POLITECNICO DI TORINO
Repository ISTITUZIONALE

Preliminary Investigation of Women Car Sharing Perceptions Through a Machine Learning Approach

(Article begins on next page)

10 April 2024

Preliminary Investigation of Women Car Sharing Perceptions through a Machine Learning Approach

Andrea Chicco[0000-0003-0266-4735], Miriam Pirra*[0000-0002-4110-1102] and Angela Carboni[0000-0003-1780-1157]

DIATI, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
miriam.pirra@polito.it
andrea.chicco@polito.it
angela.carboni@polito.it

Abstract. Mobility studies have shown that travel patterns and means use vary a lot comparing women and men behavior. In recent years, new solutions have been introduced in the urban mobility offer and the interest raised in investigating how they can help in reducing the gender mobility gap. The current work analyzes 2934 responses collected through a car sharing survey proposed in Italy with the precise objective of considering women and men like different kinds of users to delineate characteristics that could influence car sharing modal choice. A hierarchical clustering technique is applied to the dataset collecting a selection of questions, mainly focusing on socioeconomics features, travel patterns and individual habits. The algorithm identifies 8 clusters in the male dataset and 9 clusters in the female one, defined according to characteristics aggregating the survey respondents. Thus, a selection of these groups of respondents is analyzed in more detail according to their percentage of car sharing users, also comparing the results among male and female datasets. Many common attributes are found in clusters irrespective of the gender, showing how the interest (and its lack) toward this service affects women and men similarly. At the same time, this analysis helps in identifying the features characterizing the users to investigate how this new mobility offer can help in reducing the gender mobility gap.

Keywords: Women in Transport, Car Sharing, Smart Mobility, Travel Behavior, Users Profiling, Machine Learning, Clustering.

1 Introduction

The gender gap is a challenging and increasingly discussed topic in many fields, which is observing a recent increase in interest in the smart mobility and transport declination. As depicted in recent studies, gender seems to influence attitudes to new vehicle technologies and preferences concerning their adoption. In general, women seem less prone to emerging technologies: for example, reduced driving experience of electric vehicles is found among them usually joined with a lower interest in their purchase [1]. The previously cited technologies insert in the broader context of smart mobility, a new

concept in transport domain involving four main aspects: vehicle technology, Intelligent Transport System, data and new mobility services [2]. The latter element of the list commonly include ride-sourcing services, real-time ridesharing services, multimodal trip-planning apps, smart traffic control, self-driving vehicles and shared mobility services [3].

On the whole, the interest of a specific target of users, such as women, towards shared mobility services is just starting to be investigated. Research suggests that men make up the majority of car sharing (CS) users [4]. Reasons could be various: a lower affinity of women for technology, reduced use of a single mode of transport, avoidance of the stress associated with driving an unfamiliar car, fear for their security.

The current work points at inserting in this context combining the knowledge gained from two H2020 projects, namely TInnGO (https://www.tinngo.eu/) and STARS (http://stars-h2020.eu/). The former project aims at creating a framework for a sustainable game change in transport through to the development of methods and tools for gender and diversity mainstreaming in transport planning [2]. STARS project studies the diffusion of car sharing in Europe and its connections to technological and social innovations. The project plans the collection of new data through the development of a survey investigating car sharing user and non-user profiles [5]. The analysis of the present work represents the starting point for the activities of the so-called TInnGO Italian hub. The hub focus is the study of how shared mobility services can contribute to reducing the gender mobility gap and the suitability of sharing mobility services to satisfy special mobility needs. We started from the responses collected through the STARS survey, and we conducted a detailed analysis with a specific objective of investigation the female perception of a facet of smart mobility, namely car sharing.

2 Dataset description and statistics

The data used in this study comes from a survey developed and distributed between April and June 2018, as a part of the STARS project activities. The survey, conducted in the form of web questionnaire, was addressed to both users and non-users of car sharing services living in European cities where that kind of shared mobility service was active [5]. The questionnaire aimed to collect information about individual characteristics of the respondents (such as gender, age, and education level), household dimension and composition, travel behavior (such as the use of the private car) and relationship with the car sharing, but also sensibility about climate change and the use of technologies (such as the smartphones) [5].

About 6800 respondents took part in the survey, mainly from Italy, Sweden and Germany. As part of the Italian hub activities, only the respondents of this country will be considered. They account for a representative sample of the Italian populations living in 17 cities where at least one car sharing service was available at the time of the survey [5]. An additional number of interviewees have targeted thanks to the support of two car sharing operators which distribute the survey among their customers. The questionnaire was also circulated among the staff and students of Politecnico di Torino. As a result, 2934 completed questionnaires were retained. A first overview of the dataset showed that 46% of respondents were women, both users and non-users of car sharing

services. About 29% of males belong to the former group, while this percentage dropped to 21% among females respondents. At the same time, 61% of women declared not having had any experience with the service or not knowing this concept at all, while the corresponding percentage was 53.7% in the men sample. Respondents were also allowed to declare a past use of car sharing: this choice seemed not varying between genders (17.3% of men and 18% of women). Any specific description of the datasets is provided at this stage; however, it is worth observing that the variables investigated are similarly distributed in the two datasets.

3   Methodology

The preliminary overview of the data can provide a general picture of the two samples, namely women and men. However, this study aims to investigate the possibility of discovering some features that characterize different groups of respondents together with their use and attitude toward car sharing. In order to do so, a hierarchical co-clustering approach is adopted. Differently from traditional hierarchical clustering techniques, the proposed algorithm builds two coupled hierarchies, one on the objects and one on features, thus providing insights on both them [6]. The proposed methodology does not require a pre-specified number of clusters and produces compact hierarchies because it splits each cluster into n subclusters ($n \geq 2$), where n is automatically determined [7].

Before running the algorithm, a selection of the most significative variables from the original 41 questions is extracted from the original dataset. Besides, factor analysis is applied to a selected number of items (five questions with multiple statements that had to be assessed through 7-points Likert-scale) to define the underlying structure among the variables in the study [8]. These questions reduce to only three features, and a summated scale approach is thus applied to produce the values for the following steps of the methodology. Table 1 below presents the variables used to feed the co-clustering algorithm. Further preliminary transformations of the dataset are necessary to guarantee a proper application of the co-clustering technique. Therefore, all categorical variables of Table 1 are converted as dummy variables to gain consistency using the one-hot encoding, while numerical ones are normalized (0-1 scaling). The algorithm is applied to two different subsamples of the original dataset, considering men and women separately. The analysis is thus based on two datasets made up of 33 variables and 1584 or 1350 items, respectively.

The hierarchical co-clustering provides a partition of the datasets that can be represented through a dendrogram with different levels. 17 levels are found in the male sample, while 15 levels result in the female one. Obviously, the size of each cluster (i.e. the number of items) reduces while their number increased level by level in the dendrogram exploration. Hence, it is essential to focus on some criteria that allow checking how much in-depth it is worth analyzing the dendrogram obtained. We check the percentages of car sharing users and non-users, both in a cluster and in all those originated from that one. A small variation in the rates is considered a good motivation to stop the dendrogram investigation at a proper level. At the same time, the dimensions of the clusters obtained are evaluated to avoid identifying interesting groups that, however, collected a too little number of respondents with a loss of representativeness. After

considering the two mentioned criteria, we decide to stop our analysis at the third level in the dendrogram obtained for both samples. This procedure produces a total of 8 clusters in the male case (CM1, CM2, …, CM8 in the following) and 9 clusters in the female one (CF1, CF2, …, CF9).

Table 1. Selected input variables for co-clustering algorithm

Variable	Description	Variable	Description
CitySize	City dimension (C, I)	Age	Age (N, I)
HHsize	Household dimension (C, H)	Gender	Gender (C, I)
Child_0-3	Presence of children 0-3 years (C, H)	Education	Level of education (C, I)
Child_4-6	Presence of children 4-6 years (C, H)	HHdrivLic	Number of driving licences (C, H)
Child_7-15	Presence of children 0-3 years (C, H)	HHcar	Number of cars (C, H)
Child_>16	Presence of children 16 years and more (C, H)	CSmembership	Car sharing membership (C, I)
FreqCarDriver	Car as driver monthly use frequency (M, I)	EnvConcTravelB	Factor: environmental concerns related to travel patterns (C, I)
FreqCarPassenger	Car as passenger monthly use frequency (M, I)	EasyUtilityCS	Factor: ease of use and utility of car sharing (C, I)
FreqPublicTransport	Public transit monthly use frequency (M, I)	PrivateCarAffinity	Factor: private car affinity (C, I)
FreqTaxi	Taxi monthly use frequency (M	M1	Motive for using CS: CS parking accessibility (C, I)
FreqActive	Active modes monthly use frequency (M, I)	M2	Motive for using CS: to reduce expenses (C, I)
SmartWV	Use of smartphone in connection with travels:value (C, I)	M3	Motive for using CS: to travel more sustainably (C, I)
SmartPr	Use of smartphone in connection with travels: practicality(C, I)	M4	Motive for using CS: more comfort (C, I)
CSnearHome	Availability of car sharing stations or operational areas near home (C, I)	M5	Motive for using CS: convenience of a car only when necessary (C, I)
CSnearJob	Availability of car sharing stations or operational areas near work/study place (C, I)	M6	Motive for using CS: to avoid responsibilities with maintenance of private car (C, I)
CStechApp	Easiness of booking a shared car with a smartphone App/website (C, I)	M7	Motive for using CS: parking accessibility (C, I)
CSexpensive	Using car sharing is expensive (C, I)		

C = categorical; N = numeric; M = metric; I = individual; H = household

4 Preliminary results and discussion

This section describes more in detail a selection of clusters among those resulting from the application of the hierarchical co-clustering algorithm. Four clusters coming from the female dataset will be analyzed: CF6 collects car sharing users, CF4 gathers past users, while CF1 and CF3 define two groups of non-users. Similarly, in the male sample, CM7 and CM8 are the labels of two clusters of users, CM3 refers to a group of past users and CM4 to a cluster of non-users.

Clusters of car sharing members
The three clusters show substantial percentages of car sharing users (about 70%) and are characterized by a not so high number of respondents. More precisely, CM7 collects 162 items, 133 are found in CM7 while only 87 women belong to CF6. People of these clusters share many common characteristics concerning their sociodemographic profiles, but also show some peculiarities when it comes to travel habits and perceptions towards car sharing. In general, both male and female car sharing members live in cities with more than 500k inhabitants (100% of respondents in CM7, about 98% in CF6 and 86% in CM8). Concerning the sociodemographic profile, individuals belonging to CM7 and to CF6 are young, 34.6 years for the men cluster and 33.2 years for the women on average. Both groups have a very high level of education, especially if compared to the whole sample: all males in CM7 have a university degree, and 17.3% has an even higher title (postgraduate master, Ph.D); in the women group almost 90% of individuals have got a degree, however there is a small percentage of individuals with a diploma (11.5%). These two clusters share the highest rate of individual household: 39.5% of CM7 and 58.6% of CF6 is single, which is far higher compared to the general trend of the whole men and women samples (only 14.6% and 13.8% lives alone, respectively). Consequently, in these clusters, there are almost no children. Differently from CM7 and CF6, the CM8 is composed of men slightly older (37.5 years old on average), that mainly live in the two-members household (probable couples without children). In this group, the percentage of children is also very low compared to the whole men dataset (18% vs 50%), but a bit higher compared to CM7. More similarly to the women cluster, in CM8 there are also people with a secondary school diploma, but the group still has a higher education level compared to the male group.

In both men clusters, the percentage of a car-free household is quite high (19.1% and 29.3% in CM7 and CM8 respectively). Similarly, most respondents of CF6 live in a car-free household too (54%): this percentage is very high compared to the general trend in the women dataset and even compared to male car sharing members.

Regarding travel habits, individuals belonging to the car sharing members' clusters (both men and women) are more multimodal: they use private cars less frequently and use public transport and active modes (bike and walks) more regularly than respondents belonging to other clusters. Contrarily to the general trend, most of the men in CM7 and CM8 stated having no affinity with the private car. This affinity is even lower in CF6: this might be the reason behind the lowest frequency of use of private car found in this group (only 3 days per month on average). Therefore, women car sharers seem to use private vehicles even less than men car sharing members.

The use of technologies is positively evaluated in all clusters, whose members consider smartphones useful and practical for journey planning. Furthermore, all these respondents show high environmental concerns related to their travel patterns; however, this is a general trend observed in almost the whole sample. The three clusters show nearly the same characteristics when considering car sharing elements investigated within the survey. Car sharing stations or operational areas are generally available near home and workplace (or study place). Therefore, service availability seems to be one of the key elements that trigger the service registration. As already mentioned above, these individuals have a good feeling with technologies and smartphones; consequently, most of them do not find it difficult to book a shared car through an app or website. Besides, they find car sharing easy to use, useful to reach their activities and not expensive. Finally, the main motives that entice both men and women to register for car sharing services are the availability of a car when they need, the absence of responsibilities with the maintenance of the private vehicle and the car sharing parking accessibility.

Clusters of previous car sharing members
Two clusters collect the highest percentage of past users: one is made up of 95 items and comes out from the male dataset (CM3); the other collects 82 respondents of the female dataset (CF4). The comparison of these two clusters reveals interesting similarities. Many respondents live in medium and large cities, most of them belonging to a household with three or more members. 77.9% of men declare having at least a child under 16, while this number increases in CF4. In both cases, at least half of the respondents state owning two cars within the household, while almost all the remaining hold only one. The education qualification seems not changing comparing the genders, with a predominance of secondary school diploma (72.6% among men, 68.3% among women) and few respondents having a degree. The analysis of travel habits reveals a quite high frequency of use of all different modes in both clusters and strong reliability on the car as the driver. As an oddity, these groups of respondents are the only ones revealing the most significant monthly use of the taxi, compared to both datasets. Considering these results, it is somehow weird finding a not valuable tendency regarding the private car affinity. Another peculiarity lies in the lowest environmental concern, which is in sharp contrast with all other groups. A focus on the variables connected to car sharing reveals that these two clusters had strongly different opinions compared with all other clusters. This aspect could help in shedding light on the reasons behind their state of non-users. For instance, most of them see smartphones as worthless and impractical during travel activities. At the same time, the majority does not have car sharing stations or operational areas available near her/his house or job place. This latter aspect seems more relevant in CF4 (higher percentages), and further investigations could help in understanding the entity of this influence. The low satisfaction related to car sharing use is depicted from the absence of detectable motives for joining this service. Overall, CF4 and CM3 do not show any variation that could be due to gender so far. Some slight differences can be found in the CS perception when comparing the percentages in these cluster with those found in the complete datasets. A high number of women in CF4 (70%) thinks that this service is rather worthless and not easy to use,

while only 58.9% of men of CM3 has this thought. These percentages represent some kind of anomalies compared to the overall datasets, where they come to be 28.9% and 25.3%, respectively. Moreover, CM3 collects the highest rate of people believing that it is difficult to book a shared car through an app or website (53.7% against 80.3% in the whole dataset). At the same time, CF4 groups a significant number of respondents that associated a high cost to this service.

Clusters of car sharing non-members
The investigation of clusters with low percentages of people joining car sharing services bring us to focus on three groups of respondents: two with no women users (CF1 and CF3) and one with only 1% of users in the male dataset (CM4). They collect 189, 123 and 208 items, respectively. All these clusters show many people living in medium and large cities. The household characterization reveals the total absence of singles and a presence of 16 years older children, more significant than the datasets trends. The average age in all clusters is similar and assesses at 48 or 47, which represent the highest values for the females' dataset. The investigation of travel habits starts with observing that CF1, CF3 and CM4 collect the majority of household owning two cars. The massive adoption of a personal vehicle seems confirmed by the highest monthly frequency associated with this mean, compared to the entire datasets. In addition, the private car affinity reaches the highest values in all clusters: they are higher than 90%, while they assess around 56-57% on average. At the same time, these clusters reveal the lowest use of public transport and active means. The latter result is somehow coherent with the highest monthly frequency that is found in the car sharing users' clusters: this travel mode can be, in fact, commonly associated with multimodal people. Some differences start rising while analyzing the trends of variables more related to technological aspects and car sharing perception. CF1 and CF3 reveal a very different opinion about the use of the smartphone as support for travelling. CF1 thinks it could give a valuable help (90.5% of respondents compared to 82% of the whole dataset), while this number reduced to 61% in CF3. This latter value is similar to the male case, where it stopped at 64.4% against an amount of 79.5% found for the entire dataset. A further element of interest characterizes CF3: almost half of its respondents think that smartphones are not so practical for travelling purposes. This aspect is somewhat in contrast with the overall dataset trend, where the great majority of women (77.4%) appreciates the tool practicality. At the same time, CF3 collects the broader number of respondents recognizing a difficulty in booking a car using the technology (36.6% against 21.3% of the entire female dataset). Similar values are found in CM4, demonstrating that people not so practical with technology would not be potential car sharing users. The investigation of all these characteristics is a crucial step because it can provide useful hints on the motivations behind the low interest in joining this mobility service. For instance, the percentages of people declaring not having a car sharing station in the proximity of their house or job place are very high in these clusters. Some of the personal perceptions about this mobility service are different in the various clusters. The men seem, for example, more concerned about the cost of the service: this group of respondents reveals the highest percentage, with 41.8% of people declaring that car sharing is expensive compared to 27.4% in the whole male dataset.

5   Conclusion

The work presented in the current paper includes the preliminary analyses conducted on a rich dataset, including car sharing users and non-users, considering women and men as different groups of respondents. Further investigations will try to investigate in more detail the factors that influence the creation of the clusters obtained so far. The present analysis was mainly focused on the identification of features found in different groups of respondents to depict common characteristics. Further studies will help in delineating more precisely the contribution of each variable in the car sharing modal choice investigated at gender level.

References

1. Sovacool, B.K., Kester, J., Noel, L., de Rubens, G.Z.: The demographics of decarbonizing transport: The influence of gender, education, occupation, age, and household size on electric mobility preferences in the Nordic region. Glob. Environ. Chang. (2018). https://doi.org/10.1016/j.gloenvcha.2018.06.008

2. Pirra, M., Carboni, A., Diana, M.: Assessing gender gaps in educational provision, research and employment opportunities in the transport sector at the European level. In: Proceedings of TRA2020, April 27-30 2020, Helsinki, Finland

3. Singh, Y.J.: Is smart mobility also gender-smart? J. Gend. Stud. 1–15 (2019). https://doi.org/10.1080/09589236.2019.1650728

4. Alonso-Almeida, M. del M.: Carsharing: Another gender issue? Drivers of carsharing usage among women and relationship to perceived value. Travel Behav. Soc. 17, 36–45 (2019). https://doi.org/10.1016/j.tbs.2019.06.003

5. Martins Ramos, Silva, É., Bergstad Jakobsson, C., Chicco, A., Diana, M.: Mobility styles and car sharing use in Europe : attitudes , behaviours , motives and sustainability. Eur. Transp. Res. Rev. 12, (2020). https://doi.org/10.1186/s12544-020-0402-4

6. Ienco, D., Pensa, R.G., Meo, R.: Parameter-free hierarchical co-clustering by n-ary splits. In: Buntine, W., Grobelnik, M., Mladenić, D., and Shawe-Taylor, J. (eds.) ECML PKDD 2009, LNCS, vol. 5781. pp. 580–595. Springer, Heidelberg (2009)

7. Pensa, R.G., Ienco, D., Meo, R.: Hierarchical co-clustering: Off-line and incremental approaches. Data Min. Knowl. Discov. (2014). https://doi.org/10.1007/s10618-012-0292-8

8. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate Data Analysis Seventh Edition. (2014)