# Big Data in Service Delivery System by Communication Operator

Globa L.S., Svetsynska I.S., Volvach I.A.
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine
Email: e.svetsinskaya@hotmail.com
Email: lgloba@its.kpi.ua

Библиотека БГУИР

***Abstract*—In work complex analysis specialties of degree of subscribers' satisfaction taking into account both the services' technical parameters of and economic characteristics are considered when communication operator provides services. Object is achieved by methods of fuzzy logic for collaborative accounting of the impact of clear and fuzzy parameters. And it uses and a transition from unstructured data to the fuzzy knowledge base with a well-structured rules, that allows to implement a strategy typical for Big Data: data - information - knowledge, and thus significantly reduce the amount of calculations when data transmitting and processing.**

***Keywords—data, information, parameters, Big Data, metagraph, knowledge base.***

## I. Introduction

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

Accumulating data from network nodes rapidly increase every year. We need powerful computers to increase processing speed and data access. In connection with this, necessity to improve large amounts of data processing algorithms becomes more urgent.

According to research were conducted by a number of leading companies in the world [1] 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 of high processing speed in real time in order to reduce the volume of storage;
- heterogeneity (can be structured, unstructured, semi-structured);
- necessity respond reliability requirement (may be disrupted due to the variety of data sources and processing methods, violations of safety requirements);
- importance (using of forecasted methods and methods of analysis allows to predict the direction of companies development).

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

Therewith data analysis is often performed based on data, which is obtained as a result of the economic operator's activity, or base 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) to achieve the objectives. If we consider the degree of users'satisfaction by services (it offered by telecom operator) problem, it is quite obvious that the frequent technical failures affect degree of satisfaction, pricing operator's policy and the services quality also have an impact on the final evaluation of the customer services quality.

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 [1]:

- 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.

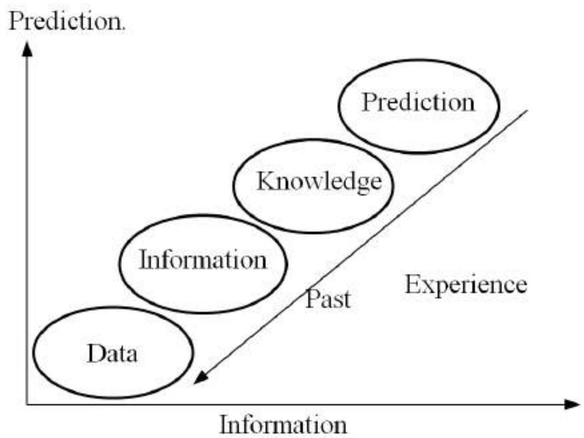


Figure 1. Scheme of the transition from unstructured data to information and to knowledge

## II. Existing mathematical methods for Big Data

To implement the collection processes, storage, processing, analysis and forecasting such DB processing methods are now widely used:

- association mining;
- classification;
- clustering;
- neural networks;
- support vector machines;
- decision tree learning;
- others.

But there no data analysis algorithm, by which it would be possible to to combine all the components (technical, economic and social).

## III. Proposed approach for collection, analysis and processing Big Data for communication operator

In this work joint assessment of all unbound clear and fuzzy parameters is proposed to use for constructing a fuzzy rule base, which would allow to produce a complex analysis of the expected quality of service delivery by telecom operator.

According to the considered features of Big Data processing, which are widely discussed in several papers [2; 3], the approach is proposed. This approach is based on fuzzy logic methods for taking account the impact of clear and fuzzy parameters, and using the transition from unstructured data to the fuzzy knowledge base with well-structured rules, that can significantly reduce the amount of calculations in the transmission and data processing. The main steps of this approach:

Step 1 Data stream forming (functional and sense dependencies extraction, meta description forming, parameters' impact rate determination).

Step 2 Fuzzy Knowledge Base (FKB) forming.

Step 3 FKB reorganizing and learning with the help of metagraph theory.

Step 4 Fuzzy conclusion.

Step 5 Fuzzy prediction.

One of the most difficult steps of the proposed approach is precisely the first step, where you should create define, organize and classify the data streams to get the possibility of applying algorithms of fuzzy knowledge base rules formation. It will allow to use this base in future for making decisions and predictions.

For this purpose, it is necessary to determine the parameters for forming a knowledge base, and methods for obtaining the values of these parameters.

Note, that telecom operators have already carried out analysis of economic efficiency, an analysis of the technical network infrastructure by monitoring system, also conducted a polls of subscribers for communication quality evaluation and services satisfaction. But the analysis processes of three components are isolated and are not included in the complex.

## IV. Parameters definition of the knowledge base

In the context of the degree satisfaction with the quality of services assessment by end-users and the possible churn influencing this process parameters are divided into the following groups:

- parameters characterizing the technical state of the system;
- the cost of using voice and Internet services;
- quality of provided services.

To collect data as the sources of initial statistical data for the task solution is proposed to use:

- for technical parameters - operator's functioning monitoring system;
- for economic parameters - available estimates of economic indices at companies;
- for sociological parameters - available subscribers' opinion poll at companies.

These data are combined into one or several tables in accordance with the timing parameters for the formation of a fuzzy base rule. To get an array of statistical data for each parameter parameters calculation is carried by using existing data, which were collected by the system. Structuring and classification of these data allows getting a table (Fig. 2).

Based on the statistical data (data were presented in the table) rules represent knowledge are formed (for example, the forecast of degree change of customer satisfaction in the future periods). These rules have the form of:

Библиотека БГУИР

| Subscriber's satisfaction (Y1) | Technical parameters (Y2) | | | | | Economical parameters (Y3) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Network Operability | Call Setup Failure Rate | Subscriber Perceived Congestion | Call Drop Rate | ... | Churn rate | Appetency | Gini index | ... |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

Figure 2. Structuring and classification of data streams to form the rules of fuzzy knowledge base (FKB)

<< IF Y2 AND Y3 ..., THEN Y1 >>.

Using these rules the analysis of the social component of the network for the next period can be carried.

To solve the objective of satisfaction degree assessment of quality of service by subscribers it is necessary to determine integral index, which can characterize the quality of the services and which can be obtained from the data network monitoring system. Such parameter may be considered the integral index of quality of network functioning.

### *A. Integral quality index calculation*

To determine the integral quality index (KQI_D) it is necessary to calculate such additional parameters:

- Connection Success Rate - successful data connections 2G/3G, %
- Connection Block Rate - the percentage of line lockout due to overload
- Connection Drop Rate - the percentage of breakages data connections 2G/3G, %
- PS Attach Success Rate - the percentage of successfully completed procedures Attach 2G/3G, %
- PDP Context Activation Success Rate - the percentage of successfully completed activation procedures PDP Context 2G/3G, %
- Speed DL - the average daily speed HSDPA of transmitting data to the subscriber, kbit/s
- Iub Congestion - the share of 3G BS with high overloads on Iub interface, %;
- Backhaul Accessibility - zonal transport network availability, %
- DNS Success Rate - successful DNS resolution, %
- DNS Response Latency - DNS resolution time, ms
- Peering Utilization - peering interface load level, %.
- Backbone Abnormal Latency - exceeding the normal delay between trunk transport network's nodes, %
- $w1 \dots 13$ - weighting coefficients of parameters.

Fig. 3 shows the structure of mutual influences parameters, which are components of the integral quality index.

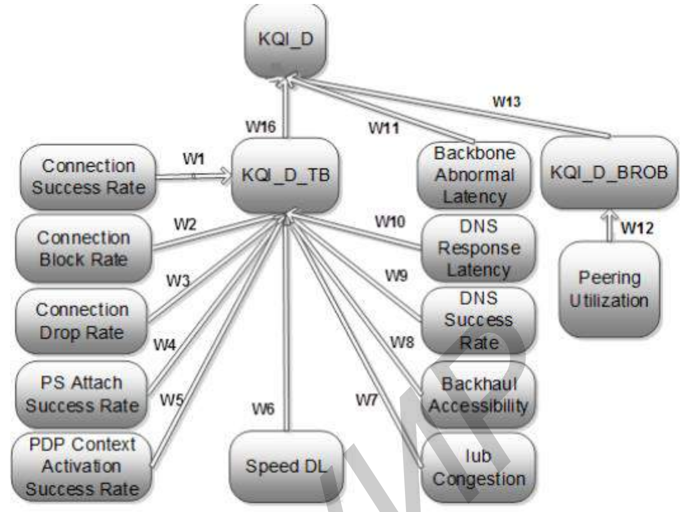


Figure 3. Scheme of integral quality index calculation

### *B. Average means of use voice services of 1 minute for all subscribers coefficient calculation*

Average means of use voice services of 1 minute for all subscribers coefficient (KVS) defined by (1):

$$KVS = \frac{P}{t}, \frac{USD}{min}, \tag{1}$$

where $P = \sum P_i$, $USD$ - the total cost of all voice services; $t = \sum t_i$, $min$ - the total time of all conversations.

The cost of one conversation is offered to calculate by the following formula:

$$P_i = k_t * (p_c + t_i * k_d * p_o)\,,\ USD,$$

where $k_t$ - subscriber's tariff plan coefficient; $p_c, USD$ - connection cost; $t_i, min$ - one conversation time; $k_d$ - call destination coefficient; $p_0, USD$ - one minute conversation cost.

### *C. Average means of use internet services of 1 MB for all subscribers coefficient calculation*

Average means of use internet services of 1 MB for all subscribers coefficient (KVS) defined by (2):

$$KIS = \frac{I}{m}, \frac{USD}{MB}. \tag{2}$$

where $I = \sum I_i$, $USD$ - the total cost of all internet services; $m = \sum m_i$, $MB$ - total amount of used traffic.

The cost of one internet session is offered to calculate by the following formula:

$$I_i = k_t * k_{type} * m_i * k_w * k_{dn} * I_0,\ USD,$$

where $k_t$ - subscriber's tariff plan coefficient; $k_{type}$ - traffic type coefficient; $m_i, MB$ - amount of used traffic; $k_w$ - day of the week coefficient; $k_{dn}$ - time of the day coefficient; $I_0$, $USD$ - cost of using 1 MB traffic.

Библиотека БГУИР

### D. Quality of services provided coefficient calculation

Quality of services provided coefficient is calculated by (3) [4].

$$F = \sum_{i=1}^{k} h_i * F_i \qquad (3)$$

where $F_i$ - partial quality index; $h_i$ - weightiness coefficient of partial quality index $F_i$ , which is defined, as a rule, based on subjective considerations and is normalized by $k$ partial quality index $\sum_{i=1}^{k} h_i = 1$.

## V. Knowledge base formation

FKB is formed according to metagraph theory, that allows forming knowledge base rules based on clear and fuzzy values of its parameters [5].

Metagraph is generalized concept of graphic structures. Formally, metagraph is range

$$< V, MV, E, ME >$$

where $V$ - range of vertices; $MV$ - range of metavertices; $E$ - range of edges; $ME$ - range of metaedges.

Fig. 4 shows a modified decision making algorithm which use FKB [6; 7]. This algorithm is the most common method of inference in fuzzy systems using minimax composition of fuzzy ranges, which is effective for ordinary (non-adaptive) fuzzy systems.

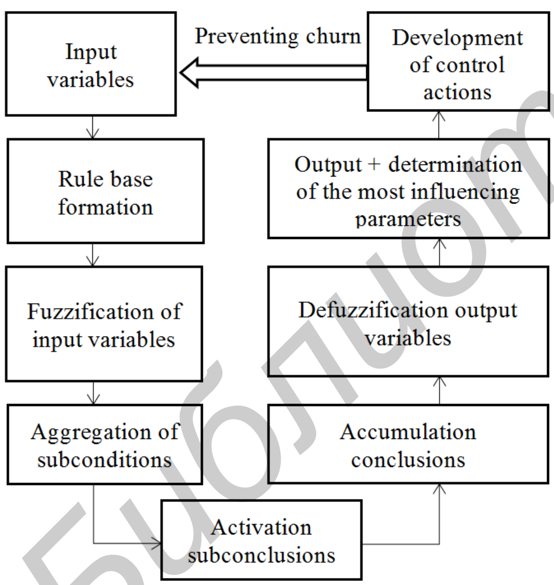


Figure 4. The process of preventing churn

Thus, the object is achieved by fuzzy logic methods, which use observed data in communication operator's network as input data. These methods take into account the combined influence of clear and fuzzy parameters and use transition from unstructured data to the FKB with clearly structured rules. It allows to implement typical strategy for BD: data - information - knowledge, and consequently significantly to reduce the amount of calculations when transmitting and processing of data.

## VI. Conclusion

An approach, that allows to complex analyze the degree of subscribers' satisfaction based on the change technical parameters of the operator's infrastructure, and economic performances of the offered services is proposed.

The advantage of the approach is the use of the knowledge base and rules generated by automated path instead of the large amounts of processed information.

Further work will be connected with analysis and the determination of specific parameters (technical, economic, social), that have the greatest impact on the process of subscribers' satisfaction by provided communication services.

## References


[1] M. Ulema, Big Data and Telecommunications Telecom Analytics, Tutorial II BlackSeaCom'2016.

[2] A. Mauro, M. Greco and M. Grimaldi, What is Big Data? A Consensual Definition and a Review of Key Research Topics, 4th International Conference on Integrated Information, AIP Proceedings, 2014.

[3] B. Marr and J. Wiley, Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance, Sons Ltd, 2015.

[4] L. P. Scherbina, Osnovyi teorii voennoy svyazi (Principles of Military Communication Theory), L.: 170 p., 1984.

[5] A. Basu and R. Blanning, Metagraphs and Their Applications, Springer, 2010.

[6] S. Shtovba, Proektirovanie nechetkih sistem sredstvami MATLAB (Design of fuzzy systems by MATLAB tools), M: Goryachaya liniya-Telekom, 2007. - 288 p.

[7] V. Dyakonov and V. Kruglov, Matematicheskie paketyi rasshireniya MATLAB. Spetsialnyiy spravochnik (Mathematical packages expansion of MATLAB. Special reference), Sankt-Peterburg: Piter, 2001 - p. 307-309.


## BIG DATA В СИСТЕМЕ ПРЕДОСТАВЛЕНИЯ УСЛУГ ОПЕРАТОРОМ СВЯЗИ


Глоба Л.С., Свецинская Е.С., Вольвач Е.А.



В работе рассматриваются особенности комплексного анализа степени удовлетворенности абонентов при предоставлении услуг оператором связи с учетом как технических параметров инфраструктуры предоставления услуг, так и экономических характеристик самих услуг. Поставленная задача решается методами нечеткой логики для совместного учета влияния как четких, так и нечетких параметров. Также использует переход от неструктурированных данных к нечеткой базе знаний с четко структурированными правилами. Данный подход позволяет реализовать стратегию, характерную для Big Data: данные – информация - знания. Такая стратегия позволяет существенно уменьшить объем вычислений при передаче и обработке данных.



Библиотека БГУИР