Recommender systems for carer guidance

Đức The Trần[1], Evgeny Cherkashin[1,2,3], Viktoria Kopylova[2], Nikita Lukyanov[2]
[1] *Institute for Informational Technologies and Data Analysis*,
National Research Irkutsk State Technical University, Irkutsk, Russia
[2] *Laboratory of Complex Informational Systems*,
Matrosov Institute for System Dynamics and Control Theory, Siberian Branch of Russian Academy of Sciences, Irkutsk, Russia
[3] *Chair for Informational Technologies*,
Institute for Mathematics and Information Technologies, Irkutsk State University, Irkutsk, Russia
eugeneai@icc.ru

***Abstract*—The research of the paper is devoted to present of an R&D of a recommender system in carer guidance for Việt Nam middle school students. The system is built on the base of John Holland's six personality types theory and collecting university students' opinions on their study conditions. The RS relates new user (an entrant or a pupil) to a student group and then presents [[top-N]] universities/departments as a recommendation. Techniques of overcoming "cold start" problem and other ones in this domain are considered.**

***Index Terms*—recommender system, cold start, carer guidance, TODO:**

I. Introduction

In the countries with predominantly young population, such as Việt Nam, the demand of young specialists is high. While still a high school student, one have to make a decision about a career. Career choice depends on many factors such as personal interests, academic ability and employment opportunities available after the graduation. Therefore, choosing the right carer is not an easy task. In fact, quite often graduates remain unemployed for a long time or do not work in their specialties mastered in an educational institution. This leads to inefficient use of state funds, their irrational distribution. That's why the choice of an university after receiving a certificate of maturity is very important. Choosing a suitable university results in an increase students' study productively, as they work and practice with enthusiasm. Supporting the decision activities will give a better chance of achieving students' goals in the future.

In Tables 1 and 2, we present the result of interest assessment to career guidance and relevance of improving of the guidance system. Tables show that there is a real problem of carer guidance and middle school students are very interested in improving the corresponding activities.

TABLE I
Interest assessment to the carer guidance in Việt Nam, Hà Tĩnh city

Interest level	Count of votes	Ratio, %
Very interested	218	51.9
Relatively interested	155	36.9
Less interested	36	8.6
No interest	11	2.6
Total	420	100

TABLE II
Questionary results of middle school students on need for improvement of the carer guidance system

Answer option	Quantity	Ration, %
very necessary	272	64.8
necessary	145	34.5
no necessity	3	0.7
`Total`	420	100

At present, with the development of the Internet, the search for information about universities is sufficiently efficient. However, having a huge amount of information, selecting substantive information is a difficult task. Recommender systems (RS) have emerged as a decision support tool, providing users with the most useful and personalized variants for goods and services. The functioning of RS is based on filtering information with respect to a set of known properties of objects and users. For example, RSs are used to assess the users' preferences for goods and services (songs, films, video clips, books, articles, *etc*.), which have been not previously given ratings by the user trying to make a choice.

RSs are also quite successfully used in many business spheres, such as entertainment: offering songs to listeners (*e.g.*, the LastFM system – www.last.fm), offering films (the Netflix system – www.netflix.com), recommended videos (the YouTube system – www.youtube.com); in education and training (learning resources, books, articles, site addresses), in intelligent systems of teacher assistants (predicting students' learning capabilities). RSs are the field of active IT research since 2007.

Teaching experience of the first author of this paper in secondary school shows the relevance of the problem of choice of a profession. This research, being the results of a master thesis, is dedicated to the development of an RS that allows students (applicants) to receive recommendations when taking career decisions. The aim of the research is development of a RS for supporting students in their carer decision making.

The resulting RS prototype has two options for producing recommendations. The first one is realized on the base of John Holland theory [**?**], according to which most people have one of six personality types. After determining the type, a

set of corresponding universities are produced. The second one is an aggregation of questioning data about concrete universities obtained from students already being taught there. The recommendations, then, are generated by means of collaborative filtration (CF) based on users' profile comparison. Thus, the RS prototype combines two approaches: using expert knowledge and CF.

II. Technique of construction for the RS helping with the carer choice

Recommender systems [1] are decision support information systems designed to assess the user's level of interest in a particular product or service (object) based on available information about user and object. The RS development industry began to actively develop with the emergence of online sales services, and now it is one of the active areas of development of decision support systems, a direction of artificial intelligence, focused primarily on commercial use, as well as on solving problems of increasing the productivity of searching for relevant information. A profession is the object of RS recommendation production. Let us consider the development technique for RS construction from the point of view of solving the standard set problems and challenges.

Development of the RSs is aimed at solving the following set of problems [**?**]:

1) Increasing the sales of a product
   - a number of commodities sold,
   - organizing wider range product sales;
2) Increasing user satisfaction and/or loyalty;
3) Better understanding user needs;
4) Better products offers with respect to the user needs;
5) Selection of sets of products for users with a common properties
   - "good" ones, and
   - product groups, having a common usability properties;
6) Mining the classes of products, structuring the RS product domain, and
7) Generate a continuity of recommendations using the classification;
8) Rectifying user profile, *e.g.* with targeted questioning;
9) For the users having no goal to make choice, but searching for an expert opinions,
   - Analysis of other users impact on a choice,
   - Formalizing opinions,
   - Recommending opinions.

In this context, the RS is aimed at solving 4–7-th problems. This set defines the methods of RS proposal generation and algorithms, which have to be implemented. As we can see, the universities and their departments can be distributed on various carer directions by two ways: with the use of expert knowledge, when experts analyze a departments' properties and relate them to a direction, and with use of onions of the students (users who already "consume" a product). In the second case, a good opinion of a user on a department (product) creates or solidifies a corresponding relation. In the same time, a group of students positively characterized a department form a group of "similar" users.

The main problem solved during the R&D is "*cold-start*" *problem*. For RS based on CF, the similarity of user profiles in terms of sets of specified characteristics are performed. Content filtering RS compare products. In both cases, RS cannot generate recommendations if it does not have enough information about users or products. The cold-start problem arises at the first stages of RS functioning and when a new user's behavior had not been observed sufficient time, *i.e.*, the RS have no data about his preferences, or user's profile contains no useful information for a comparison. The same situations arises for a new product, *e.g.* there is no user bought the product, and no responses were given. In this system explicit information collection has been realized for solving cold-start problem in CF, and John Holland's theory was used as initial profile data filling in, including the input data for the Holland's questionary.

The simplest technique for user profile data collection and usage is so called *impersonalized acquisition*. Recommendations are generated for an "average" user related to a nearest group od users. An RS gives top-N products for the group, efficiently ignoring personal preferences. The approach can be developed to support personal data with construction of subgroups of the groups relating them to a set of personal characteristics. This may lead to unwillingness giving personal data of users and impossibility to generate proposals when a user cardinally changes his/her preferences. In this system, as our intention is a detection of a direction of a carer, impersonalized approach seems to give sufficient quality result.

The *scalability problem*. Scalability is a property of software systems to be able to cope with the load prescribed by design, as well as with increased load if additional information–communication and computing resources are available. For a simple RS with several users and products, there is no need to maintain scalability, and algorithms such as "nearest-neighbor" are widely used for assessing interests as the combinatorial space is not large.

Thus, to construct carer choice RS prototype, we have chosen to collect impersonalized data, structure users with respect to their university/department preference and user profiles by means of collaborative filtering, use questionary and expert approach for initial user profile filling in.

A. Holland's six types TODO

Realistic, Investigative, Artistic, Social, Enterprising, Conventional.

In order to determine the type one must answer 54 questions, the questions are evenly distributed on this six types. For each question, a 0-4 scores are to be set depending on the degree of manifestation. The store 0 corresponds to "no manifestation" was mentioned of a kind, and score 4 denotes "always manifestative". For each type the scores of the corresponding nine questions are collected, and the type is figured out as a type with the maximum score sum. Student

profiles store all answers and estimates, just for a further system development.

III. Preliminary Data Processing

In realty RS, two program agents are implemented, which are executed on an event occurrence. The first one starts by timer of web browser and implements issuing user positive evaluation, if user spends some time viewing a page of a realty object. The second one, located at the server side, is activated if the first one raised event of a positive evaluation. Server agent receives the evaluation and may start recalculation of estimates for recommendations if there are enough computational resources.

All data are stored in a database, which structure is represented as objects in Fig. **??**. The central entity is `IOffer`, representing the offer of a realty object. The information on an object, which is independent of an offer, is stored as `IObject`-object. The object structure corresponds to standard Yandex and Google form of real estate representation.

IV. Database structure

V. Implementation

The RS is a web-application implement in Django MVP[1] framework, MySQL has been chosen for data storage. Python programming language is the main glue subsystem, gathering web-application and recommender unit in one system. For carrying out experiments with data, a Jupiter subsystem was included in the system as well.

User roles are administrators, anonymous and regular user. Administrator can manipulate user accounts and any raw data. Anonymous users can register, view information on universities, taught courses and professions. Regular registered user can obtain recommendations, rate universities, add an university to his/her preference list. The use-case diagram is shown in Figure 1.

```
from django.db import models
from django.contrib.auth.models import AbstractUser

# User
class NGUOIDUNG(models.Model):
IdUser = models.AutoField(primary_key=True)
TenNguoiDung = models.CharField(max_length=20, null=False)
HoTen = models.CharField(max_length=200, null=False)
MatKhau = models.CharField(max_length=200, null=False)
Email = models.CharField(max_length=200, null=False)
Quyen = models.IntegerField(null=False)
GioiTinh = models.BooleanField(default=False)
SDT = models.CharField(max_length=15, null=True)
NgaySinh = models.DateTimeField(null=False,
  default="1999-01-01")

# Interests
class TINHCACH(models.Model):
IdTinhCach = models.AutoField(primary_key=True)
Loai = models.IntegerField(null = False)
ChiTiet = models.CharField(max_length=1000, null=False)
# List university

class TRUONG(models.Model):
IdTruong = models.AutoField(primary_key=True)
```

[1]Abbreviation of Model-View-Controller programming pattern

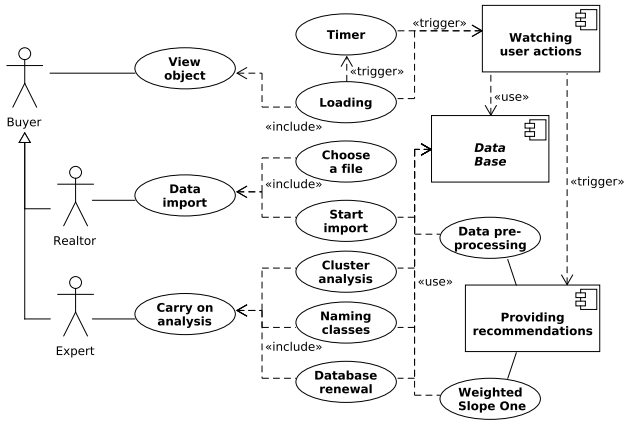

Fig. 1. Use case diagram for real estate RS [TODO: UPDATE to CDT's thesis]

```
TenTruong = models.CharField(max_length=1000, null=False,default="")
ChiTiet = models.CharField(max_length=1000, null=True)
ChiTietFull = models.CharField(max_length=1000, null=True)

# Vote for an university
class VOTE(models.Model):
IdVote = models.AutoField(primary_key=True)
IdTruong = models.IntegerField(null=False)
IdUser = models.IntegerField(null=False)
Point = models.IntegerField(default=0)

# List of favorite universities
class FAVORITE(models.Model):
IdFavo = models.AutoField(primary_key=True)
IdTruong = models.IntegerField(null=False)
IdUser = models.IntegerField(null=False)
```

For representing answers on Holland's questionary and intermediate data two tables added.

```
from django.db import models
# Create your models here.
class ListUni(models.Model):
name = models.CharField(max_length=100)
description = models.TextField()
excerpt = models.TextField(max_length=300)
pictrure = models.ImageField(upload_to='Uni_picture')
group = models.IntegerField(default=0)
def __str__(self):
return self.name
class list1(models.Model):
name = models.CharField(max_length=100)
anddress = models.CharField(max_length=200)
description = models.TextField()
excerpt = models.TextField(max_length=300)
pictrure = models.ImageField(upload_to='Uni_picture')
group = models.IntegerField(default=0)
def __str__(self):
return self.name
```

TODO Activity diagram

VI. Recommendations Generation

VII. Web Applications

VIII. Related Works

There are many research groups and commercial firms dealing with developing RS technologies for various domains. We will consider related domains and other interesting cases.

In [19] a case-based reasoning RS for real estate market is presented. Its goal is to find a similar instance in the database of existing cases, describing the cases for the user. The obtained relations are used in followed collaborative filtration and content filtering. In theoretical research [20] the users are divided on sellers and buyers. Sellers "advertise" their realty by highlighting the important features by their opinion. So the sellers play expert roles for forming data for content filtration. Research [21] showed that usage of Internet does not influence significantly on the time efficiency, flexibility and satisfaction criteria of the search for a flat to buy. There are no statistical difference between 2009 and 2011, since in 2011 88% of buyers used Internet just as the main source of information. Authors created RS with case based reasoning utilizing an ontology model.

In [21] authors highlight three key features which significantly affects the decision: location, housing unit properties and price. Final decision is made after evaluation of the environment, such as distances to shops, kindergartens, schools. This can be accounted reducing the additional data to the value of the feature of housing unit properties using ontology. The environment constraint is entered by user: user draws a circle around the preferable location, which should contain interested services and realty.

Paper [22] considers a complex logistic problem of organization of a process of real estate control, involving various people group in changing business environment. The main aim is to develop a model, where the groups will be maximally satisfied in a rational micro and macro environment. The efficiency of the realty utilization is evaluated with criteria of market, ownership, renewal prices, capacity, number of operations to be done while ownership transferring, safety, comfort, the time of physical and technical exploitations, *etc*. The software functions shift from "the most economically efficient control of the property" to multicriterial choice, raising computation efficiency.

For recent RS dealing with student's problem of choice the following papers were observed. In [23] an RS is constructed for helping entrants to choose an education direction (expressed as sets of corresponding high schools) according to his/her grades in GCE (General certificate of education), gained competing prizes, physical activity, and hobby. Three methods of content filtration were implemented: by distances between vector characteristics, axiomatic method of Pareto-set contraction, and analysis of hierarchies. Authors figured out coefficients relating study directions to the features of user using series of experimental assessments controlled by experts.

In [24], a problem of learning outcomes assessment of higher education students is considered. A course RS on the base students' graduate attributes, which describe their developing values, is proposed. RS rates improvement after each course and suggests new courses by a collaborative filtering algorithm in order to improve student's average competence profile. Students are presented as long vectors of courses they already taken with the corresponding grades. Similarity is calculated with a variation of angle cosine of the vectors. The recommendations are the prediction of the grades for the new courses. Similar by the problem statement paper [25], where RS predicts course learning trajectory patterns, the students is suggested elective curses. Source data for collaborative filtering are existing learning trajectories acquired from senior students in the image and likeliness one student advices younger one. Students and courses are divided on clusters. For each student cluster (a group), mean values are calculated. RS algorithms are based on combination of collaborative filtering and fuzzy-like rule based system, which defines process of production of a decision. The article has a good reviews of related works and three classical approaches basics.

Work [26] has the similar aim to develop RS for advising students new courses, but is focusing on description of courses, taking advantage of natural language processing over course documents to acquire descriptions. Feature descriptions contain formal course data (name, structure, lecturer) and a placement in a keyword appearance frequency space. Existing student data, represented as grades of the courses he/she already got, related to the course data. The resulting recommendations are represented as top-5 new courses, where student will gain the best grades. This is a simplistic direct approach and, by authors opinion, gives good results; some measurement supporting the opinion were carried on.

Interesting research has been done in [27], where the problem of overcoming the cold start problem in recommending genres and music compositions, as well as movies. The authors developed the RS "EZSurf" automating the process of web surfing and content filtering using a user profile on the social network "VKontakte", as well as API services `last.fm` and TheMovieDB, to obtain information about similarity of the objects. This approach greatly simplifies storage of RS data, since it does not require the designing own system of classifications. In [28], a problem of choosing top–N objects, with higher evaluations of user interest, in content filtration. Authors proposed RS model based on fuzzy sets, RS quality assessment technique, and a corresponding algorithm. The models have been tested on `last.fm` data.

In a comprehensive RS review dealing with text (scientific) document relevance grading [3], authors noted that more than half (55%, 34 of 62) RS were based on content filtration. Collaborative filtration was used only in 18% (11 of 62) cases. Stereotype based and hybrid methods are also presented. Authors concluded that 81% in of cases modeling user does not produce significant results in comparison to explicit indication the set of keywords. The papers were described by keywords and, less frequently, N–grams included in the text, as well as metadata such as authors and references. Most popular recommending model is based on vector space approach. User interest modeling is implemented with graphs, where vertices are the papers, and arcs are their relationships, and topic lists assigned to user by machine learning. The topics are organized in hierarchical catalogs using ACM classifiers. In the RS which used collaborative filtration, the explicit ratings was not collected at all as users were too lazy to supply a rating to a paper he/she looked through. Implicit ratings were obtained

by measuring the number of pages user read, document user interaction (loading, editing), co–loadings, co-view and co-citing by users of one group.

The collaborative filtration is currently more popular as it reflects practical experience: most of RS R&D have to solve "cold start" problem of initial lack of information on objects and users, as well as have adaptation to user group interest transformations due to modern trends.

IX. Discussion

Comparing these two approaches to the RS R&D ....

The first approach ... allowed collecting data in the runtime of the RS and also used J. Holland's theory, which has been used for decades for carer guidance, ...

The second approach ... is more useful for solving problems 1, 6,7,9 of the list of Section II.

X. Conclusion

In this brief narration of the results of two master theses, information recommender systems (RS) were developed to support the choice of a profession for students in Viet Nam and Irkutsk, Russia. The following tasks have been solved:

1. The subject area of RS, assistance in choosing a profession (carer guidance), have been analyzed; typical problems were recognized and their solution method were proposed.

2. The theory of John Holland in carer guidance was applied for high school students in Ki Lam High School, Ha Tinh, Viet Nam.

3. High school students of last courses tested the universities rating setting playing role of university students.

4. A methodology for calculating recommendations for carer guidance, based on the method of collaborative filtering and expert evaluation, is proposed. The technique has been implemented with Python modules.

5. Implemented an RS subsystems in the Python programming environment based on the proposed techniques and data structures; This Rs was tested by younger students of the Viet Nam school.

The resulting RS MVP[2], extended with automatic university data crawling and more efficient algorithms, is to be deployed in the school.

The source code is located at Github.com URL: https://github.com/tranducthe/diplom_CollaborativeFiltering.

Acknowledgements

The results were obtained within the state assignment of the Ministry of Education and Science of Russia, the project "Methods and technologies of a cloud-based service-oriented digital platform for collecting, storing and processing large volumes of multi-format interdisciplinary data and knowledge based on the use of artificial intelligence, a model-driven approach and machine learning", No. FWEW-2021-0005 (State registration No. 121030500071-2).

[2]Abbreviation of Minimal Valuable Product

References

[1] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich. Recommender Systems: An Introduction. Cambridge University Press, 2010.

[2] E. Charkashin, B. Shevchenko. Source code of a prototype real estate market recommender system. URL:https://github.com/eugeneai/RSPO-CSharp. (access date: 21-09-2020)

[3] J. Beel, B. Gipp, S. Langer, C. Breitinger. Research-paper recommender systems: a literature survey. International Journal on Digital Libraries (2016) 17: 305. DOI:10.1007/s00799-015-0156-0

[4] S.P. Balandina, M.V. Bautin, A.V. Maiyorov, E.R. Smirnova. Social networks as means of education and educational participants interactions. Pedagogicheskie i informacionniye tehnologii v obraxovanii. No. 15 (2016). (in Russian) URL:https://journals.susu.ru/pit-edu/issue/view/31

[5] K.S. Veber, A.A. Pimenova. Comparative analysis of social networks. Vestnik Tambovskogo universiteta. Series: Natural and technical sciences. Vol. 19, No. 2, 2014. p. 636–643. (in Russian)

[6] A.V. Rozhkova. Self-expression (self-presentation) in social Internet networks as a phenomenon of human cyber-socialization. Homo Cyberus. 2017. No. 2(3). (in Russian) URL:http://journal.homocyberus.ru/Samovyrazhenie_v_socialnyh_internet-setjah_kak_fenomen_kibersocializacii

[7] A.V. Fescshenko. Social networks in education: experience analysis and development perspectives. Gumanitarnaya informatika, Tomsk, Issue 6. 2012. p. 124–134. (in Russian)

[8] V.V. Matsuta, P.B. Kiselyev, A.V. Fescshenko, V.L. Goyko. Investigation of social networks potential for revealing talented students. Psychologiya i psihotehnika. No. 4. 2017. p. 104–121. (in Russian)

[9] Ch.C. Aggarwal Recommender Systems: The Textbook. Springer. 2016. ISBN 9783319296579.

[10] P. Brusilovsky. The Adaptive Web. 2007. p. 325. ISBN 978-3-540-72078-2.

[11] D.H. Wang, Y.C. Liang, D.Xu, X.Y. Feng, R.C. Guan. "A content-based recommender system for computer science publications", Knowledge-Based Systems. 2018. 157: 1-9

[12] S. Blanda. Online Recommender Systems – How Does a Website Know What I Want?. American Mathematical Society. 2015.

[13] J. Albahari, B. Albahari. C# 6.0 in a Nutshell: The Definitive Reference 6th Edition, 2015. 1138 p. ISBN-13: 978-1491927069

[14] Entity Framework Tutorial. URL:https://www.entityframeworktutorial.net/ (access date: 05.07.2020).

[15] VKontakte API documentation. URL: https://vk.com/dev/manuals (access date: 05.07.2020).

[16] D. Lemire, A. Maclachlan. Slope One Predictors for Online Rating-Based Collaborative. URL: http://cogprints.org/4031/1/lemiremaclachlan_sdm05.pdf (access date: 10.05.2018)

[17] A. Freeman. Pro ASP.NET MVC 5. 5th ed. Apress. 2013. 832 p. DOI:10.1007/978-1-4302-6530-6

[18] J. Chadwick, T. Snyder, H. Panda. Programming ASP.NET MVC 4: Developing Real-World Web Applications with ASP.NET MVC. O'Reilly Media; 1st Edition. 2012. 492 p. ISBN-13: 978-1449320317

[19] E. M. Alrawhani, H. Basirona, Z. Sa'ayaa. Real estate recommender system using case-based reasoning approach. Journal of Telecommunication, Electronic and Computer Engineering (JTEC). Vol. 8 No. 2. p. 177-182.

[20] E. V. Britina. Segmenting recommender system with client–server connection based on programmable configured networks and usage protocol with fast-jumping IP-address. // Sovremennye problemy nauki i obrazovaniya. No 6. 2015. URL:https://www.science-education.ru/ru/article/view?id=16875 (in Russian)

[21] X. Yuan, J.-H. Lee, S.-J. Kim, Y.-H. Kim. Toward a user-oriented recommendation system for real estate websites. Information Systems, 38 (2013). p. 231–243.

[22] T. Ginevičius, A. Kaklauskas, P. Kazokaitis, J. Alchimovienė. Recommender system for real estate management. Verslas: Teorija ir praktika (Business: Theory and Practice). 2011 12(3). p. 258–267 DOI:10.3846/btp.2011.26

[23] E. A. Belotskiy, A. V. Suetin, Construction of a recommendersystem for choosing higher education institutions for entrants, Vestnik S.-Petersburg Univ. Ser. 10. Prikl. Mat. Inform. Prots.Upr., 2016, Issue 1, 66–77. URL:http://www.mathnet.ru/links/1c01f91d57c8dc6f050c7e6d3fed3604/vspui277.pdf (in Russian)

[24] B. Bakhshinategh, G. Spanakis, O. Zaiane, S. ElAtia. A Course Recommender System based on Graduating Attributes. In Proceedings of the 9th International Conference on Computer Supported Education, Volume 1: CSEDU, ISBN 978-989-758-239-4, 2017, pp. 347-354. DOI:10.5220/0006318803470354

[25] A. Al-Badarenah, J. Alsakran. An Automated Recommender System for Course Selection" International Journal of Advanced Computer Science and Applications(ijacsa), 7(3), 2016. DOI:10.14569/IJACSA.2016.070323

[26] J. Naren, M. Z. Banu, S. Lohavani. Recommendation System for Students'Course Selection. A. K. Somani et al. (eds.), Smart Systems and IoT: Innovationsin Computing, Smart Innovation, Systems and Technologies 141. 2020. pp. 825–833. DOI:10.1007/978-981-13-8406-6_77

[27] B. R. Avhadeev, L. I. Voronova, E. P. Ohapkina. Development of recommender system on the base of profile in social network "VKontakte". Vestnik Nizhnevartovskogo gosudarstvennogo universiteta. Issue No 3. 2014. (in Russian)

[28] A. S. Amel'kin, D. M. Ponizovkin. Mathematical model of top-N problem for content recommender systems. Izvestiya MGTU "MAMI" No 3(17), 2013, Vol.2. p. 26-31. (in Russian)