

Models and Methods of Classification of Innovation Market Information

Berg Tatiana Igorevna

Department of Economics and planning, master's degree in
Applied Informatics
Siberian Federal University
Krasnoyarsk, Russia tatiyana.berg@gmail.com

Stupina Alena Aleksandrovna

Department of Economics and information technology
management
Siberian Federal University
Krasnoyarsk, Russia
AStupina@sfu-kras.ru

Erygin Yuri Vladimirovich

Department of Finance and credit
Reshetnev Siberian State University of Science and
Technology
Krasnoyarsk, Russia
yuri_erygin@mail.ru

Yurdanova Vera Nikolaevna

Department of foreign languages for engineering
Siberian Federal University
Krasnoyarsk, Russia
vyurdanova@sfu-kras.ru

Korpacheva Larisa Nikolaevna

Department of Economics and information technology
management
Siberian Federal University
Krasnoyarsk, Russia
LKorpacheva@sfu-kras.ru

Ruiga Irina Rudolfovna

Department of Economics and business process
management
Siberian Federal University
Krasnoyarsk, Russia
irina_rouiga@bk.ru

Abstract— The importance of innovation as a product on the market field in the conditions of transformation of means of promotion, digitalization, and expansion of sales channels on the Internet is discussed in the article. The authors' vision of classifying innovations by the type based on a subjective approach has been developed: manufacturers who formulate supply on the market and buyers (customers) that determine real and potential demand. The informative characteristics of the types of innovations are based on nominal features that are of practical value to the main agents of the market. The selection of informative features of innovation is based on clustering methods. The Hamming metric is proposed as an effective way to determine the measure of distances. The search for the necessary information and data processing in the innovation market is carried out using modeling tools. Clustering algorithms are proposed as adequate models of intersecting. The clustering algorithm solves the problem of dividing the sample of signs of innovation into clusters that are similar by certain criteria (scientific novelty, patent protection, etc.), depending on the type of innovation as a product (patent, technology, etc.). The proposed classification of innovations will allow to position innovation on the market as an object of sale. Modeling tools accumulate, group, process, coordinate and manage innovation market information when making management decisions.

Keywords: *innovation, innovation market, information, classification, clustering, clustering models, Hamming metric, clustering algorithm*

I. INTRODUCTION

Structural adjustment of Russian economy is a priority course of innovative development in the context of strengthening globalization processes and the organization of a single information space. The objectives of this area are intended to be the expansion of volumes, nomenclature of products and services production based on the latest achievements of science and technology. The low innovation and investment activity of economic generators, lack of information on the demand and supply of innovative products, and barriers to entry and exit to technology markets hamper the implementation of the claimed technological solutions. The existing commercial exchange of products of innovation does not provide stable economic growth rates. The reasons for the current situation are the imperfection of a structured electronic database, tools for continuous monitoring of complex information on the volumes and structure of the innovation market, which hinders the process of entering real and potential markets of national and international levels. The

solution to these problems is seen in the subjective approach of classifying innovations clustering objects according to selection criteria and types of innovations, which has economic and other benefits for producers and consumers of innovations. In addition, the data obtained make it possible to assess the real and potential demand and supply for innovative goods and services.

II. RESEARCH METHODOLOGY

Studying the socio-economic impact of institutions of The main methodological provisions of the importance of innovation as an object of commodity exchange in the market are based on a thematic discourse analysis of scientific publications presented in open sources of foreign and Russian researchers in the field of innovative economics and management.

The classification of innovations is based on the subjective approach of the importance of information for agents of the innovation market (sellers and buyers). Innovation clustering is based on clustering criteria. To calculate the measure of distances of classification objects, the Hamming metric is chosen. The task of information search and data processing in the innovation market is solved with the help of modeling tools. As adequate models of intersecting (fuzzy) clustering algorithms are proposed. The criteria for selecting innovations in a cluster take into account the importance and effectiveness of decision-making.

III. RESULTS

In recent years, the innovation market has attracted demanded attention in scientific research and is considered as a tool for ensuring the development of a new technical and technological structure of economic systems. Innovations are complicated, accelerated, and become interactive [1], they act on the market as a specific product, which has an owner (copyright holder) for which the rights to the results of intellectual activity and equivalent means of individualization are recognized. In the era of high-tech, constant and dynamic technology updates, virtually no manufacturer can provide a complete innovative product on their own. For this, various types of interactions transfer channels are used [2]. The main driving forces do not change, but tactics have become more complicated [3]. New business models and “innovative markets” reduce costs and create an environment in which they make more efficient search procedures, template agreements [4]. In this situation, innovation markets require adaptation and revision of means of product promotion, shifting the focus from the sale of a product to cooperation and communication.

Issues of promoting [5] innovations are of particular importance in the context of the digital transformation of the economy. A new type of interaction and a management system for the relationship of potential customers (buyers, customers) with manufacturers of high-tech goods is being built. This eliminates excessive mediation [6] for conducting transactions based on online technologies. The exceptions are closed deals and markets that ensure the defense and security of the national economy. Digitalization helps to increase the diversity, quality and speed of services received in the

innovation market, improves transaction security, and saves time on search, acquisition and movement.

The concept of the modern innovation market emphasizes the importance of demand, end users and interaction among users and manufacturers [7]. They are potential buyers who can best “tell” about the factors of commercial success [8]. Distinctive features of the innovation market in the current realities include:

- dynamism of quantitative and qualitative development, susceptibility to cyclical changes, determined by changes in technological structures;

- long-term and diverse cooperation of the buyer with the developer of the innovation, due to the life cycle from concept to disposal. These features require constant monitoring of innovation market information to make alternative decisions, for both the producer and the buyer (customer) of the innovation.

The main source of information for information services in modern society are databases [9]. They integrate the information of suppliers and consumers of innovative goods and services, the connections and relationships between them, the procedure and conditions for the sale and purchase of innovations. However, the information available in Russia from the innovation market data bank has large amounts of information resources that are poorly coordinated and are stored in databases of different subordination. This argumentation complicates the information exchange, the concentration of information to solve a problem.

The experience of creating a bank of data resources of innovation is based mainly on participants in the innovation process and stages of the innovation life cycle [10, 11]. This approach is justified, however, in the conditions of rapidly changing technologies and international competition, the most significant, in our opinion, is the task of creating a bank of information resources depending on the agents (sellers and buyers) of the innovation market. Since the recognition of innovation by the end user provides economic and social benefits from their implementation, it helps to return investments at all stages of the innovation process.

It is proposed to create the information base of the innovation market on the basis of the main key agents of commodity exchange in order to interact demand and supply - producers and consumers.

A. The challenge of classifying innovation information

One of the approaches to improving the quality of pattern recognition and reducing computational costs is to evaluate the information content of signs, attribute values, and highlighting the most representative objects [12, 13]. For this purpose, it is necessary to transform information containing the conditions of the offer, the capabilities of producers and the requirements of consumers of innovations, which in the future will allow us to establish relationships and interaction between them, to carry out sales and purchase transactions.

The solution to this problem is seen in the structuring of types of innovation by classification criteria. The decomposition of innovations is based on nominal

characteristics, the allocation of which determines the significance of information on the needs of customers (demand) and proposals of manufacturers. The class gives an enlarged detail of the innovation market (Table I). In this paper, we ignore innovation sellers, since they act as integrators of innovation market information.

TABLE I. CLASSIFICATION OF INNOVATION BY TYPES AND SIGNIFICANCE OF DATA OF INNOVATION MARKET AGENTS INFORMATION [COMPILED BY THE AUTHOR]

Kind of information / innovation	Type of information / innovation	Significance for agents *	
		producers	consumers
scientific	scientific regulatory documentation R&D reports bibliographic reviews annotations reports forecasts abstracts dissertations books, etc.	++	+
macro technology	aviation space nuclear shipbuilding transport engineering chemical engineering special metallurgy; special chemistry; new materials oil production and refining gas production and transportation power engineering industrial equipment; machine tool industry micro and nano electronic computer and information communication biotechnology, etc.	+	++
state producer	advanced countries countries with limited scientific potential catching up countries lagging countries, etc.	+	++
producer	Institute of Science enterprise of defense large industrial company - world leader manufacturing enterprise production laboratory business incubator industrial park technopolis, etc.	+	++
type of innovation in management (business process)	production marketing organizational technological	+	++
type of innovative commodity (product)	completed innovative technology incomplete innovative technologies at various	++	++

	stages of readiness technological innovations that improve manufacturing processes means of automation of production processes quality control systems for manufactured products process control systems equipment, instruments, tools license patent trademark service certificate, etc.		
field of activity	production trade social sphere management, etc.	+	++
after-sales service	innovation with after-sales support unaccompanied innovation	+	++
economic effect	production of new types of products (services) cost reduction for the production of products (services) improving the quality of products (services) reduced material consumption increasing energy efficiency of production mechanization and automation of production environmental friendliness of production (services) increase in labor productivity, etc.	+	++
*significance of the recipient of information: ++ - especially significant; + - significant.			

Material media of innovation market information can be documents of various forms: textual (books, magazines, reports, etc.); graphic (drawings, diagrams, schemes); Audiovisual (sound recordings, films and video films); electronic (records on electronic media, information on computer networks).

B. The task of selecting an innovation class

In solving the problem of choosing the class [14] to which the innovation belongs, the problem arises of evaluating it according to several criteria[15] and correctly considering these estimates when generalizing them or considering them together at the stage of making the final decision. The solution to the problem is possible on the basis of a discrete approach

to assessing the information content of heterogeneous characters and the analysis of statistical methods [16].

We study a number of educational objects of innovations $I = \{I_1, \dots, I_m\}$ described by the system of attributes $\{z_1, \dots, z_m\}$ belonging to certain classes. It is known that the set (I) is representable as the union of K subsets of S_1, \dots, S_K , called classes. The description of the innovation has the form $q_i = (q_{i1}, \dots, q_{im})$, $q_2 = (q_{21}, \dots, q_{2m})$, ..., $q_m = (q_{m1}, \dots, q_{mm})$, where q_i is the value of the attribute z_j for the object I_i . it is necessary to determine the class according to the presented set of attribute values (b_1, \dots, b_m) , that describes some object from I (not included in the training set I_1, \dots, I_m), which is not known which class it belongs to.

The most general definition of a class is the criterion for selecting innovation to solve the search problem. Information on the properties of innovation can be obtained from manufacturers or consumers and generated by the seller (provider) of the innovation market. Information values are expressed in a numerical scale. The objects included in the same class are considered equivalent, and each class of objects is characterized by a certain quality that distinguishes it from other classes [13]. The selection criteria for the first level for the buyer are: the novelty of the technology and its competitive advantages, scope (industry) of application; type of technology); compliance with world standards; patent protection and terms of legal protection of the main technical solutions; the availability of know-how; the volume of transferred rights to use the technology. The criteria of the second level can be considered all types of information presented in table I.

The selection of innovations is carried out not only by class (property), but also by image [17], by which we understand the type of innovation (technology, license, etc.) represented in the matrix model. The matrix model (M) of these innovations includes a description of the properties of innovations (Q) in the space of characteristic features and differences (R) of types of innovations in the space of classification features. The set of all non-repeating rows in the matrix is mapped to the set of selected images. Elements of the image are objects represented by rows of the matrix, mapped by the same columns of the matrix R. Matching rows and columns give the right search solution (Table II).

TABLE II. MATRIX (M) OF INNOVATION SELECTION AS A PRODUCT BY MARKET AGENTS [COMPILED BY THE AUTHOR]

Selection criterion/ properties of innovation (Q)	Type of innovative product (fragment) (R)					
	technology	license	patent	trade mark	device, system, etc	service
1st level						
novelty of technology	1	0	1	0	1	0
field of activity (industry)	1	1	1	1	1	1
type of technology	1	0	0	1	1	1
application in business processes	1	1	1	1	1	1
patent protection	1	1	1	1	0	0
terms of legal protection of basic technical solutions	1	0	1	1	0	0
know-how available	1	0	1	0	1	1
volume of transferred rights to use the technology	1	1	1	0	0	0
Level 2 (fragment)						
country of origin	1	0	0	1	1	1
manufacturer	1	0	1	1	1	1
type of innovation in management	1	1	0	0	1	1
economic effect	1	0	1	1	1	1

IV. DISCUSSION

The global search for files, websites, and other objects of market information is based on an “intellectual” grouping of innovations. To solve the search problem [18], a classification method using the genetic algorithm is proposed [19]. The method is based on the cluster analysis model [20], as a universal tool for data classification and processing [21-23]. Cluster analysis enables an agent of the innovation market to quickly navigate, select a knowingly more relevant subset and exclude a knowingly less relevant one. Clustering concretizes the information of the innovation market, improves the usability of the interface compared to the output of a simple list sorted by relevance.

The result of the search for the necessary innovation in the market is a set of parameters expressed by quantitative, qualitative and mixed characteristics for each innovation object in the form of numerical values [0, 1]. For each pair of objects, the degree of similarity is measured - the distance between them. As metrics, we choose the Hamming metric on features (1), applicable for nominal features with the same sets of permissible values $Dz_1 = Dz_2$, which vanishes if and only if the vectors z_1 and z_2 coincide:

$$p(z_1, z_2) = \sum_{i=1}^n [z_1(O_i) \neq z_2(O_i)]. \quad (1)$$

For nominal features with different sets of values, a correspondence $\sigma: Dz_1 Dz_2 \rightarrow$ is searched for, at which the Hamming distance is minimal (without loss of generality, it is assumed that $Dz_1 Dz_2 \geq$).

The methods of cluster selection of innovations according to various criteria make it possible to break down the sample into clusters consisting of several characteristics and distinguish one most typical innovation in each group. For example, the technology should be novel, applied in mechanical engineering, be an additive technology with patent protection, for which the terms of legal protection of the main technical solutions are established, have “know-how”, produced in an advanced country, the manufacturer is a world leader, practically applicable in grocery business -processes, ensures the production of new types of products, reduces production costs, automates processes, increases the environmental friendliness of production and goods, etc.

Classification by the principle of minimum distance gives an additional advantage of the application of clustering methods. Since the clusters corresponding to different classes can be separated far enough from each other [24], and the values of features describing objects from different images are quite far from each other [13].

The task of cluster analysis is to find k-means (centers, clusters of innovation in the d-dimensional space of characteristics (d-number of agents). To determine the centers of each of the clusters representing innovation. The task of k-means is to minimize the following objective function (2) for known data vectors A_1, \dots, A_n [19].

$$F(X_1, \dots, X_k) = \sum_{i=1}^n \min_{X \in (X_1, \dots, X_k)} \|A_i - X\|, \quad (2)$$

where $\{A_1, \dots, A_n\} - \{A_1, \dots, A_n\}$ is the set of data vectors in d-dimensional space; X_1, \dots, X_k — the desired points of the clusters;

To solve the problems of local search, the ALA algorithm (Alternating Location-Assignment) is used [25, 26]. The local search procedure begins with the determination of the set of clusters, which is used as a subset of k data point vectors (selection criteria). Next, the sets of data vectors for each cluster for which this cluster is the closest are alternately found, and a new cluster is found for each of these sets.

A. Algorithm. ALA procedure.

Given: data vectors A_1, \dots, A_n , k of the initial centers of clusters X_1, \dots, X_k .

1. For each center X_i , define a cluster C_i - a subset of data vectors for which the center X_i is the closest.

2. For each cluster C_i , redefine its center X_i (3):

$$X_i = \arg \min_x = \sum_{y \in C_i} \|x - y\|. \quad (2)$$

3. Repeat from step 1 if at least one cluster has changed in steps 1, 2.

The practical application of the classical method of k -means with a quadratic Euclidean distance measure ($L_{2/2}$) has an advantage: when using a quadratic Euclidean measure, determining the center of the cluster is the simplest problem that can be solved in one step — the average value of each

coordinate of the data vectors in the cluster is determined, the obtained values are the coordinates of the center of the cluster.

If the center of the i cluster $X_i = (x_{i,1}, \dots, x_{i,d})$ is a vector in the d -dimensional space and data vectors $A_j = (a_{j,1}, \dots, a_{j,d}), j = \overline{1, n}$, respectively, also have d measurements, then the new center is determined by the formula (4) [27]:

$$x_{i,k} = \frac{\sum_{y \in C_i} y_k}{|C_i|}, \quad k = \overline{1, d}. \quad (4)$$

When organizing a global search, either the multiple start of the local search procedure with randomly selected initial solutions is used, or various recombination algorithms of the initial solutions are used.

The set of parameters of the center of the cluster may not completely coincide with the parameters of one of the data vectors, but the value of each individual parameter necessarily coincides with the value of this parameter of one of the data vectors (that is, one of the type of innovation). But at the same time, the accuracy of the value is preserved.

The proposed decomposition of dividing innovations into classes and clusters will allow us to compare innovation in a cluster, find the most characteristic object and make a decision.

Information on innovations is of particular importance in solving such problems as:

- formation of a market of supply and demand for innovative products for sellers, manufacturers and consumers;
- building cooperative ties for manufacturers of innovative products;
- organization of development supplies and technology transfer.

V. CONCLUSION

Improving the positioning of innovative products ensures the growth of production and marketing, the development of economic relations between producers and consumers. The task of classifying innovations by type makes it possible to quickly, easily obtain information from subjects of the innovation market. Models and methods of innovation clustering will allow solving the problems of information search and data processing in the innovation market. Hamming's metric reduces the measure of distance choice of information on innovation. The application of the model of fuzzy clustering algorithms in the selection of information takes into account the properties and types of innovation, as an object of sale, and determines the final search solution. A justified choice of methods and models for classifying innovations, a data processing method and practical application will significantly reduce the labor costs for preparing management decisions for innovation at various levels.

ACKNOWLEDGMENTS

The presented research was funded by Krasnoyarsk Regional Fund of Science in the framework of the scientific project: «Methodological approaches to the formation of a new system for innovative investment sustainability for resource-type regions based on elaboration of the cluster policy mechanisms (as exemplified by the Krasnoyarsk territory).

The study was financially supported by the Krasnoyarsk Regional Fund for Support of Scientific and Science and Technical Activities, a regional state autonomous institution, as part of research project No. 2019051705126, “Methodological approaches to the formation of a system for ensuring innovation and investment sustainability for resource-type regions based on improving cluster policy implementation mechanisms (on the example of Krasnoyarsk territory).

REFERENCES

- [1] H. Chesbrough, *Open Innovation: The New Imperative for Creating and Profiting from Technology*, 2003.
- [2] G. Prause, T. Turner, "Consumer communities - drivers of open innovation", *Foresight*, vol. 1, 2014.
- [3] J.A. Moore, *Bridging the abyss. How to bring a technological product to the mass market*, 2013.
- [4] A. Bingham, D. Spradlin, *The open innovation marketplace: Creating value in the challenged driven enterprise*, 2011.
- [5] V. Serra-Sastre, A. McGuire, "Information and diffusion of new prescription drugs", *APPLIED ECONOMICS*, vol. 45(15), 2013, pp. 2049-2057, <https://doi.org/10.1080/00036846.2011.648321>.
- [6] M. Cui, S.L. Pan, S. Newell, L. Cui Strategy, "Resource Orchestration and E-commerce Enabled Social Innovation in Rural China" *The Journal of Strategic Information Systems*, vol. 26(1), 2017, pp. 3-21, doi: 10.1016/j.jsis.2016.10.001.
- [7] B. Gregersen, B.A. Lundvall, "The Public Sector as a Pacer in National Systems of Innovation", *National Systems of Innovation - Towards a Theory of Innovation and Interactive Learning*, 1992, pp. 129-145.
- [8] B. Jaworski, A.K. Kohli, "Market orientation: Antecedents and consequences", *Journal of Marketing*, vol. 57, 1993, pp. 53-70.
- [9] I.N. Korabeynikov, J.A. Ermakova, "Development of the market for information services: theoretical aspects", *Bulletin of OSU*, vol. 4(140), 2012, URL: <https://cyberleninka.ru/article/n/razvitie-rynka-informatsionnyh-uslug-teoreticheskie-aspekty> (accessed: 02.10.2019).
- [10] Yu.M. Arsky, V.A. Tsvetkova, S.P. Yashukova, "Information resources to support innovation", *Information support of innovation*
- [11] E.V. Ioda, S.Yu. Ponomarev, "Information support for the interaction of subjects of innovation in the region", *Competitiveness. Innovation, Finance*, vol. 2, 2011, pp. 17-20.
- [12] A.I. Dmitriev, Yu.I. Zhuravlev, F.P. Krendelev, "On the mathematical principles of the classification of objects or phenomena", *Discrete analysis*, vol. 7, 1966, pp. 1-17.
- [13] S.I. Kolesnikova, "Methods of analyzing the information content of heterogeneous characters", *Tomsk State University Journal. Tom. state un-that. Management, computer engineering and computer science*, vol. 1(6), 2009, pp. 69-80.
- [14] V.S. Govada, I. Thomas, S.K. Samal, S.K. Sahay, "Distributed Multi-class Rule Based Classification Using RIPPER", *IEEE International Conference on Computer and Information Technology (CIT)*, 2016, pp. 303-309.
- [15] S. Asadi, J. Shahabi, M.C. Rip, "RIPPER for multiclass classification", *Neurocomputing*, vol. 191, 2016, pp.19-33.
- [16] L.D. Meshalkin, "Some mathematical methods for the study of noncommunicable diseases", *Proc. 6-th Intern. Meeting of Uses of Epidemiol. in Planning Health Services. Yugoslavia, Primosten*, vol. 1, 1971, pp. 250 - 256.
- [17] A.E. Yankovskaya, "Test pattern recognition with the use of genetic algorithms", *Pattern Recognition and Image Analysis*, vol. 9(1), 1999, pp. 121-123.
- [18] D. Dua, C. Graff, *UCI "Machine Learning Repository"* [<http://archive.ics.uci.edu/ml>], Irvine, CA: University of California, School of Information and Computer Science, 2019.
- [19] L.A. Kazakovtsev, V.I. Orlov, A.A. Stupina, I.S. Masich, "The task of classifying the electronic component base", *Bulletin of SibSAU*, vol. 4(56), 2014, pp. 55-61.
- [20] P.N. Tan, M. Steinbach, V. Kumar, "Cluster Analysis: Basic Concepts and Algorithms", *Introduction to Data Mining*, Addison-Wesley, vol. 8, 2006, pp. 487-567.
- [21] I.S. Masich, E.M. Kraeva, "Selection of the decision rule in logical recognition algorithms", *Sistemy upravleniya i informatsionnye tekhnologii*, vol. 51(1.1), 2013, pp. 170-173.
- [22] A.V. Egorov, N. Kupriyanova, "Features of data clustering methods", *Bulletin of the Southern Federal University. Technical science*, vol. 11, 2011, URL: <https://cyberleninka.ru/article/n/osobennosti-metodov-klastersizatsii-dannyh> (accessed September 22, 2019).
- [23] X.L. Xei, G.A. Beni, "Validity Measure for Fuzzy Clustering", *IEEE Transaction on Pattern Analysis and Machine Intelligent*, vol. 3(8), 1991, pp. 841-846.
- [24] J. Tu, R. Gonzalez, *Principles of Pattern Recognition*, 1978, p. 415.
- [25] L. Cooper, "Location-allocation problem", *Oper. Res.*, vol. 11, 1963, pp. 331-343.
- [26] R. Kuzmich, I. Masich, A. Stupina, L. Kazakovtsev, "Algorithmic Procedure for Constructing the Truncated Basic Set of Characteristics in the Method of Logical Analysis of Data", *VISION 2020: SUSTAINABLE ECONOMIC DEVELOPMENT, INNOVATION MANAGEMENT, AND GLOBAL GROWTH, VOLS I-IX*, 2017, pp. 5592-5597.
- [27] R.Z. Farahani, M. Hekmatfar, *Facility Location Concepts, Models, Algorithms and Case Studies*, 2009, p. 550.