The Eurasia Proceedings of
Educational & Social Sciences (EPESS)
ISSN: 2587-1730

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2016

Volume 5, Pages 309-312

ICRES 2016: International Conference on Research in Education and Science

NEW RECOMMENDER SYSTEM USING NAIVE BAYES FOR E-LEARNING

Mehmet Ozcan
Anadolu University

Tansu Temel
Anadolu University

Abstract: Coming into prominence at the present time, e-learning is a great opportunity for learners. It provides tremendous assets most valuable of which is distance free learning. Besides, there is a great deal of e-learning resources on the web that causes information overload. Accordingly, it turns into a requisite that you ask for recommendation so as to find the resource you surely need. There are readily available recommendation services arranged for that purpose. Such systems have various rating systems; furthermore users tend to rate the materials in different manners. Our goal with this paper is to generate confidential referrals thanks to Naive Bayesian algorithm for e-learning materials rated multifariously by learners. We also researched the effects of several data preprocessing techniques on achieving this goal.

Keywords: Naive bayes, data preprocessing, e-learning, recommendation systems

Introduction

The purpose of a recommender system is to generate meaningful recommendations to a collection of users for items or products that might interest them (Melville & Sindhwani, 2011). Recommender systems have a wide usage area in our daily life a significant example of which is e-learning. E-learning has become more of an issue recently. Distance education supports traditional education yet more began to replace it. Besides, amount of resources related with e-learning is huge. Given the increasing number of e-learning platforms, learners are often overwhelmed with the large amount of learning resources available online (Souali, El Afia, Faizi, & Chiheb, 2011). Therefore, having a right material in right time is also difficult. Our goal in this paper is to implement Naïve Bayes algorithm for e-learning materials rating from learners with different ways. Several data preprocessing operations are applied before applying Naïve Bayes Classifier. The vestigial of this paper is regulated as follows. Section 2 presents related works. Section 3 exhibits proposed architecture. Section 4 includes experimental results. Section 5 gives a short conclusion and future works.

Related Work

In this section we present some of the research literature related with e-learning recommender systems. Bayesian Network is utilized in order to detect learner's learning style and discover their preferences (Carmona, Castillo & Millán, 2007; García, Amandi, Schiaffino & Campo, 2007). Ueno and Toshio (2007) created learner model via Bayesian Network. Using the learner model, learner's final status (Failed, Abandon, Successful, Excellent) is predicted. Next, active learner's learning processes are compared with excellent learners' learning processes, and appropriate messages to the learner are generated. Colace and De Santo (2010) studied on the role of ontologies in the context of e-learning. A novel algorithm for ontology building with Bayesian Networks is presented in their work. Analyzing students' learning performances, their proposed method can analyze the courses' ontology and propose corrective actions. Thus, teachers better understand the requirements of their students and can redesign their courses appropriately. Moreover, an ontological basis is provided to determine learning paths to personalize learning. Chang, Kao, Chu and Chiu (2009) proposed a learning style classification mechanism to

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: Mehmet Ozcan E-Mail: mehmet_ozcan@anadolu.edu.tr

classify and identify students' learning styles. The proposed method improves k-nearest neighbor classification and combines it with genetic algorithms. The proposed method is implemented on an open-learning management system. García Amandi and Schiaffino (2008) detected a student's learning style automatically from the student's actions in an e-learning system using Bayesian Networks. E-teacher uses the information contained in the student profile to proactively assist the student by suggesting him/her personalized courses of action that will help him/her during the learning process. Ozpolat and Akar (2009) addressed the problem of extracting the learner model based on Felder–Silverman. Using Naïve Bayesian Tree in conjunction with Binary Relevance classifier, the learners are classified according to their interests. Learners' learning styles are defined using these classification results.

Proposed Architecture

Our proposed architecture basically aims to produce effective predictions with Naïve Bayes Classifier for e-learning systems. It is necessary to have a quality dataset in order to get efficient predictions. Thus, we take the advantage of some preprocessing operations. Researchers also slog on maintaining their studies at generating predictions from binary data because of not having a binary dataset with great amount of data. They are constrained to convert continuous or discrete datasets into binary datasets. In these conversions, researchers make some assumptions to decide the rating scales to be converted into 'true' and 'false'. If we notate the possible minimum and maximum ratings as R_{min} and R_{max} respectively, common technique is selecting a threshold value t as $(R_{min}+R_{max})/2$ then converting the ratings greater than t as 1 and less than t as 0. In a 1~5 rating scenario, converting 1, 2, 3 into 0 and 4, 5 into 1; is another frequently used technique. We proposed some new approaches to convert continuous data into binary data in the hope of creating more accurate predictions.

Experimental Work

Datasets & Evaluation Metrics

In this paper we use jester data set because we do not have a real e-learning data set with continuous values. This dataset contains 4.1 million continuous ratings (-10.00 to +10.00) of 100 jokes from 73,496 users. We present our solution with accuracy, specificity, precision, recall, f-score and g-mean metrics.

Experimental Design

We select a subset containing 1000 users each rated 100 items. We get rid of the missing values by filling them with the mean of overall ratings. Then we selected 5 random items for each user and produced predictions for them with leave-one-out technique.

Accuracy & Performance Analysis

We conducted 3 groups of experiments which have different concepts to convert discrete ratings into binary.

In the first group; we make the conversion with respect to quartiles of the rating domain. The quartiles are selected as thresholds and higher and lower values are converted to 1 and 0 respectively. For the selected dataset, values of Q1, Q2 and Q3 are -5, 0 and 5 respectively. As mentioned before, Q2 refers to the widespread approach. The results are shown in Table1.

In the second group; conversion is made considering the ratings in the dataset. Consecutively, thresholds are defined for each user and each item separately so the conversions are made separately as well. Besides, threshold value is assigned to overall mean of the ratings and conversion is made with that threshold for all ratings. These results are shown in Table2.

Table 1: Performance of quartiles

	Accuracy	**Specificity**	**Precision**	**Recall**	**F-Score**	**G-Mean**
Q1	0.8412	0.5132	0.9382	0.8803	0.9083	0.6721
Q2	0.7642	0.8071	0.6375	0.6785	0.6574	0.7400
Q3	0.8210	0.8532	0.3869	0.6091	0.4732	0.7209

In the third group; same technique is used for the conversion but with their absolute values. Results of these experiments are demonstrated in Table3

Table 2: Performance of means

	Accuracy	Specificity	Precision	Recall	F-Score	G-Mean
UM	0.7552	0.7897	0.6648	0.6975	0.6808	0.7422
IM	0.7650	0.8092	0.6946	0.6942	0.6944	0.7495
OM	0.7732	0.8177	0.6469	0.6823	0.6641	0.7469

Table 3: Performance of absolute means

	Accuracy	Specificity	Precision	Recall	F-Score	G-Mean
AUM	0.8030	0.8356	0.4567	0.6499	0.5365	0.7369
AIM	0.7954	0.8312	0.4894	0.6513	0.5589	0.7358
AOM	0.8014	0.8387	0.4938	0.6479	0.5604	0.7371

According to the experimental results, F-score of the Q1 results the best among all techniques. Here Q1 can be thought as an outlier because of the characteristics of the dataset. As we can see from the tables, techniques used in the second group of experiments are all resulted in better f-scores than Q2 while techniques in third group of remain deficient according to f-scores. In the view of g-means, techniques in second group can be selected as the bests. Specificity remains stable in the third group of techniques which may be used for different purposes. Q3 has the best accuracy except Q1 but, there is a critical fall in f-score. Hence we may select the techniques used in the third group for accuracy concerns.

Conclusion

E-learning is a leading practice for every kind of learners with its tremendous opportunities. On account of the fact that there is a huge amount of e-learning resource on the web, it is inevitable to benefit from a recommender system in order that one can determine the right material to study. We proposed to take advantage of Naïve Bayesian algorithm to achieve this goal. Our study includes the evaluations of several data preprocessing operations applied in continuous to binary conversion step. It is inferred from the results that preprocessing techniques considering the rating means are the best regarding f-measure. The other preprocessing techniques can be preferred to apply through different aspects.

Experiments in this study are held on a different kind of dataset instead of a real e-learning one. In future work, we desire to use real e-learning data set with continuous and discrete values, and improve our approach in this way.

References

Carmona, C., Castillo, G., & Millán, E. (2007). Discovering student preferences in e-learning. In *Proceedings of the international workshop on applying data mining in e-learning* (pp. 33-42).

Chang, Y. C., Kao, W. Y., Chu, C. P., & Chiu, C. H. (2009). A learning style classification mechanism for e-learning. *Computers & Education*, *53*(2), 273-285.

Colace, F., & De Santo, M. (2010). Ontology for E-learning: A Bayesian approach. *Education, IEEE Transactions on*, *53*(2), 223-233.

García, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles.*Computers & Education*, *49*(3), 794-808.

Melville, P., & Sindhwani, V. (2011). Recommender systems. In *Encyclopedia of machine learning* (pp. 829-838). Springer US.

Özpolat, E., & Akar, G. B. (2009). Automatic detection of learning styles for an e-learning system. *Computers & Education*, *53*(2), 355-367.

Schiaffino, S., Garcia, P., & Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students. *Computers & Education*, *51*(4), 1744-1754.

Souali, K., El Afia, A., Faizi, R., & Chiheb, R. (2011, April). A new recommender system for e-learning environments. In *Multimedia Computing and Systems (ICMCS), 2011 International Conference on* (pp. 1-4). IEEE.

Ueno, M., & Okamoto, T. (2007, July). Bayesian agent in e-learning. In*Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on* (pp. 282-284). IEEE.