Success Rate, DNS Response Latency, SIM_PRIORITY, COUNT_DAYS_OVER_1MB, COUNT_DAYS_OVER_5MB, N_SERVICES, Oblast, City DUAL_SIM_PROBABILITY.

Classes Telephony, Internet and Television are subclasses of class Service. Class Bill is parent class for classes: Receipt, Balance, and Invoice.

Let us describe relations and attributes of basic classes (Provider, Service, Tariff Plan, Account).

The Class ”Provider” is associated with the ”Service” class with the ”has a Service” relation, with the tariff plan class relation ”has a Tariff Plan” and with the ”User” class the relation ”has a relation”. The class has data type relations:

– ”has a name” with the data type string (the name of the ”Provider”);

– ”has a date of creation” with a data type dateTime (the date of ”Provider creation”);

– ”has a description” with the data type string (additional description of ”Provider”).

The ”Service” class has the following data type relations:

– ”named” with string data type (name ”Services”);

– ”has a description” with the data type string (description of ”Services”).

The ”Account” class is related with the ”Account” and ”Tariff Plan” classes in relation to ”Has a Bill” and ”Uses the Tariff Plan” respectively. Also contains the following relations data types:

– ”has login” with data type string (Login ”Account”);
– ”has a password” with the data type string (Account password);
– ”has an email” with the data type string (registered e-mail ”Account”).

The Class ”User” is associated with the ”Account” class with the relation ”has an Account”, with the class ”Service” with respect to ”using the Service” and with the ”Provider” class, the relation ”has a relation”. In the class description, there are the following relations of data types:
– ”has a full name” with string data type (username ”User”);
– ”has passport data” with string data type (Passport data of ”User”);
– ”has address” with data type string (address of ”User”).

Described ontology was developed using one of the tools for semantic modelling – visual editor Protégé 5 [13]. Visual methods of designing ontologies help quickly and fully understand the structure of the subject domain knowledge, which is especially valuable for researchers working in the new subject domain. Protégé 5 supports all phases of the ontology life cycle in accordance with ISO / IEC 15288: 2002 [14] requirements – from the development of a semantic network and the creation of a knowledge base on its basis, to the formation of user requests to these bases in order to obtain knowledge.

The ultimate hierarchical structure of the telecom Big Data ontology was developed by means of the ontologies editor and the framework for constructing knowledge bases Protégé. A component of such ontology is represented on fig. 3.

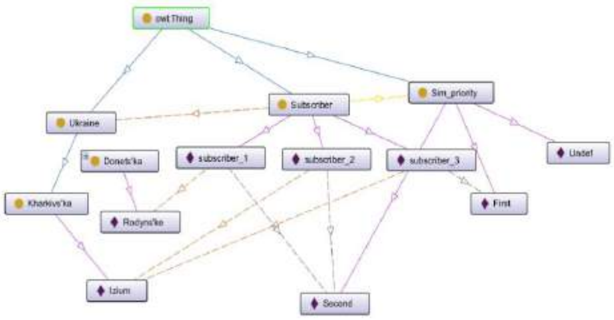

Figure 3. Component of telecom Big Data ontology

V. Conclusions

In the paper the approach to telecom operator Big Data systematizing and structuring is described. It is proposed to use ontology model for this purpose. Different measurement tables provided by telecom operators were analysed. As a result the set of parameters of telecom Big Data were identified. They formed the basis of ontology model.

Selection of Big Data groups and identification of semantic connections between them using ontology model allows to gaine more effective Big Data analises and proccesing.

Further research will be oriented on more detailed telecom Big Data ontology investigation, description of its elements and identification of relationships and regularities. As soon as ontology will be filled with all data the search engine based on logical inference will be tested.

References

[1] Y.Y. Bolyubash, “Methods and tools for processing large data resources in information systems of territorial administration”, Ph.D. dissertation, software for computer systems, Lviv Polytechnic, Lviv, Ukraine, 2017.
[2] M. Ulema. “Big Data and Telecommunications Telecom Analytics” – Tutorial II, BlackSeaCom’2016.
[3] T.O. Ayodele, “Types of Machine Learning Algorithms”, New Advances in Machine Learning. INTECH Open Access Pub., – 2010 – pp. 19–48.
[4] H.A.H. Ibrahim, S.M. Nor, A. Mohammed, A.B. Mohammed, International Journal of Computer Networks and Wireless Communications, – 2012 – 2(1):69–73.
[5] A.K. Jain, M. N. Murty, P. J. Flynn, “ACM Computing Surveys”, – 1999 – 31(3):264–323.
[6] W.A. Barbakh, Y. Wu, C. Fyfe, Studies in Computational Intelligence, Springer, 2009, 249 p.
[7] S. Russell, P. Norvig, Artificial Intelligence: A modern approach, New Jersey, Upper Saddle River, 2010, p. 1154
[8] M. Tim Jones, Artificial Intelligence: A Systems Approach, Hingham, Massachusetts, Infinity Sci. press LLC, 2008, 485 p.
[9] D. Kriesel, A Brief Introduction to Neural Networks. Available at: http://www.dkriesel.com/en/science/neural_networks (accessed 2017, Oct)
[10] Volvach, I.,Globa, L., Method of generalized quality index calculation in mobile networks, Proceedings of 14th International Conference “The Experience of Designing and Application of CAD Systems in Microelectronics”, 2017
[11] Yu. A. Zarogulko, O. I. Borovikova, An approach to constructing knowledge portals, Optoelectronics, Instrumentation and Data Processing, 2008, vol 44, no 1, pp. 75–82
[12] L. Globa, R. Novogrudska, O. Oriekhov Method of heterogeneous information resources structuring and systematizing for Internet portals development, 978-1-4673-2232-4/13/$31.00 ©2013, IEEE, EuroCon 2013, 1-4 July 2013, Zagreb, Croatia, p. 319-326
[13] Protégé – Free, open-source ontology editor and framework for building intelligent systems: Available at http://protege.stanford.edu (accessed 2017, Sept)
[14] Web Ontology Language. Available at: http://www.w3.org/TR/owl-features/ (accessed 2017, Sept)

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ДЛЯ ПРЕДСТАВЛЕНИЯ БОЛЬШИХ ДАННЫХ ТЕЛЕКОМ ОПЕРАТОРА

Глоба Л.С., Новогрудская Р.Л.
Национальный технический университет Украины "Киевский политехнический институт имени Игоря Сикорского"

Аннотация – В статье представлено подход к описанию Больших Данных телеком оператора с использованием онтологической модели. Элементы онтологической модели вводят не только описание различных атрибутов данных, но и описание семантической связи таких данных. Используя онтологическую модель становится возможным организовать поиск информации, а также реализовать логический вывод по данным. Описано специализированные параметры Больших Данных телеком оператора. Предложена оценка качества услуг предоставляемых телеком оператором с использованием нечеткой логики, а также, описана ее взаимосвязь с метаописаниями Больших Данных за счет использования онтологии.