Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 161 (2016) 1282 – 1288

Procedia Engineering

www.elsevier.com/locate/procedia

World Multidisciplinary Civil Engineering-Architecture-Urban Planning Symposium 2016, WMCAUS 2016

Discovering Relationships between Factors of Round-trip Car Sharing by Using Association Rules Approach

Dahye Lee[a], Luca Quadrifoglio[a,*], Benedetta Sanjust di Teulada[b], Italo Meloni[c]

[a]*Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843, USA*
[b]*CRiMM - Center of Research on Mobility Models, University of Cagliari, Via San Giorgio 12, 09124 Cagliari, Italy*
[c]*University of Cagliari | DICAAR | Dept. of Civil and Environmental Eng. and Arch., Via Marengo 3, 09123 Cagliari, Italy*

Abstract

To maximize private vehicle usage efficiency and alleviate urban congestion, many studies and actual field operations on vehicle sharing have been done since the mid '90s. The classic carsharing system, which is known as a round-trip, is operated out of fixed stations so that customers can pick up and drop off the vehicle at the same station. Although many private carsharing companies offer one-way or free-floating trips currently for customers' convenience, studying characteristics of round-trip carsharing is still significant in making cost-beneficial, fuel-efficient, and less-congested driving environments in urban areas. The main objective of this research is a comprehensive analysis for discovering relationships between critical factors of round-trip carsharing operations based on the city of Cagliari, Italy, by analysing data retrieved from the local carsharing providing company, PlayCar, with the association rules approach. This paper investigates round-trip carsharing behaviour characteristics from various angles, including demand analysis of reservation by hourly and daily manner, geographic analysis, and connectivity to public transportations. The association rules technique was used to discover the relationships between the characteristics and understand their attributes.

Peer-review under responsibility of the organizing committee of WMCAUS 2016
Keywords: carsharing; association rules; round-trip; playcar; Cagliari.

The demand analysis shows that customer's carsharing usage does not follow typical hourly and daily flow patterns in urban areas and even one-way carsharing systems at some point. Results of association rules analysis show that the strongest dependent

* Corresponding author. Tel.: +1-979-458-4171.
E-mail address: quadrifo@tamu.edu

1877-7058
Peer-review under responsibility of the organizing committee of WMCAUS 2016
doi:10.1016/j.proeng.2016.08.573

variables do not have high correlations with the variables of distance from customers' residence locations, and the rule that has peak hour driven travels and yearly subscription customers has considerably high support value. Although the results gave an idea of connections of round-trip operations characteristics with various combinations, the degree of impact of each variable still need to be investigated. The ultimate goal for future studies is to use this information to improve the performance of these services for providers and customers and maximize connectivity to public transportation to eventually help in reducing congestion and pollution.

1. Introduction

In an effort to maximize private vehicle usage efficiency and to alleviate urban congestion, many studies and actual field operations on vehicle sharing have been done since the mid '90s. Starting in Europe, private carsharing companies started to develop their own way to provide service efficiently to customers and came up with three operational methods: round-trip (two-way, station-based), one-way, and free-floating. The local carsharing company, PlayCar, also started their station-oriented business in May 2014 in the province of Cagliari, Italy.

The most significant objective of providing station-based carsharing service is to maximize customers' efficiency by optimizing locations of stations and distributing vehicles properly. Investigating critical factors of round-trip carsharing and understanding relationships between them are important to achieving the ultimate goal and grasping customers' demands and travel behavior. Factors of round-trip carsharing operations for this study are determined based on extensive literature review. This paper investigates behavior characteristics from various angles to understand factors' attributes and then discover the relationships between factors by using the association rules technique.

The paper is organized as follows. In background studies, carsharing operation methods, methodology of the association rules, and introduction to the case study in Cagliari, Italy, are included. Following that, the data set is described, and the results of analysis are presented and discussed. A summarized conclusion of the study and comments for future research are mentioned at the end of the paper.

2. Background

2.1. Carsharing

Ever since people started to think of a more effective usage of cars, there has been a paradigm shift in urban transportation. As a solution to the excessive air pollution, energy consumption, and traffic congestion that private-owned vehicles cause, the carsharing system has been proposed, which serves as a half-way point between private and public transport (Jorge and Correia, 2013). The state of Washington defines carsharing as "*a membership program intended to offer an alternative to car ownership under which persons or entities that become members are permitted to use vehicles from a fleet on an hourly basis*." (Kittleson & Associates, Inc. et al., 2013) The classic carsharing system, such as Montreal's Communauto and Cagliari's PlayCar, which is known as a round-trip, operates out of a fixed station so that customers can pick-up and drop-off the vehicle at the same station. However, due to the less flexibility and efficiency, a one-way system was proposed. In the one-way operation system, customers can pick-up the vehicle at one station and drop it off at a different station. Thereby, several studies showed that the one-way carsharing system has better market penetration than the round-trip system both in theory and case studies (Barth and Shaheen, 2002; Firnkorn and Muller, 2011; Jorge and Correia, 2013). The free-floating carsharing system provides more flexibility to customers by allowing the collection and return of vehicles to any location in a service area such as Car2Go and DriveNow in San Francisco (Wielinski, G. et al., 2015).

2.2. Association rules

The association rules technique is one of the important data-mining tools in big data analysis and is used to effectively discover interesting relationships between variables in large amounts of data, and every relationship between variables that is earned as a result of analysis is called a rule. Agrawal et al. (1993) first presented the algorithm to generate association rules. According to his research, the association rules technique is defined as follows: let I, items, be a set of binary attributes and D, database, be a set of transactions. Each transaction ID has a unique ID and contains a subset of the items in I. An association rule can be expressed as a form X (antecedent) →Y (consequent), where X and Y are disjoint item sets (Tan, P. et al., 2005). There are two rule evaluation criteria: *support* and *confidence*. Support indicates a fraction of transactions that contain both X and Y. In other words, it means how often a rule is applicable to a given data set. This is an important metric because it eliminates uninteresting rules, such as rules with low support value, which may not be considered to be profitable in the business perspective. Additionally, confidence is the measurement of how often items in Y exist in transactions that include X, which means the higher the confidence, the more likely it is for Y to exist in transactions that contain X. Thereby, unlike support, confidence explains the reliability of inference made by a rule (Geurts, K. et al., 2003; Tan, P. et al., 2005). However, sometimes rules with high-confidence ignore the support of the item set and lead to the wrong conclusion. To address this problem, a criterion called Lift is applied. Lift measures the dependence between two item sets in the association rule, which computes the ratio of the rule's confidence and the support of the item set in the rule subsequent in case of non-binary variables.

$$Lift = \frac{c(X \to Y)}{s(Y)} = \frac{s(X \Rightarrow Y)}{s(X) \times s(Y)} \quad (1)$$

If the value of Lift is equal to one, then it implies that X and Y are independent. If Lift is greater than 1, it can be concluded that there may be a complementary effect in terms of economics between two item sets, which means the two sets are dependent on one another. On the other hand, two item sets may have substitution effect if the Lift value is less than one (Tan, P. et al., 2005; Hahsler, M., 2005).

2.3. A Case study of carsharing in Cagliari, Italy

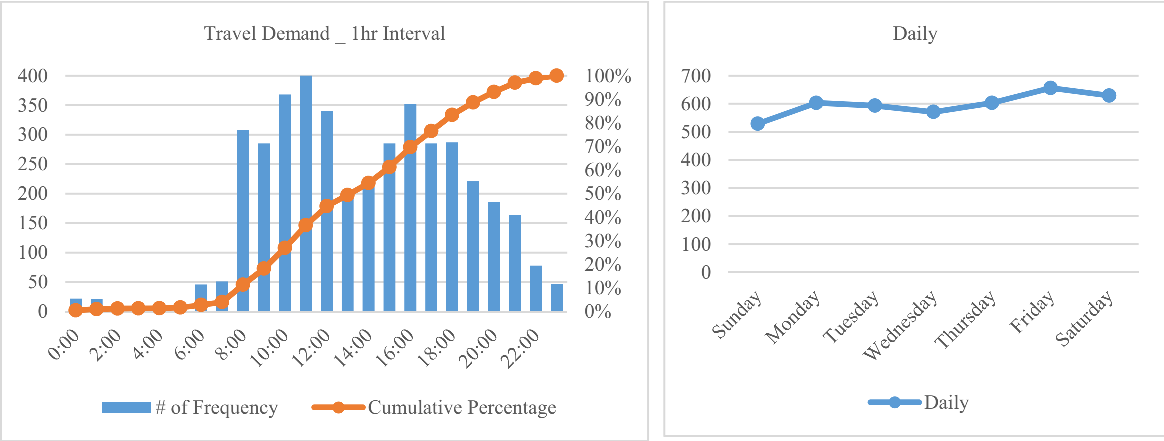
The data for analysis is retrieved from the local carsharing company PlayCar whose started station-oriented activity from May 2014 at the province of Cagliari, Italy, and now they are operating 12 stations with 7 different types of vehicles (Nissan Micra, Nissan Note, Renault Kangoo, Renault Trafic, Scooter, VW Maggiolino, and VW Polo). The service zone is in central Cagliari, the 26th largest city in Italy with a total population of 154,543 in 2014 (Wikipedia, 2014). Since the company started their business, more than 200 members have made around 4700 trips with 21 vehicles (as of January 2016).

3. Data

The received data were collected from May 2014 to January 2016. Each trip has a unique trip ID, member ID, vehicle type, reserved lot ID, reserved start- and end-time, actual trip start- and end-time, and actual billing time and distance, etc. Instead of actual traveled data, the data set of reservation was mainly considered. This is because vehicles are blocked for others even if those are not "in use" and actually ready to be picked up by customers who booked the vehicle. The reservation time is limited for 30-minute increments. Among the reservation dataset, cancelled and maintenance trips are excluded, and only normal, unconfirmed, and void status are considered. Actual traveled data set is used to collect billed trip time and distance, which correspond to each trip ID. In the end, a total of 2985 transaction data and 7 variables are used for the analysis.

The variables used for analysis are rate plan (Rate.Plan), vehicle type (Vehicle.Type), peak hour travel (Peak), distance between customers' residence and each station (Dist), billed trip time (TripT), total billed distance (TripD), and public transportation connectivity (PTC). The rate plan variable has 7 different levels: B (exclusively for company), Base (monthly subscription plan for a year), Cyclists Plan, One (no subscription), UNI (for college students), XL (yearly subscription plan), and XS (three months' subscription). Also, 6 different types of vehicles are analyzed excluding scooter.

The peak hour travel variable (Peak) is a logical data. If the trip is performed during morning or afternoon peak hour, the value of variable is true; otherwise, the value will be false. Fig. 1 - (a) shows the distribution of reserved pick-up times and the following cumulative percentage. The observed cumulative percentage implies that more than 80% of daily usage are between 8 a.m. and 7 p.m. According to the frequency graph, the morning peak hours are from 8 a.m. to 12 p.m., and the afternoon peak hours are from 3 p.m. to 6 p.m., which are slightly different from the renowned peak hours (7-9 in the morning and 4-6 in the afternoon). Thereby, the peak hour variable is decided based on this frequency analysis. Additionally, when it compared to the research from Cervero and Tsai (2004), daily timeframe of car sharing usage seems almost similar, and also Saturday and Friday are two most popular days of reservation in both the results (Fig. 1 – (b)).



a) b)

Fig. 1. Distribution of Reserved Pick-up Times and Cumulative Percentage (a) by hourly and (b) by daily.

According to the literature (Kittleson & Associates, Inc. et al., 2013), although the walking distance varies between cities and the circumstance, around 80% of passengers are willing to walk 400 m (0.25 mi) or less to bus stations at a walking speed of 5 km/h, which is a 5 minutes walking period at max. However, in Cagliari, Italy, distances between bus stations are closer than that of the U.S. Thereby, the data of public transportation connection is obtained by counting the number of bus stations and light rail stations within a radius of 200m from each lot, which is half the distance of average passenger walking distance to bus stations.

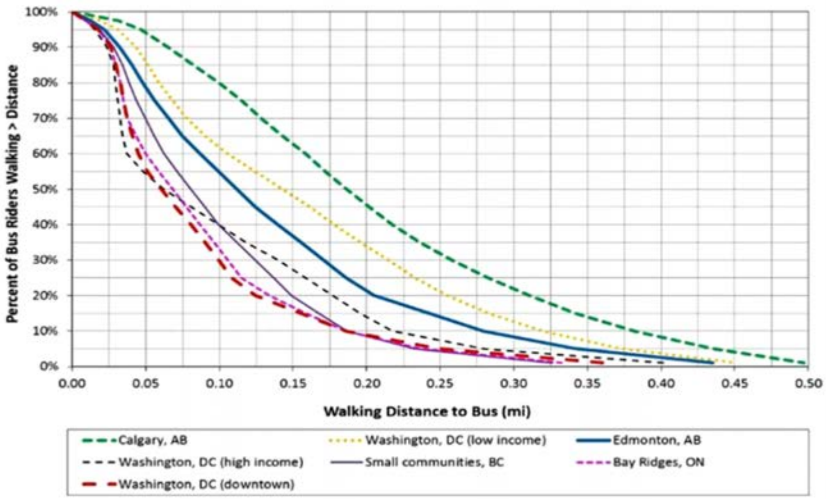


Fig. 2. Walking Distance to Bus Stops.

Fig. 3. Data Collected Area for Public Transportation Connectivity.

The numbers of bus stations and light rail stations for each carsharing lot ID are indicated in the table below. The lot number 2 and 15, which have only 2 public transportation connections, are located in the northern part of Cagliari. Lot ID 4 has the largest number of bus and light rail stations within a radius of 200m, and it is located right in front of the city hall, so a lot of traffic is expected in that area. The average of public transportation connectivity of all PlayCar's stations is 5.09, and the sample standard deviation is 2.3.

Table 1. Number of public transportation connection for each lot ID.

Lot ID	1	2	3	4	5	6	9	11	13	14	15
PTC	8	2	6	9	3	5	5	4	5	7	2

4. Results and discussions

The objective of this study is discovering the relationships between factors of round-trip carsharing operations by understanding variables' dependencies and frequencies. The association rules analysis is performed by using "apriori" function from "arules" library in R software. The data frame consists of 2985 observations and 7 variables, and all the integer or number variables are changed into factor or logical variables for the analysis. From the given dataset, 21 rules are found by the association rules analysis.

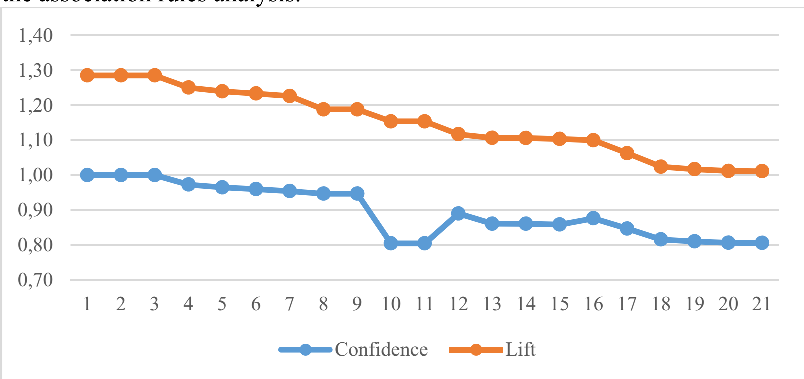


Figure 4. Rules Aligned by Descending Lift Values.

Figure 4 indicates the trend of rules' confidence values according to the descending lift values. Largely, those two values show similar patterns, however, the rules number 10 and 11 need to be noted that the consequent of rules, pick-hour travel, is not likely to exist in the trips that contain the vehicle type of Nissan Micra and the lot with 7 public transportation stations nearby, despite of the mid-high dependency between the variables.

The results of the 10 selected significantly dependent rules are shown in Table 2 by descending lift values. Including the first three rules from the Table 2, rules that are including vehicle type of Nissan Micra as a consequent have high lift values and high confidence values also.

Table 2. Ten significant rules by descending lift values.

Rules	Support	Confidence	Lift
{Peak, PTC=7} ⇒ {Vehicle.Type = Nissan Micra}	0.1062	1.0000	1.2850
{Rate.Plan=XL, PTC=7} ⇒ {Vehicle.Type = Nissan Micra}	0.1250	1.0000	1.2850
{Rate.Plan=XL, Peak, PTC=5} ⇒ {Vehicle.Type = Nissan Micra}	0.1090	0.9647	1.2396
{Dist=1.2} ⇒ {Rate.Plan=XL}	0.1709	0.8901	1.1168
{Dist=0.8} ⇒ {Vehicle.Type = Nissan Micra}	0.1534	0.8609	1.1062
{Rate.Plan=XL, Dist=0.8} ⇒ {Vehicle.Type = Nissan Micra}	0.1075	0.8606	1.1058
{Peak, Dist=0.8} ⇒ {Vehicle.Type=Nissan Micra}	0.1139	0.8586	1.1033
{Peak, Dist=1.2} ⇒ {Rate.Plan=XL}	0.1162	0.8763	1.0995
{Peak} ⇒ {Rate.Plan=XL}	0.5652	0.8103	1.0167
{TripT=2} ⇒ {Rate.Plan=XL}	0.1015	0.8059	1.0111

According to their confidence values, when customers book Nissan Micra, they tend to subscript PlayCar's plan yearly and drive during peak hours from the stations that have 5 to 7 public transportation connectivity. Furthermore, the overall results show that the strongest dependent variables among round-trip carsharing factors, which can be distinguished by high lift values such as pick-hour travel, vehicle type of Nissan Micra, and yearly rate plan (XL), do not have high correlations with the distance variable that shows the accessibility of carsharing stations from customers' residence locations. The rule with peak hour traveled vehicle and yearly subscription customers has remarkably high support value of 0.5652, which means that the probability of trip that contains both variables is more than 50% of the whole data set. Thereby, the relationship between them is highly prevailing.

5. Conclusions

Carsharing services have flourished in the last few decades and have proved to be successful and growing in many instances. In this paper, the relationships between round-trip carsharing factors from the large amount of data collected from Playcar (provider in the city of Cagliari, Italy) are analyzed by using the association rules technique in this paper. The results gave some important insight for understanding the connections of round-trip operations characteristics with various combinations. This is the first step towards investigating the degree of impact of each variable, which will be the scope of future research to eventually improve the performance of these services for providers and customers.

Acknowledgements

The data is provided from carsharing company, PlayCar, in the province of Cagliari, Italy.

References

[1] Agrawal, R., Imielinski, T., and Swami, A., 1993. Mining Association Rules between Sets of Items in Large Databases. ACM SIGMOD Conference. Washington DC, USA.

[2] Barth, M. and Shaheen, S., 2002. Shared-Use Vehicle Systems: Framework for Classifying Carsharing, Station Cars, and Combined Approaches. Transportation Research Record: Journal of the Transportation Research Board, No. 1791: 105-112.

[3] Cervero, R. and Tsai, Y., 2004. City CarShare in San Francisco, California: Second-Year Travel Demand and Car Ownership Impacts. Transportation Research Record: Journal of the Transportation Research Board, No. 1887: 117-127.

[4] Firnkorn, J. and Muller, M. 2011. What will be the environmental effects of new free-floating carsharing systems? The case of car2go in Ulm. Ecological Economics, 70 (8): 1519-1528.

[5] Geurts, K., Wets, G., Brijs, T., and Vanhoof, K., 2003. Profiling of High-Frequency Accident Locations by Use of Association Rules. Transportation Research Record: Journal of the Transportation Research Board, No. 1840: 123-130.
[6] Hahsler, M., Grün, B., Hornik, K., 2005. Introduction to a rules -- A computational environment for mining association rules and frequent item sets. Journal of Statistical Software, 14 (15):1-25.
[7] Jorge, D. and Correia, G., 2013. Carsharing systems demand estimation and defined operations: a literature review. European Journal of Transport and Infrastructure Research, 13 (3): 201-220.
[8] Kittleson & Associates, Inc., Parsons Brinckerhoff, KFH Group, Inc., Texas A&M Transportation Institute, and Arup, 2013. Transit Capacity and Quality of Service Manual: Third Edition. TCRP Report Number 165. Transportation Research Board, Washington, D.C., United States. (www.TRB.org)
[9] Tan, P., Steinbach, M., and Kumar, V., 2005. Introduction to Data Mining. Addison-Wesley Co., United States, 329-374 p.
[10] Wielinski, G., Trepanier, M., and Morency, C., 2015. What About Free-Floating Carsharing? : A Look at the Montreal, Canada, Case. Transportation Research Record: Journal of the Transportation Research Board, No. 2536: 28-36.
[11] Wikipedia, 2014. Cagliari. https://en.wikipedia.org/wiki/Cagliari/