Recommender System for Big Data in Education

Surabhi Dwivedi[1], Dr Kumari Roshni VS[2]
(1, 2) Centre for Development of Advanced Computing
Bengaluru, India
{surabhi[1], roshni[2]}@cdac.in

***Abstract*—With the advent of web based e-learning systems, a huge amount of educational data is getting generated. These massive data gave rise to Big data in educational sectors. Currently, big data analytics techniques are being used to analyze these educational data and generate different predictions and recommendations for students, teachers and schools. Recommendation systems are already very helpful in e-commerce, service industry and social networking sites. Recently recommendation systems are proved to be efficient for education sector as well. In this work we are using recommendation system for Big data in education. This work uses collaborative filtering based recommendation techniques to recommend elective courses to students, depending upon their grade points obtained in other subjects. We are using item based recommendation of Mahout machine learning library on top of Hadoop to generate set of recommendations. Similarity Log-likelihood is used to discover patterns among grades and subjects. Root Mean Square Error between actual grade and recommended grade is used to test the recommendation system. The output of this study can be used by schools, colleges or universities to suggest alternative elective courses to students.**

Keywords—Educational data mining; recommender systems; big data analytics;

I. INTRODUCTION

A huge amount of educational data is getting generated from massive open online courses (MOOC), online learning management system, web based educational resources, use of mobile apps in educational technology, use of social networking sites by students etc. These massive data cannot be handled using the traditional learning management systems. So it is required to handle and utilize such data using big data technologies. Big data is described as data sets with very big size that can not be handled by commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time [20].

The mining of educational data using data mining techniques gave birth to Educational data mining (EDM). Educational data mining deals with educational data and develops methods to explore those data to better understand the students and their learning environment [1]. EDM uses statistical, machine-learning, and data-mining algorithms over different types of educational data to improve the quality of education [22]. The main objective of education data mining is to find the patterns of system usage by teachers, students and to discover the students' learning behavior patterns [11]. Various methods like clustering, classification, relationship mining, prediction, discovery with models, and distillation of data for human judgment etc can be used for EDM [5].

The recommendation systems are very popular in e-commerce, entertainment; social networking and services like travel hospitality etc to recommend some items based on the preference of other similar users or based on their previous preferences of the same user. Recently recommender system have gained popularity in education sector as well to generate various kind of recommendations for students, teachers and schools etc.

This work is focused on the recommendation system techniques for big data, generated from educational sector. The system discussed in this paper uses big data technologies Hadoop and item based recommendation from Mahout, to generate recommendation for students to choose elective subjects. We are using similarity Log-likelihood to discover patterns among grades and subjects and Root Mean Square technique to test the model.

The paper is organized as follows. Section II introduces the literature survey; Section III gives an overview of recommendation systems and techniques of recommendation. Section IV gives an introduction to the relevant big data tools and techniques. The experimental set up for item based recommendation system using Mahout Machine learning library is explained in section V. Conclusion and future work is presented in section VI and acknowledgement in section at the end.

II. LITERATURE SURVEY

This section explains about the related work done for educational recommendation systems.

Collaborative filtering technique and content based method has been used by the author Mei-Hua Hsu in his work personalized English learning recommender system for students [14] to set basic score of lessons. Clustering technique is then used to classify students into various subjects. Finally association rule has been used to generate the recommendation for various lessons.

Educational data has been mapped to user/item by Nguyen, Lucas, Artus and Lars in their research work Recommender system for predicting student performance [17]. They are using matrix factorization technique to generate the recommendation and logistic regression to validate their approach.

An automated recommender system for course selection [3], uses collaborative recommendation technique to recommend elective courses to students by using association rule mining to generate course association rules. They are using precision and recall to evaluate the performance of the model.

978-1-5386-1922-3/17/$31.00 © 2017 IEEE

Boticario, Olga and Jesus have used a conceptual approach which can be used as personalized recommender in e-learning scenarios in their work Semantic educational recommender systems in formal e-learning scenarios [6].

Almost all the related work explored for this work do not discuss much about a model to generate the recommendations if data size is big. Since the size of data generated by current educational system is huge, we will use big data techniques Hadoop and Mahout to generate the recommendation for the students. We found collaborative filtering technique more suitable for our purpose, because we do not have much information about attributes of students and subjects. The number of items i.e subjects is relatively small as compared to number of users, i.e students. So we will be using item based collaborative recommendation technique.

III. RECOMMENDER SYSTEMS

In this section we will discuss about the recommender system techniques.

Fig. 1 shows the steps involved in recommendation system for educational data. First step is to obtain data from the educational resources. This involves integrating multiple databases, data cubes, files etc and removing the inconsistencies of data. Data and attribute redundancy is taken care at this stage. Next step is to select the required attributes of data using various feature selection and extraction techniques. This step reduces the volume of data such that analytical results are not affected. As data can originate from different sources, it is susceptible to noise, missing values and inconsistency. In order to get the correct result it is required to preprocess it. Following tasks is performed at data preprocessing stage

- Data Cleaning- Data is cleaned at this stage. Missing values are handled by filling in those values. Noisy data i.e those data which are not conveying meaningful information are handled using smoothing techniques. Smoothing techniques creates a function to extract meaningful information from data and ignore the noise.
- Transformation of Data- Data discretization or hierarchy generation can be used for data transformation. Data discretization is to convert expression data into finite value. Hierarchy generation reduces the data by converting low level concept for example age into higher level concept for example youth.

Next step is to use appropriate data mining technique to obtain the desired result. At this stage identified data mining technique is used to generate the recommendation for users.

The common tasks in recommender systems are Top-N item recommendation and rating prediction. The recommender suggests a ranked list of items to a user. Rating prediction predicts a preference score for a given user-item combination [16]

As explained in Fig-2, the recommendation techniques can be broadly categorized into collaborative filtering [13],

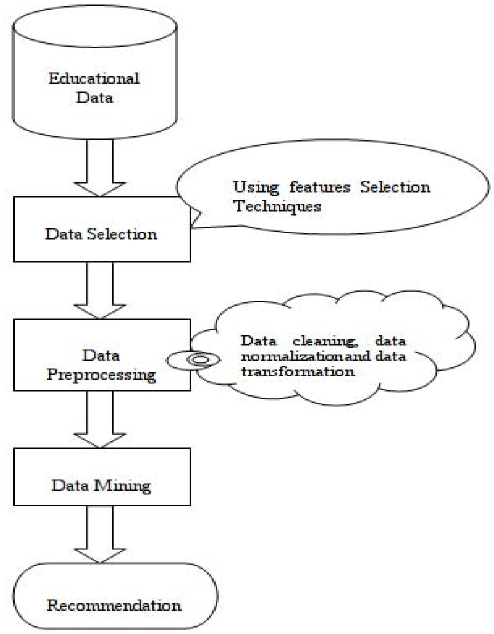

Fig -1 Steps in Recommendation system

content-based filtering [2] knowledge-based system [7], hybrid systems etc.

A. Content Based Recommendation -

Content-based recommendation systems find out items of interest for users by analyzing item descriptions [18]. These systems generate list of item profiles for the users, based on data provided by users. It uses a metric of term-frequency (TF) and inverse document frequency (IDF). The product of TF*IDF is used to identify the importnace of the item. Term-frequency determines how many times the item is occurring in a document. IDF identifies the importance of the item. It is log of the ratio of total number of documents and number of documents containing the item.

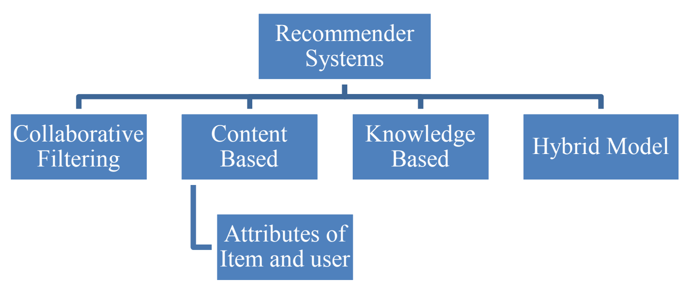

Fig 2. Classification of recommender systems

After calculating TF*IDF, vector space model is used to identify preference of a user to an item. In vector space model, vector is created for the item and corresponding attributes. Users vector is also created based on his preference on the attributes of the items. Finally cosine angle between these vectors are calculated to identify the similarity between user and item vectors. Correlation based approach can also be used for content based recommendation [2].

B. Collaborative filtering based recommendation

These recommendations are based upon preference of similar users. Item will be recommended to the user based on the preference of other similar users for the same item. If set of users have strongest correlation in the past, they will be identified as 'nearest neighbor'. Score of the new items will be predicted based upon the scores of nearest neighbor. Many collaborative filtering algorithms exists for data sets where number of items are less than number of users [13]. In collaborative filtering Pearson correlation or Log-likelihood ratio can be used to identify preferred items for the user. Pearson correlation can be calculated as following [19]

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_j)^2}}.$$

where $R_{u,i}$ is the rating given user u to item i; $\bar{R}i$, $\bar{R}_j$ is the average rating of i^{th} and j^{th} item . The log-likelihood ratio (LLR) measures of how two events occur together, more than once which are unlikely to be independent [23].

Item based recommendation of collaborative filtering is purely based upon ranking given by a user to a particular item whereas content based recommendations are focused on attributes of the items as well as attributes of users.

C. Knowledge Based Recommendation Sytem

In these systems knowledge about the user's need, his preferences etc are used for recommendation. These systems are based upon the knowledge of user's need for a particular item and can therefore reason about the relationship between a need and a possible recommendation [7]

D. Hybrid Recommendation Sytem

The features of content based, collaborative filtering and knowledge based systems can be combined together to improve the recommendation accuracy. Burke describes hybrid recommender systems based on the hybridization methods [7]. The weighted hybridization method combines the scores of several recommendation techniques together. Switching hybrid system switches to different recommendation techniques based on the current situation. Feature combination combines features from different recommendation data sources into a single recommendation algorithm. Cascade recommendation uses series of recommendations, the recommendations given by one recommender will be refined by the other. Feature augmentation uses output from one technique as an input feature to another. In meta-level system, model learned by one recommender acts as input to another.

IV. TOOLS AND TECHNIQUES

This section describes about few of the big data tools like Hadoop, Mahout , MLlib, R, Python – Crab etc that can be used to set up a recommendation system for big data.

A. Hadoop

Hdoop is an Apache project. Hadoop framework allows the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage [12]. The main components of Hadoop are as following

- Hadoop Common supports the other Hadoop modules.
- Hadoop Distributed File System (HDFS) provides access to application data and high-throughput.
- Hadoop YARN is a framework for scheduling the job resource management of the hadoop cluster.
- Hadoop MapReduce is based on YARN (Yet Another Resource Negotiator) system. It provides parallel processing for large data sets.

B. Mahout

Apache Mahout is an open source, scalable machine learning library from Apache Software foundation. It implements many machine learning algorithms in different categories like collaborative filtering, clustering, classification and dimensionally reduction [16]. Mahout also makes use of Hadoop for processing big data.

C. MLlib By Spark

MLlib is developed as part of the Apache Spark project. Spark is an engine for big data processing. It has two packages spark.mllib and spark.ml [21]. MLlib is a machine learning (ML) library which includes algorithms such as classification, regression, clustering, collaborative filtering, and dimensionality reduction. It is new as compared to other data mining and processing tools. The data processing using MLlib is very fast as the processing takes place in memory.

D. Crab- A Recommender framework in Python

Crab - scikits.recommender is a recommender engine for Python. Crab is a Python framework for building recommender engines integrated with the scientific Python packages (numpy, scipy, matplotlib). It has support for recommender algorithms like user-based and item-based collaborative filtering [24]. The recommender interfaces can be easily combined with more than 10 different metrics, like cosine, tanimoto, pearson, euclidean using, Scipy and Numpy based optimized functions [28].

E. R

R is an integrated suite of software facilities for data manipulation, calculation and graphical display under GNU General Public License (GPL). R is the most popular tools as per KD nuggets software poll for Data Mining for year 2016 [27]. It provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering etc.) and graphical techniques [26]. R supports numerous data mining algorithm and can also be easily integrated with Hadoop for processing big data. R can be used for recommendation using association rules, collaborative techniques etc [25].

V. EXPERIMENTAL SETUP AND RESULTS

We have used a sample data from different schools of Central Board of Secondary Education (CBSE) across India. We assume that data size will grow huge if collect data from all schools across India, so we have used Hadoop framework to handle the data. The data set has list of grade points of students in different subjects, which can be compared with preference of items for users in item based collaborative filtering recommendation system. We have used item based recommendation technique from Mahout machine learning library. Mahout recommender model uses the input data in a format (user, item, value) triple in a preference object.

We used Mahout 0.10, Hadoop 2.6.0, Personal computer (Ubuntu 14.04.3, 64-bit, RAM 16 GB, CPU: Intel Core i7-4770) for experimental set up.

The steps to generate the recommendations are as follows, first start the Hadoop clusters (dfs and yarn) and then copy the required input file (for which the recommendation has to be generated) to Hadoop file system and finally generate the recommendations using Mahout. The parameters used for Mahout based recommendation system is shown in Table I. The parameter recommenditembased is for item based recommendation. SIMILARITY_LOGLIKELIHOOD signifies log of the probability that subject will be recommended. –I and –o are path to input and output files.

TABLE I. PARAMETERS USED FOR RECOMMENDATION SYSTEM IN MAHOUT

Parameters	**Description**
recommenditembased	For item based recommendation
-s SIMILARITY_LOGLIKELIHOOD	Depending upon the grade points, log of the probability that the subject will be recommended
-i hdfs://localhost:9000/student_data/student_data.csv	Path to input file at HDFS
-o hdfs://localhost:9000/student_test/	Path to output file at HDFS
--numRecommendations 25	number of recommendations per student

```
1100001 [303:9.0,2:9.0,18:9.0,122:9.0,10:8.715225,184:8.674078,165:8.429365]
1103899 [165:6.737872,303:6.5180254,10:6.348352,18:6.2804165,122:6.240944,184:6.232204,85:6.0328064]
1104009 [85:8.714407,165:8.6246,184:8.611737,18:8.432377,10:8.4217615,303:8.413658,2:8.318124]
1104088 [184:6.132938,10:5.9831204,2:5.8276424,303:5.586342,18:5.567623,122:5.5203714]
1100004 [18:9.73358,122:9.713336,2:9.690513,184:9.628476,303:9.586342,10:9.49252,165:9.375459]
1100005 [184:8.481094,2:8.45707,18:8.287207,10:8.16442,122:8.156421,165:8.112036,303:8.068316]
1100006 [303:8.518025,18:8.2804165,10:8.24915,122:8.240944,2:8.224806,165:8.173222,184:8.0]
1100007 [303:8.0,2:8.0,18:8.0,122:8.0,10:7.715225,184:7.6740785,165:7.429365]
1100008 [165:6.570635,184:6.3259215,10:6.284775,18:6.0,303:6.0,2:6.0,122:6.0]
1100009 [122:7.315971,18:7.1659565,184:7.1473813,165:7.090202,10:7.0789495,2:7.008636,303:7.0]
1100010 [303:7.1043673,165:7.029113,10:6.947496,2:6.9066825,18:6.8480396,184:6.8070164,122:6.638309]
1100011 [184:7.9543977,165:7.9460936,10:7.777295,18:7.73358,122:7.713336,2:7.6905127,303:7.586342]
1100012 [165:6.8558917,184:6.8070164,10:6.6983457,2:6.681876,303:6.586342,18:6.567623,122:6.397365]
1100013 [184:6.9543977,165:6.9460936,10:6.777295,18:6.73358,122:6.713336,2:6.6905127,303:6.586342]
1100014 [303:9.0,2:8.991364,10:8.92105,165:8.909799,184:8.852618,18:8.8340435,122:8.68403]
1100015 [2:7.681876,303:7.586342,18:7.567623,184:7.481095,10:7.413571,122:7.397365,165:7.285257]
1100016 [2:7.7665577,18:7.553627,184:7.5266967,303:7.4819746,122:7.443085,10:7.3871255,165:7.165942]
1100017 [303:7.0,2:6.991364,18:6.8340435,122:6.684029,10:6.636276,184:6.5266967,165:6.3391633]
1100018 [303:7.5180254,18:7.2804165,10:7.2491503,122:7.240944,2:7.2248063,165:7.173221,184:7.0]
1100019 [165:6.743856,10:6.533925,303:6.5180254,184:6.3259215,18:6.2804165,122:6.240944,2:6.2248063]
```

Fig . 3. Sample output file generated by the system

The input file contains student's roll number, subject id and the grade point in the particular subject. The format of input file is similar to the format required for class FileDataModel of Mahout. The roll number corresponds to user ID, subject id corresponds to item ID and grade point corresponds to preference.

We are using only few samples of input and output data for illustration purpose here. Table-II shows sample records of five students containing different grade points in different subjects. The output file generated by the system contains two columns roll number and array of subject id's with their corresponding grade points as shown in Fig 3.

Table -III shows the part of the output file generated by recommender system for the same five students. The output can be further illustrated as follows, if we assume that initially the student with roll number 1100001 has opted for five subjects with id's 101, 85, 86, 87 and 41. If he wants to choose other subjects the output of recommender system can be used depending upon the grade recommended by the system. If student wants to choose other subjects, the recommender system recommends subject id 303, with grade point 9.0, subject id 18, with grade point 9.0, subject id 165 with grade point 8.5, subject id 10 with grade point 8.7 and so on. On observation it is found the score of recommendation for subject varies among the students. For example, subject id 303 is highly recommended with grade point 9 to student with roll no 1100001, whereas for student with roll no 1103899, it is recommended with lesser grade point of value 6.5.

In order to test the recommendation system, we removed the data corresponding to subject id 87. We then run the recommendation system on the remaining data set to evaluate the variation between actual grade points obtained and the grade points recommended by the system. Table IV shows the difference between actual grade and recommended grade for same set of students as taken previously to generate recommendation. It also calculates root mean square for these samples.

TABLE II. SAMPLE INPUT FOR THE RECOMMENDER SYSTEM

Roll_no	sub_id	grade_point
1100001	101	9
1100001	87	9
1100001	86	9
1100001	85	8
1100001	41	9
1103899	87	6
1103899	86	6
1103899	184	6
1103899	101	7
1103899	2	7
1104009	87	8
1104009	165	8
1104009	86	9
1104009	41	9
1104009	122	9
1104088	85	7
1104088	165	6
1104088	87	5
1104088	101	6
1104088	303	5

TABLE III. SAMPLE RECOMMENDATION

Roll_No	**Sub**	**GP**	**Sub**	**GP**	**Sub**	**GP**	**Sub**	**GP**	**Sub**	**GP**
1100001	303	9.0	18	9.0	10	8.7	122	9.0	165	8.5
1103899	165	6.8	303	6.5	10	6.3	18	6.2	184	6.2
1104009	85	8.7	165	8.6	184	8.6	18	8.4	303	8.4
1104088	184	6.2	10	5.9	2	5.8	303	5.5	122	5.5

To evaluate the result we are using Root Mean Square Error. The value of Root Mean Square Error between actual and recommended data is much less than 0.5. It indicates good fit of the generated recommendations. There is acceptable level of difference between recommended grades and the actual grades. So the above method can be used to generate recommendation for students to choose elective subjects. Since we are using big data technologies, it will be efficient even if data size is huge. The administration can decide a minimum threshold of the grade point suggested by the recommendation system. Then all the subjects, recommended by the system, whose grade points are above the threshold level, can be suggested to students to take as an alternative elective subject.

TABLE IV. TEST OF RECOMMENDATION SYSTEM

Roll No	1100001	1103899	1104088	1104009
Actual grade	9	6	5	8
Recommended grade	8.8	6.4	4.6	8.7
Difference (Actual - recommended)	0.2	-0.4	0.4	-0.7
Absolute Average Difference	0.425			
Root mean square	$\sqrt{((.04+.16+.16+.49)/4)} = 0.46$			

VI. CONCLUSION AND FUTURE WORK

Recommendation system can be proved to be very helpful to students to select the elective courses. The educational institute can design their syllabus to give more options to the students to choose subjects according to the specific skills and expertise of the students. Big data comes up with its own challenge to handle the data, but if it is appropriately managed, it can be beneficial to improve the quality of current education system and process. The suggestions generated by such systems can be useful to the educational institute to improve the performance of students, schools and teachers etc.

We used collaborative filtering, item based recommendation system to suggest elective courses to a student. These recommendations are based on his grade obtained in other subjects. Through this work we were able to identify applicability of recommender system for huge size of educational data. We also identified how education data can be mapped to item in recommender systems.

We are planning to use hybrid model based recommendation system in future to enhance the accuracy of the recommendation. We are also planning to use recommender systems to suggest career options to college students, based on some historic data of previous batch students.

ACKNOWLEDGEMENT

We are thankful to MEITY and CBSE for giving us an opportunity to work on the project e-Governance Analytics Platform (e-GAP), during the research phase of the project we could test all these experiments.

We are grateful to Dr N Sarat Chandra Babu, Executive Director Centre for Development of Advanced Computing (C-DAC) Bangalore, to give us an opportunity to perform this study

REFERENCES

[1] www.educationaldatamining.org (accessed April 2017).

[2] Adomavicius G., Sankaranarayanan R., Sen S., and Tuzhilin A. "Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach." *ACM Transactions on Information Systems Vol. 23, No. 1*, January 2005: 103–145.

[3] Amer Al-Badarenah, Jamal Alsakran. "An Automated Recommender System for Course Selection ." *(IJACSA) International Journal of Advanced Computer Science and Applications,*, 2016: 166-175.
[4] Baker, E. *International Encyclopedia of Education (3 rd edition).* Oxford, UK: Elsevier, 2010.
[5] Baker, R.S.J.d. (2010) Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education (3rd edition)*, vol. 7, pp. 112-118. Oxford, UK: Elsevier
[6] Boticario, Olga C. Santos and Jesus G. "Requirements for Semantic Educational Recommender Systems in Formal E-Learning Scenarios." *Algorithms*, 20 July 2011: 132-154.
[7] Burke, Robin. "Hybrid Recommender Systems: Survey and Experiments." *User Modeling and User-Adapted Interaction Volume 12, Issue 4*, November 2002: pp 331–370.
[8] Burke, Robin. "Integrating Knowledge-Based and Collaborative-Filtering ." *AAAI Workshop on AI in Electronic Commerce,*, 1999: 69-72.
[9] Burke, Robin. "Knowledge-based recommender systems." In *Encyclopedia of Library and Information Science*, 180-200. 2000.
[10] "Content-Boosted Collaborative Filtering for Improved Recommendations." *AAAI -02 Proceedings.* Canada, 2002. 187-192.
[11] Félix Castro, Alfredo Vellido, Àngela Nebot and Francisco Mugica. "Applying Data Mining Techniques to e-Learning Problems." *Evolution of Teaching and Learning Paradigms in Intelligent Environment.* Springer. pp 183-221.
[12] *Hadoop.* http://hadoop.apache.org/ (accessed Apri 2017).
[13] Herlocker J., Konstan J., Tervin L., and Riedl J. "Evaluating Collaborative Filtering Recommender Systems." *ACM Transactions on Information Systems*, 2004: 5-53.
[14] Hsu, Mei-Hua. "A personalized English learning recommender system for ESL students." *Expert Systems with Applications*, January 2008: 683–688.
[15] Juntae, Choonho Kim. "A recommendation algorithm using multi-level association rules." *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003).* 2003. 524-527.
[16] *Mahout.* http://mahout.apache.org/users/basics/algorithms.ht (accessed April 20, 2017).
[17] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe , Lars Schmidt-Thieme. "Recommender System for Predicting Student Performance." *1st Workshop on Recommender Systems for Technology Enhanced Learning, Procedia Computer Science* 2010: 2811-2819.
[18] Pazzani M.J., Billsus D. "Content-Based Recommendation Systems." *The Adaptive Web. Lecture Notes in Computer Science, vol 4321.* Berlin, Heidelberg: Springer, 2007. pp 325-341.
[19] Sarwar, Badrul and Karypis, George and Konstan, Joseph and Riedl, John. "Item-based collaborative filtering recommendation algorithms." *Proceedings of the 10th international conference on World Wide Web.* Hong Kong, Hong Kong: ACM, May 01 - 05, 2001 . 285-295.
[20] Snijders, C., Matzat, U., & Reips, U.-D. "Big Data': Big gaps of knowledge in the field of Internet science." *International Journal of Internet Science*, 2012: 1-5.
[21] *Spark MLlib.* http://spark.apache.org/docs/latest/mllib-guide.html (accessed April 25, 2017).
[22] Ventur, C. Romero and S. "Educational Data Mining: A Review of the State of the Art." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Nov 2010. vol. 40, no. 6, pp. 601-618.
[23] Isoni, Andrea, " Machine learning from Web", Packt Publishing Limited
[24] http://muricoca.github.io/crab/tutorial.html (accessed May 2017).
[25] Greg Linden, Brent Smith, and Jeremy York. "Amazon.com Recommendations Item-to-Item Collaborative Filtering." In IEEE INTERNET COMPUTING, 76-80. Published by the IEEE Computer Society, 2013.
[26] Introduction to R. https://www.r-project.org/about.html (accessed January 2016).
[27] "KD nuggets software poll for Data Mining." Analytics software/tools used in past 12 months. http://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html (accessed April 2017).
[28] Marcel Caraciolo, Bruno Melo, Ricardo Caspirro. "Crab: A Recommendation Engine Framework for Python." 10th PYTHON IN SCIENCE CONFERENCE. Austin, Texas, 2011. 11-17.