A Recommendation System for Shared-Use Mobility Service through Data Extracted from Online Social Networks

Eduardo Lucio Lasmar Junior, Renata Lopes Rosa, and Demóstenes Zegarra Rodríguez, Senior Member, IEEE

Abstract—In recent years, the shared mobility service has increased in many countries across the world because of its low cost and several shared-use mobility applications on mobile devices [1]. Commonly, if a ride is shared between people with similar preferences, users likely feel both more comfortable and safe. In this context, the main goal of this article is to classify users with similar preferences, in automatic manner, to improve user’s quality of experience in ridesharing service. To obtain initial data, subjective tests are carried out using questionnaires and their results are used to determine ridesharing profiles. Then, some basic user profile information is extracted from Online Social Networks (OSN) to determine an user profile based on preferences in ridesharing service. The user profile classification is performed through different machine learning algorithms, which use as input the data extracted from OSN. Two case studies of shared-mobility are treated, (i) sharing a ride with a passenger with a similar hobby [2], and (ii) sharing a ride with people that support an opposite football team. The novelty of this work consists in the application of the Hybrid Discriminative Restricted Boltzmann Machines (HDRBM) algorithm for classification whose performance outperforms widely used algorithms such as Random Forest, SVM and DRBM. The experimental results present a correctly classified instance of 96.9% and 97.3% for the cases of sharing a ride with people with similar hobby and support different football team, respectively. Finally, a Recommendation System (RS) is proposed, which efficiency is compared with a basic RS, obtaining a Pearson correlation coefficient of 0.97 and 0.71, respectively.

Index Terms—Recommendation system; ridesharing; mobile applications, online social networks, machine learning, HDRBM, social web analysis tool.

I. INTRODUCTION

The mobile devices and applications have facilitated the manner of people request services. The urban mobility is a kind of service that became very popular, in it is common that people share the rides among others. The advantages of this service are related to benefits for urban traffic, such as reduction of environment pollution, and low cost.

The sharing-based consumption reinforces value to experience over the notion of possessing something. Currently, people are looking for more goods and services that a commercial exchange offers, they seek to satisfy their real needs through intangible experience. Collaborative consumption is the reinvention of old market behavior, sharing and exchange, trade and rent, through new technologies in scale and forms never before possible [3].

The ridesharing service takes advantage of the comfort and availability of some private vehicles, reducing the costs related to a private transport service. The transformations of the conventional transport modes have gradually created a shared transport market and the concept of shared economy [4] in case of cars and houses.

In a service available for the network not only the Quality of Service is important [5], [6] but also the user Quality of Experience (QoE) [7] is an important factor to be considered. The users QoE is a measure used to evaluate the degree of pleasure, or discomfort, when trying out a product or service. Hence, QoE concept can be applied in ridesharing service.

Many people prefer sharing services with users of similar characteristics, according to studies [8]. For suggestion of sharing services exist the Recommendation Systems (RS), which are a mechanism that includes a database of user characteristics and the services or products available to the users [9].

RS solutions have been very used in last decade, both in academic and commercial areas. Among the benefits of RS we can highlight the capacity to offer the users more personalized suggestions [10] and it can be also used for ridesharing suggestions. It is important to note that for an appropriated suggestion, the user data need to be available. In this case, the online social networks (OSN) [11] have been very popular among people, playing a key role in RS solutions. The proper treatment of this information can provide useful data to build models or user profiles, and thus suggest content with a higher accuracy.

The data extracted from OSN associated to the number of urban mobility applications have created and ideal scenario and opportunities to improve the existing RS. A RS is studied in [12], however, authors do not study the user preferences according to his/her personal characteristics, such as age and gender.

The machine learning algorithms have been used for discov-
ering or classifying user characteristics. Currently, the Convolutional Neural Network (CNN), the Restricted Boltzmann Machine (RBM) and their variations are very popular methods used in profile classification [13], presenting confident results.

This paper is an extended contribution to the work [2], in which a RS is performed based on the passenger preferences to share a ride with another potential passenger with same or different hobbies using the Discriminative Restricted Boltzmann Machine (DRBM) algorithm. In the present paper we evaluate and additional case study and an additional machine learning algorithm. Thus, the proposed RS will be able to obtain better performance results. In general, the new contributions of this work are stated as follows:

- The inclusion of a scenario in which a passenger shares or not a ride among potential passengers of opposite football teams, in order to evaluate the impact on users’ QoE.
- The use of Hybrid Discriminative Restricted Boltzmann Machine (HDRBM) algorithm, to improve the RS performance in ridesharing service.
- The use of only basic user profile characteristics, such as age, gender and relationship, which are automatically extracted from OSN. Hence, the proposed RS is more useful and easier to be implemented of ridesharing service providers.

In this paper, the performance of the proposed RS that uses HDRBM algorithm for classification is evaluated, in both study cases, if the user agrees to share a vehicle with people have a different hobby and with people that support opposite football clubs. Initially, subjective tests are performed [14] in which users answer some questions about his or her preferences regarding the ridesharing service. The users responses are used to build a user profile model for ridesharing. In the RS, the model is used to suggest passengers who could share the service. Experimental results show that the correct classification performed by HDRBM algorithm obtained 96.9% and 97.3% for the option of sharing a ride with a person with similar hobby and sharing a ride with a person with different football team, respectively. These results overcome those reached by SVN, Random Forest and DRBM algorithms. In a cultural or sport events is common people share a ride, but it is necessary to know user’s preferences on ridesharing.

Finally, additional subjective tests were performed in order to evaluate the performance of the proposed RS in relation to a basic RS that not consider user preferences on ridesharing service.

It is important to note that, currently, the existing related studies do not cover an RS with focus on increasing user satisfaction or user’s QoE, using deep learning algorithms and with focus on ridesharing for cultural and sport events.

The remainder of this paper is structured as follows. Section II presents an overview of sharing services, HDRBM, RS, and data extraction from OSN are treated; also, related works are presented. Section III presents the architecture of the proposed RS for shared-use mobility. Section IV presents the methodology followed to build the proposed RS. Section V presents the results and discussions. Finally, the conclusions are presented in Section VI.

II. MAIN CONCEPTS DEFINITIONS AND RELATED WORK

This section presents the main concepts definitions and studies about sharing services, Machine Learning Algorithms and Discriminative Restricted Boltzmann Machine, recommendation systems and OSN data extraction.

A. Sharing Services

The consumer population of digital applications have increased their interest in the shared services systems. This fact has caused financial and social changes in the world. The increased use of the OSN has allowed the access to groups, in which the sharing of similar interest is present. Thus, new solutions about sharing services have been occurring, bringing the costs saving. These new solutions have affected both the market and individual interests [15].

The shared economy emerged according to the sharing services, which are related to the practice of sharing, exchange, loan, rent, and donation [16]. This kind of economy has many advantages, such as the reduction of costs in general. The service offered by Airbnb is an example of sharing services, in it a sharing platform for hosting enables worldwide sharing, increasing the attention to collaborative activities [17].

In the same context, the urban mobility emerged thanks to the automation of vehicles and other innovation about the vehicle market. Some definitions of the models of ridesharing and ridehailing are presented in [18], which occurs the sharing of a ride among different users. A study was performed in [19], in which, the question “If you use ridehailing services, how will your usage evolve in the next 2 years?” was used in the tests. In the results, more than 60% of respondents believe that this model will increase. Additionally, according to [20] the traditional models of mobility, such as buses and taxi, could migrate to models similar to Ridehailing with the arrival of autonomous vehicles.

Fig. 1 presents the main topics the people like when using ride-hailing services, according to a study with 582 online respondents of US population.

![Fig. 1. Topics the people like when using ride-hailing services.](image)

Authors [21] discussed the question about what really qualifies sharing services, in it is treated the concepts of shared economy and collaborative consumption.
B. Machine Learning Algorithms and Hybrid Discriminative Restricted Boltzmann Machine

Machine Learning algorithms allow the computer to automatize and improve the performance of some tasks in diverse areas [22], highlighting the RS, pattern recognition, time series prediction, search engines, and others [23], [24]. Machine learning algorithms can be classified in supervised and unsupervised classification. In general, some algorithms such as support vector machines (SVMs) and neural networks are very common and they are used in diverse applications.

The SVM and Naive Bayes algorithms are used for performing classification in ridesharing context [25]. In order to predict classes or attributes in RS the Bayesian networks, Decision trees, Random Forest, and SVM are used [26], [27]; Artificial Intelligence (AI) algorithms has also been used to predict online behavior [28] about customer services [29].

RBM is a neural network, capable of generating data according to a probability distribution. The RBM can be trained using the contrastive divergence (CD) algorithm, minimizing the generative loss function \( l_{gen} \).

\[
 l_{gen} = -\sum_{t=1}^{\alpha_t} \log P(y^{(t)}, x^{(t)})
\]  

where \( \alpha_t \) is the training samples number and \((x^t, y^t)\) are the t-iesim training sample. It has the input \( x^{(t)} \) and its respective class \( y^{(t)} \).

The DRBM algorithm has a discriminative property, which can minimize the discriminative loss function \( l_d \).

\[
 l_d = -\sum_{t=1}^{\alpha_t} \log P(y^{(t)}|x^{(t)})
\]  

The Fig. 2 presents the DRBM, in which the connections are weighted by a matrix \( W_{m \times n} \) and the connections among the label and hidden units are weighted by the matrix \( U_{k \times n} \).

![Fig. 2. Architecture of the Discriminative Restricted Boltzmann Machine.](image)

A hybrid discriminative/generative approach, \( f_h \) can be adopted by combining the respective training criteria.

\[
 f_h = f_d + \alpha f_{gen}
\]  

The HDRBM can be trained using stochastic gradient descent and adding the gradient contribution due to the \( f_d \) with \( \alpha \) times the gradient estimator for \( f_{gen} \). In which \( f_{gen} \) is cost function for generative training by the RBM [30].

In the present work, the DRBM and HDRBM implementation was developed in the Python and it is based on library Theano. The processor used is an Intel Xeon with 4 physical cores, with a frequency of 2.4 GHz and 8 GB of RAM.

Studies use AI to predict online behavior [28], but there is no studies about prediction of user preferences on ridesharing services using HDRBM algorithm. In this paper the HDRBM is tested for preferences classification according to basic profiles obtained from a social network. The machine learning algorithms used in this paper were HDRBM, Random Forest and SVM. Preliminary tests were carried out using diverse machine learning algorithms. The previously mentioned algorithms presented better classification results of user preferences in the ridesharing service.

C. Recommendation System

RS are responsible to suggest items for diverse areas to many users, it is capable of processing a large amount of data. RS enable the generation of personalized suggestions about products or services of company customers [31]. These suggestions are part of the e-commerce, such as occurs with Internet TV services [32].

The use of RS aims to increase of satisfaction and fidelity of their users, it is used in diverse applications, such the suggestion of books, movies and other products and services, based on past searches and purchases.

According to [33] is possible to provide customized without overload a system with a large data services. In general, in RS, there are three main different approaches according to the similarity between items, they are: Content-Based Filtering, Collaborative Filtering, and Hybrid Filtration.

In ridesharing applications, [12] presents a RS, named of ComeWithMe. The RS is used in Carpooling services, which records the activities of users. However, in [12] it is not studied how the user basic characteristics can influence the user preference in a ridesharing.

D. Online Social Network Data Extraction

In OSN the users can exchange their experiences and opinions. This fact contributes to knowledge and generation of relevant information, according to [34]. The OSN allow creating user profiles that will represent them. The people can interact with others through comments, posts and shares.

Authors [35] show how is possible extract some user features, obtaining some relevant information about the contents accessed by the user. This emphasizes the importance of OSN for sharing and generating information about users in online world.

Studies create systems [36], which accesses the OSN data for a extract some user data and other possible passengers. The application indicates a common link between users. However, in [36] the objective is to find a relationship among users, not covering simple user characteristics, such as age and gender and the user behavior. It differs from our study, which has the objective to avoid ridesharing among users with different characteristics or behavior.
III. RECOMMENDATION SYSTEM ARCHITECTURE

The proposed RS is implemented according to the Fig. 3. Initially, based on a questionnaire, represented by (1) in Fig. 3, which is applied to a group of volunteers, user profiles are extracted. After the training and testing of the machine learning algorithms, the one that presented the best results is applied in the RS, represented by (3) in Fig. 3. The RS proposed in this study is processed through data extracted from OSN and is developed in Hypertext Markup Language (HTML) and Cascading Style Sheets (CSS) language, with Hypertext Text Preprocessor (PHP) and the social networking application. Finally, the user data obtained in OSN are validated by the model through tests, represented by (4) in Fig. 3.

IV. METHODOLOGY OF THE RECOMMENDATION SYSTEM FOR SHARED-USE MOBILITY

The methodology for obtaining the RS for Shared-use Mobility is the following:

- Subjective tests are performed by questionnaire;
- An user preference model for ridesharing is performed by machine learning;
- The extraction of user characteristics are performed in OSN;
- The RS suggests the similar users for the ridesharing.

Initially, face-to-face subjective tests are performed to discover the user preferences with focus in ridesharing. With the results of the subjective tests a model is built by machine learning technique, the user characteristics are extracted from OSN and the recommendation for the ride according to the model of machine learning is applied.

A. Face-to-face Subjective Tests

In face-to-face tests, the volunteers answered questions about the main user’s characteristics that would influence in a good ridesharing experience. The questions in the majority answered by the users were selected to be part of the questionnaire. Thus, a user preference model for shared mobility was obtained, with the users information. The focus of the study is to discover if people would like to share a ride with people of similar or opposite hobbies and football team.

296 volunteers participated of the tests (150 women and 146 men aged between 16 and 69 years old). The questions were about some characteristics, such as:

- age;
- gender (M for man and F for woman);
- relationship (married, single, divorced, in a relationship, or widower);
- if the person would share a ride with someone with opposite hobbies and football team (no or yes), in which yes represents an acceptance of the sharing condition with a person of opposite hobby or football team.

The face-to-face subjective tests are used to create training and test set with labels (share a ride with people who have similar hobbies and share a ride with people who cheer for rival teams).

B. Machine Learning Classification

The data extracted from the questionnaires were analyzed by machine learning through the algorithms of HDRBM, DRBM, SVM, and Random Forest. For the machine learning applications was used the data analysis package Waikato Environment for Knowledge Analysis (WEKA). In WEKA the attributes and classes are defined to be used in the extraction of OSN data.

Fig. 4 presents the attributes used for the machine learning classification in WEKA, considering the same or opposite football team. The age of the users was grouped in:

- range 1 is from 16 to 19 years
- range 2 is from 20 to 29 years
- range 3 is from 30 to 49 years
- range 4 is from 50 to 69 years

The other attributes are: the exact age, the gender (M for man and F for woman), the relationship (married, single, divorced, serious relationship or widower). The classes sharingwithsimilarhobby and sharingwithdifferentfootballteam match the questions:

- Would you share a ride with people who have similar hobbies?
- Would you share a ride with people who cheer for rival teams (e.g. Palmeiras versus Corinthians or Atletico versus Cruzeiro)?

C. Online Social Network Data Extraction

An application for extracting the user characteristics on OSN was developed and it is used on Facebook, an OSN very used in other works [37]. Firstly, the application need the users authorization for data extraction approved by Facebook after it needs the user authorization for the data extraction.
The RS uses the application created on the Facebook platform to obtain the user profile data. When the user authorizes the application to access and extract his/her data from OSN, a Software Development Kit (SDK) library uses the OSN Application Programming Interface (API). Fig. 5 shows the several characteristics of the user profile that can be configured by the social network Application.

D. Recommendation System

The RS presents a friendly interface to the user, showing the users with similar preferences and characteristics. Though the RS is possible to find the user peers on the OSN. The PHP programming language and JSON was used in the RS. The user need to provide an OSN username to login on it, because it captures some basic users characteristics extracted from the OSN, such as name, gender, kind of relationship, and age.

Fig. 6 shows the result of the proposed RS for Shared-use Mobility, presenting the user peers.

V. RESULTS AND DISCUSSIONS

In this section is presented the results about the subjective tests, the machine learning algorithms and the proposed RS.

A. Subjective Tests

According to the test results, 179 users prefer to share a ride with people of same football team and 117 users prefer to share a ride with people of opposite team.

B. Machine Learning Model and Classification

The machine learning algorithms were used in the phase to classify user preference by sharing a ride with a person of same or opposite football team. Many algorithms were used in the classification phase and those which obtained better precision were HDRBM, DRBM, Random Forest and SVM. The convolutional neural networks were used in preliminary tests but the results did not present a good precision.

In the experiments, the DRBM and HDRBM topology presented a learning rate of 0.01. The training presented best results with 100 number of training epochs for DRBM and 150 epochs for DHRBM. The DRBM and HDRBM was tested with values of 10, 25, 50, 100 and 200 of hidden units, and the best performance was achieved with a value of 50 for DRBM and 100 for HDRBM. It is important to note the supervised machine learning method is used in this work, and the attributes are labeled after passing through the DRBM and HDRBM algorithms.

Table I shows the correctly classified instances each machine learning algorithm used in this work for the hobby classification class, according to the results the DRBM and HDRBM present better results, the discriminative factor helped to classification of the the user preferences. Table II presents the correctly classified instances for the sport class and the same occurred with the DRBM and HDRBM algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>CC</th>
<th>DRBM</th>
<th>HDRBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>91.5</td>
<td>94.1</td>
<td>96.9%</td>
</tr>
<tr>
<td>SVM</td>
<td>89.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE I

% OF CORRECTLY CLASSIFIED INSTANCES (CC) OF THE ALGORITHMS USED IN THE CLASSIFICATION PHASE FOR HOBBY PREFERENCE [2]
In this work, 10-fold cross-validation was used to evaluate the classification precision. The F-Measure was also used as harmonic mean between precision and recall. The precision, recall and F-measure were calculated comparing the classified classes by the machine learning algorithms and the classified classes obtained by the subjective tests. The precision indicates how many of the items that have been identified as positives are really positives and recall indicates how many of the real positives have been found.

Table III presents the performance results for the hobby classification, the F-measure near the value of 1 represents an ideal value, presenting a good relation between precision and recall. The results of Table III show superior results of the DRBM and HDRBM algorithms for the answer of yes (question: Would you share a ride with people who have similar hobbies?). Table IV presents the results of the sport event classification and the DRBM and HDRBM present a good result for answer yes and no (question: Would you share a ride with people who cheer for rival teams?).

The subjective tests were important to extract what information (user profile characteristics) influence the ridesharing preferences.

The proposed RS compares the user profile to other profiles and the results indicate that people prefer a ridesharing to others with similar preferences. To evaluate the results of the RS, 50 assessors analyzed the proposed RS and a generic RS. The generic RS presents rides according to the gender and age of the users, not considering the preferences by opposite or same hobbies and football team.

In the test, the recommendations about ridesharing were scored. An 5-point satisfaction scale was used, in this work, which is showed in Table V.

The results of Pearson Correlation Coefficient (PCC) are presented in Table IV to compare both RS, the proposed and the generic. According to the results the evaluators gave a higher score, with scores of 5 and 4 in the majority, to the proposed RS compared to a generic RS that was evaluated with scores of 3, 2 and 1.

VI. CONCLUSIONS

The results of this work highlight that it is important to consider the users profile and preferences in an RS for ridesharing. The subjective tests showed what parameters influenced in a better suggestion for ridesharing. The proposed RS reached an PCC of 0.97 and the generic RS reached a value of 0.71.

The HDRBM algorithm allows to improve the classification performance in the case of football teams preference. It is important to note that with basic user profile characteristics is possible to classify automatically users preferences. The factors that influence user QoE are present in social or cultural context, and they are important to be considered in a RS about shared mobility services.

It is worth mentioning that services such as urban mobility provide benefits not only to the user, but also to traffic and to enhance environment conditions. Thus, it can be concluded that shared services can have an important role in how society uses services today, and how shared services are becoming a trend, especially among young people.

REFERENCES


Eduardo L. Lasmar JR. is currently pursuing his M.S. degree in systems and automation engineering from Federal University of Lavras, Brazil. Specialized in automotive engineering from University of São Paulo in 2016, and B.S degree in industrial engineering. He has a solid knowledge in automotive area. His main research interest are intelligent systems for mobility solutions and recommendation systems.

Renata L. Rosa received the M.S. degree from the University of São Paulo in 2009 and the Ph.D. degree from the Polytechnic School of the University of São Paulo, in 2015 (EPUSP). She is currently an Adjunct Professor with Department of Computer Science, Federal University of Lavras, Brazil. Her current research interests include computer networks, quality of experience of multimedia service, social networks and recommendation systems.

Demóstenes Z. Rodríguez (M’12-SM’15) received the B.S. degree in electronic engineering from the Pontifical Catholic University of Peru, the M.S. degree and Ph.D. degree from the University of São Paulo in 2009 and 2013. He is currently an Adjunct Professor with the Department of Computer Science, Federal University of Lavras, Brazil. He has a solid knowledge in Telecommunication Systems and Computer Science based on 15 years of Professional experience in major companies. His research interest includes QoS and QoE in Multimedia services and architect solutions in Telecommunication Systems.