Identifying car-sharing quality determinants: a data-driven approach to improve engineering design

Barravecchia, F.[1], Mastrogiacomo, L.[1] and Franceschini, F.[1]

[1] Politecnico di Torino, DIGEP (Department of Management and Production Engineering),

Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

STRUCTURED ABSTRACT

Purpose - This study aims at identifying the quality determinants of car-sharing services, analyzing unstructured User-Generated Contents (UGCs) and, more specifically, online reviews generated by users of the same car-sharing service. Moreover, this paper discusses the implication of the proposed data-driven approach on engineering design.

Methodology - A large dataset of car-sharing users' online reviews was analyzed by means of the Structural Topic Model (STM), i.e. a variant of Latent Dirichlet Allocation (LDA) technique which discovers underlying topics in a collection of documents also using document-level covariate information.

Findings - This paper reports an analysis of UGCs related to different car-sharing services. The analysis unveils 20 determinants of car-sharing quality: *customer service (physical office); accident & damages management; registration process; charges & fees; parking areas; app reliability; end trip issues; car condition; convenience; use rates; car proximity; car availability; efficacy; sharing benefits; customer service responsiveness; intermodal transportation; car start-up issues; customer service courtesy; billing and membership; car reservation.*

Originality – This paper proposes a novel approach to identify quality determinants by analyzing UGCs. The study of the quality determinants of a car-sharing service is a scarcely discussed field of research although the car-sharing sector is an increasingly important part of the transport economy.

Keywords: Car-sharing, Quality determinants, User-Generated Contents, Topic modelling.

Paper type: Research paper

INTRODUTION

Business models combining the offering of products and services are becoming considerably more widespread (Mastrogiacomo et al., 2019; L. Mastrogiacomo et al., 2020; Luca Mastrogiacomo et al., 2020). In this context, car sharing, a form of shared mobility, has gained increasing popularity in recent years (Shaheen and Cohen, 2007). Given its promise to reduce traffic congestion, parking demands and pollution, this mode of shared transportation has spread especially in urban contexts, so much so that several new competitors are recently entering this market designing and proposing new service solutions (Shaheen and Cohen, 2013). The number of users of carsharing services is growing rapidly: 15 million people in Europe (about 2% of the population) are expected to use carsharing services in 2020, compared to 7 million in 2015 (Frost & Sullivan, 2016). This increase of users is expected to increase profits from approximately $1 billion in 2013 to $10.8 billion by 2025 (Prescient & Strategic Intelligence, 2019). Generally, carsharing schemes can fall into one of four models:

- *one-way*, when members are allowed to begin and end their trip at different locations, through free-floating zones or station-based models with designated parking locations;
- *roundtrip*, when members are required to begin and end their trip at the same location;
- *peer-to-peer*, when the vehicles are typically privately owned or leased with the sharing system operated by a third-party;
- *fractional*, if the users to co-own a vehicle and share its costs and use.

Among others, the most successful model in terms of users over time is the "one-way" model in both free-floating and station-based configuration (Boyaci et al., 2015).

Despite the emerging importance of this type of mobility, to the best of authors’ knowledge, no structured analysis has been performed to understand the most critical determinants of the quality of a car-sharing service by means of UGCs (Illgen and Höck, 2019). Apart from traditional approaches to assess quality of car-sharing, e.g. questions/focus groups/interviews (Möhlmann, 2015), there are a couple of interesting works on the subject which relies on the use of UGCs. Guglielmetti Mugion et al. (Guglielmetti Mugion et al., 2019) discusses the antecedents of the use of car-sharing through an analysis of UGCs limited to service users in the city of Rome. Jeong et al. (Jeong et al., 2019), only investigates online car-sharing reviews on the google play website, which is designed to collect users' opinions mainly related to the application of the service.

The focus of this study is to extend previous analysis to the evaluation of the quality determinants (i.e. the most significant characteristics that influence perceived quality) of one-way and roundtrip car-sharing using information available online in the form of user reviews. A recent approach to determine the quality determinants is the analysis of UGCs and, more specifically, of online reviews which can offer a low-cost source of information for understanding customer's expectations and requirements. The identification of quality determinants is based on the in-depth analysis of such data, leveraging text mining approaches capable of obtaining information through text documents written in a natural language (Aggarwal and Zhai, 2012). To this end, topic modelling approaches are used. Such approaches are based on unsupervised machine-learning algorithms that can detect latent topics running through a collection of unstructured documents (Müller et al., 2016). Given a big set of documents, topic modelling algorithms deals with the problems of: (i) identifying a set of topics that describe a text corpus (i.e. a collection of text document from a variety of sources); (ii) associating a set of keywords to each topic and (iii) defining a specific mixture of these topics for each document (Blei et al., 2003).

The logic of these approaches is that if a topic is discussed (within the UGCs), then it is critical to the definition of the quality of the object (product, service or product-service system) under investigation.

In this study, we use a probabilistic topic modelling method, named as Structural Topic Model (STM), an extension of well-established probabilistic topic models, such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003) or Correlated Topic Models (CTP) (Blei and Lafferty, 2007). A significant advantage of STM is that it allows the connection of arbitrary information, in the form of covariates (such as customer ratings, date and place of publication of the review, service provider, etc.), with the degree of association of a document with a topic (topic prevalence) as well as the degree of association of a word with a topic (content prevalence). Roberts et al. (Roberts et al., 2014, 2019) provide a good overviews of the STM algorithm.

This paper has been organized in three sections. The first section deals with the methodology applied to identify car-sharing quality determinants, with particular reference to how the experimental dataset was obtained and processed. The second section highlights the preliminary results of the analysis. Implications for engineering design are discussed in the third section. Finally, the concluding section explores the limitations and future directions of this research.

METODOLOGHY

The analysis herein presented has been carried out using the Structured Topic Modelling (STM) package of R software (R Core Team, 2017). Its application consists of the five steps, described by the following sections (see Figure 7):

(i) dataset extraction;

(ii) pre-processing;

(iii) identification of the optimal number of topics:

(iv) labelling;

(v) validation of results (see Figure 7).

Figure 7 – Activity flow of the methodology.

Dataset extraction

Analyzed data are reviews and relevant metadata (car-sharing providers, nationality, rating, date, source) retrieved in December 2019 from different review aggregators: Yelp, Google, Trustpilot, Facebook and Playstore. Reviews were published from January 2010 to December 2019. Only English-language reviews were selected, with a total of almost 17,000 reviews from 22 carsharing providers (Car2go, DriveNow, Maven, Zipcar, Goget, etc.), distributed in 3 countries (US, Canada and UK). Each provider was related to the type of car sharing (station-based or free-floating). The average length of the obtained reviews is about 500 characters.

The information concerning review ratings, types of carsharing (station-based or free-floating) and countries was used to define the topic prevalence in the STM model, i.e. the overall frequencies of words associated to each topic.

Pre-processing

According to previous approaches (Meyer *et al.*, 2008; Guo *et al.*, 2017), the text corpus was pre-processed and unified in order to improve the efficiency of the topic modelling algorithm. In detail, the text corpus was pre-processed as follows:

- the text was converted to lowercase in order to eliminate ambiguity with uppercase words;
- punctuation and numbers were removed since adding little topical content;
- English stop words (e.g. “the”, “and”, “when”, “is”, “at”, “which”, “on”, etc.) were removed;
- words shorter than 2 or longer than 15 characters were removed;
- words with an extremely low frequency (less than 15 occurrences in the whole text corpus) were excluded from the text corpus since their inclusion would confound results or would not

be representative of any specific topics.

- the text was normalized using Porter stemmer (or ‘Porter stemming’) to reduce similar words to a unique term. Stemming removes the commoner morphological and inflexional endings from words in English (Jivani, 2011). For example, the words “likes”, “liked”, “likely” and “liking” were reduce to the stem “like”;
- words generally not related to topical content (such as: “another”, “mean”, “etc”, “problem”, “review”, “made”) were removed;
- All the n-grams, i.e. contiguous sequences of n items from a given sequence of text were replaced by a single term. For example, the n-grams “customer service” were replaced by the term “customerservice”.

Identification of the optimal number of topics

An essential parameter for the STM method is *T*, i.e. the number of topics able to describe the analyzed text corpus. The literature discusses a number of possible alternatives to define T (Wallach *et al.*, 2009). To the purpose of this analysis, the held-out likelihood has been selected as measure of performance of the topic model. The held-out likelihood evaluates how well the trained model explains the held-out data (i.e. a portion of data not used to develop the topic model). It can be seen as a measure of how the developed topic model is able to explain the overall variability in the text corpus (Scott and Baldridge, 2013; Roberts *et al.*, 2014). In the proposed application, only the 90% of available UGCs was used to train the topic model and the remaining 10% was used to test the developed topic model. Held-out likelihood (L) is formally defined as le the log probability (p) of the held-out data ($W_{held-out}$) given the trained model ($M_{trained}$):

$$L = \log p(W_{held-out} | M_{trained}) \quad (1)$$

The graph in

Figure 8 show the values of the Held-out likelihood as a function of T (from 5 to 100).

From the graph, we can observe that starting from the value of T equal to 20 there is an almost stationary Held-out Likelihood. Considering this, an optimal number of T = 20 topics was identified.

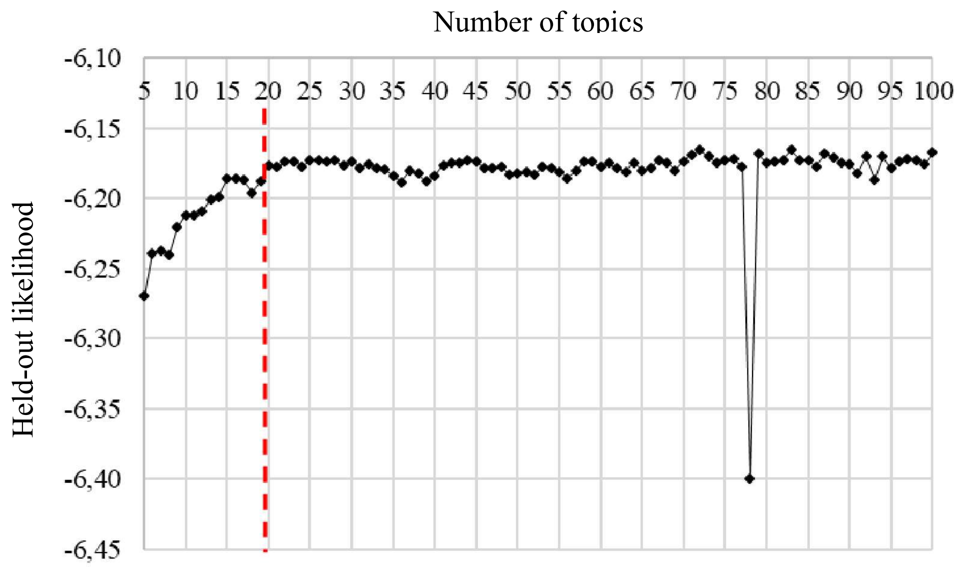

Figure 8 – Held-out likelihood by number of topics [5-100].

Labelling

For each topic, the STM approach identifies the most relevant keywords, however to generate a relevant semantic label the method still requires some human input (Blei, 2012). To date, no automatic labelling techniques have yet been developed. Table 6 shows the identified labels and the relevant lists of keywords as defined by the authors. After a first phase of independent analysis, which led to partially different labels, a joint brainstorming allowed to settle the differences and obtain the final list of labels listed in Table 6. Finally, to test their reliability, the defined topic labels were submitted for confirmation to an external panel familiar with quality research and practice.

Table 6 – Top keywords and related semantic labels of the identified topics.

Number of topic	Keywords (Highest probability)	Topic label
1	help, phone, call, person, office, answer, number	Customer service (physical office)
2	damage, report, accident, fault, member, enterprise, claim,	Accident & damages management
3	sign, process, website, license, drive, driver, registration	Registration process
4	charge, fee, late, return, time, pay, hour	Charges & fees
5	park, lot, spot, find, ticket, street, space	Parking areas
6	app, work, update, book, map, reserve, time	App reliability
7	trip, end, time, make, actual, take, system	End trip issues
8	gas, dirty, rent, clean, tank, card, tire	Car condition
9	need, convenient, quick, recommend, awesome, clean, perfect	Convenience
10	hour, price, rate, cost, expense, mile, cheaper	Use rates
11	minute, reservation, walk, wait, home, time, away	Car proximity
12	car, available, location, vehicle, area, change, time	Car availability
13	use, time, now, far, user, review, star	Efficacy
14	city, year, insurance, member, gas, need, month	Sharing benefits
15	service, custom, issue, company, terrible, problem, experience	Customer service responsiveness
16	way, drive, little, take, get, town, bus	Intermodal transportation
17	time, start, location, turn, lock, pick, key	Car start-up issues
18	call, member, cancel, ask, rep, refund, manage	Customer service courtesy
19	account, card, email, credit, month, day, membership,	Billing and membership
20	reservation, plan, time, need, book, cancel, advance	Car reservation

Data Verification

Obtained results were verified by comparing the assigned topic of a randomly selected sample composed of 100 reviews with a manual topic assignment performed by the authors. For each of the 100 reviews, the authors were requested to agree in the association of one or more of the 20 topics identified by STM. The so-defined topic assignment was then considered as reference and compared to that obtained by STM. For each review and topic, the following four cases can occur (see examples in Table 7):

- True Positive (*tp*), i.e. agreement between authors and algorithm in the assignment of a review to a topic.
- True Negative (*tn*), i.e. agreement between authors and algorithm not to assign a review to a topic.
- False Positive (*fp*), i.e. misalignment between the assignment of the review to a topic by STM and the non-assignment by the authors (type I error).
- False Negative (*fn*): i.e. misalignment between the assignment of the review to a topic by the authors and the non-assignment by STM (type II error).

Table 7 – Examples of verification procedures. Total number of topics equal to 20.

	STM topic assignment	Manual topic assignment	True Positive	True Negative	False Positive	False Negative
Review 1	20 - 11	20 - 4	1	17	1	1
Review 2	7	7	1	19	0	0
Review 3	5 – 8 - 7	5 - 8	2	17	1	0
Review 4	14 - 16	11 – 14 - 16	2	17	0	1

According to Costa et al. (Costa *et al.*, 2007), three verification indicators have been calculated (see Table 8). *Accuracy* is the most intuitive performance measure and it is equal to the ratio of correctly predicted observation to the total observations. It measures how often the algorithm produce a correct topic assignment. Accuracy assumes equal costs for both kinds of errors. Further metrics should be calculated in order to evaluate more accurately the performance of the applied method. To fully evaluate the effectiveness of a topic modeling algorithm, two indicators should also be considered: Recall and Precision. *Recall*, also known as sensitivity or true positive rate, can be defined as the ratio of the total number of correctly predicted observation (true positive) with the sum of true positive and false negative observations. Recall metric answers to the questions: “If a topic is present in a review, how often is the algorithm able to detect it??”. *Precision*, also known as positive predictive value, is equal to the ratio between the total number of correctly classified positive examples by the total number of predicted positive prediction (true positive + false positive). This metric answers to the question: “What proportion of positive topic assignments was actually correct?”.

These three metrics show a generally good correspondence between the assignment produced by STM and the authors. The accuracy of 94% proves good effectiveness of the method to predict the content of the reviews, correctly identifying true positive and true negative. According to Nassirtoussi et al. (Nassirtoussi *et al.*, 2014), accuracy values above 55% can be accepted as “report-worthy”. According to Zaki and McColl-Kennedy (Zaki and McColl-Kennedy, 2020), in most cases, accuracy is between 50% and 80%. The Recall and Precision indicators, respectively equal to 73% and 65%, show that the method performs well in terms identification of the topics (true positive).

Table 8 – Verification indicators (Costa et al., 2007).

Indicator	Definition	Value
Accuracy	$A = (tp + tn)/(tp + tn + fp + fb)$	0.94
Recall	$R = tp / (tp + fn)$	0.73
Precision	$P = tp / (tp + fp)$	0.65

PRELIMINARY RESULTS

Figure 9 shows the proportion (i.e. the average weight) of the 20 identified topics in the analyzed reviews. The most discussed topics are topic 6, concerning the reliability of the mobile application, topic 9, related to service convenience, and topic 15, related to the responsiveness of the customer service. The less discussed topics are those related to the tangible component of the carsharing service: topic 2, relating to the management of accidents and damage to vehicles, and topic 8, relating to the internal condition of vehicles. Note that this does not mean these topics are more “critical to quality” than others. The difference in proportions may depend on a number of factors, including the review aggregators used for the analysis, which may be more (or less) oriented towards collecting specific information on certain topics. For example, the Playstore commonly collects information related to the user experience with respect to the applications.

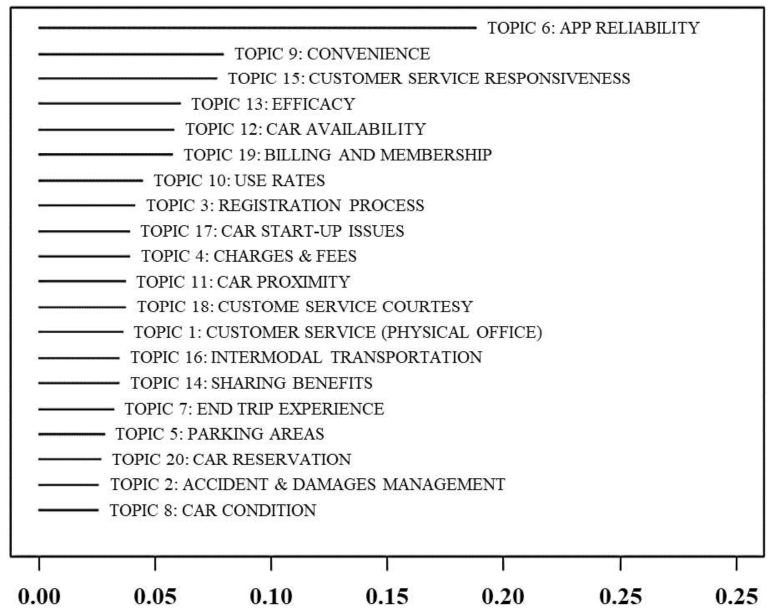

Figure 9 – Topic proportions.

Topics vs. Car sharing scheme

Figure 10 shows the association between identified topics and the scheme of car sharing (station-based or free-floating). In detail, the figure shows the marginal effects in the change of expected proportions of topic prevalence: dots on the chart depict expected difference in topic proportions, while the horizontal bars represent uncertainty ($\alpha = 95\%$) in these estimates. As an example, see how topic 14 (sharing benefits) is more discussed (by about 4%) for station-based than for free-floating car-sharing services.

The following considerations can be made by analyzing the figure:

- topic 12 (car availability) and 17 (car start-up issues) do not seem to be influenced by the type of car-sharing scheme;
- reviews related to topic 3 (registration process), 5 (parking areas), 6 (app reliability), 7 (end trip issues), 11 (car proximity), 13 (efficacy) and 16 (intermodal transportation) are prevalent for free-floating scheme. It is noteworthy how topic 6 has an important prevalence of 15%.
- topic 1 (customer service – physical office), 2 (accident and damages management), 4 (charges and fees), 8 (car condition), 9 (convenience), 10 (use rates), 14 (sharing benefits), 15 (customer service responsiveness), 18 (customer service courtesy), 19 (billing and membership) and 20 (car reservation) are mainly discussed for station-based car-sharing scheme.

Despite the differences highlighted, it is possible to state that all the topics identified are associated with both types of car-sharing.

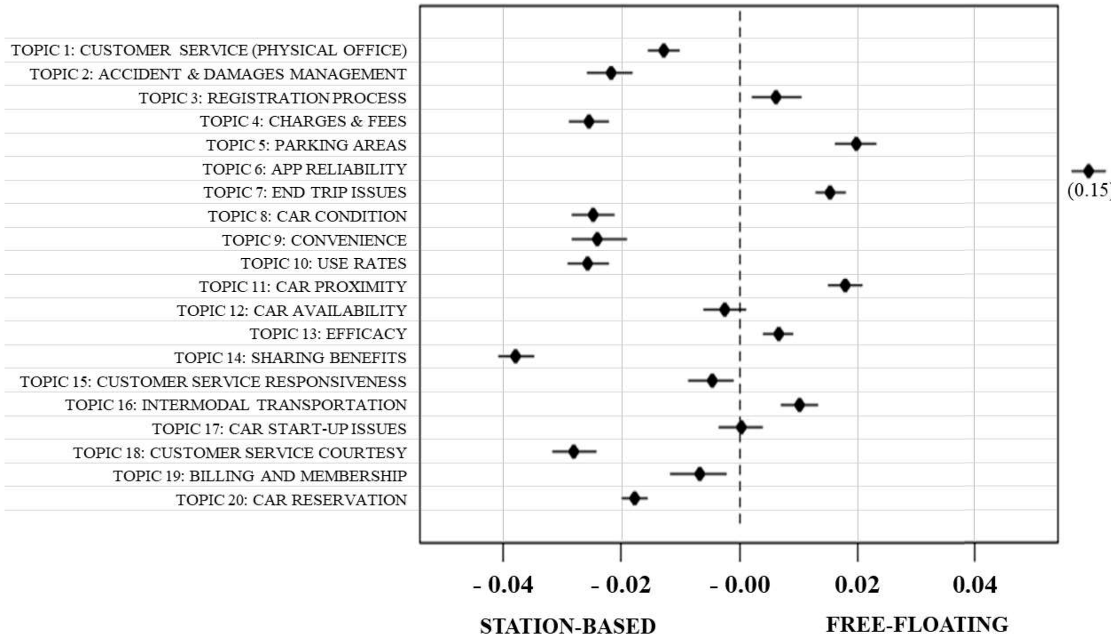

Figure 10 – Marginal effects in the change of expected proportions of topic prevalence based on the scheme of car-sharing. The dotted line represents the zero effect.

Topics vs. Ratings

Figure 11 shows the marginal effects in the change of the expected proportion of topic prevalence based on low (1 and 2) and high (4 and 5) review ratings. As the distance from the dashed axis increases, the probability of specific topics becomes more dominant. Three different cluster of topics can be qualitatively identified:

- topic 1 (customer service – physical office), 8 (car condition), 17 (car start-up issues) and 20 (car reservation) are "neutral" with respect to the review rating;
- topic 5 (parking areas), 6 (app reliability), 7 (end trip issues), 9 (convenience), 10 (use rates), 11(car proximity), 12 (car availability), 13 (efficacy), 14 (sharing benefits) and 16 (intermodal transportation), seem to be driver of high ratings as the topic are becoming dominant for satisfied customers;
- topic 2 (accident & damages management), 3 (registration process), 4 (charges & fees), 15 (customer service responsiveness), 18 (customer service courtesy) and 19 (billing and membership) are critical factors that lead to customer dissatisfaction when occurring and if not appropriately addressed.

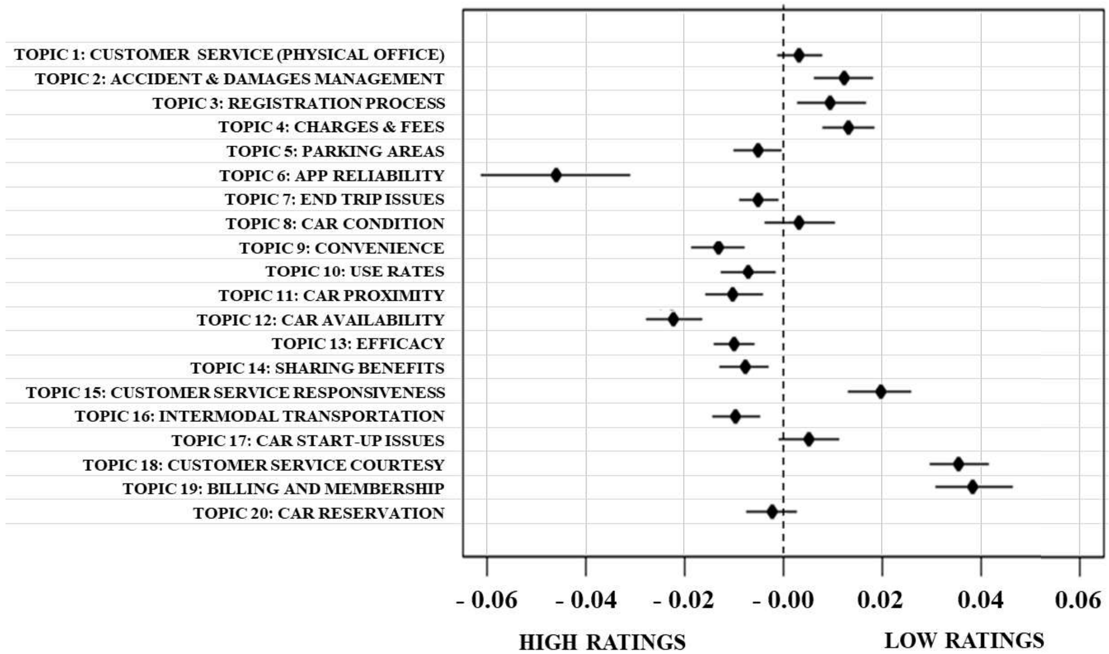

Figure 11 – Marginal effects in the change of expected proportions of topic prevalence based on low and high review ratings. The dotted line represents the zero effect.

IMPLICATIONS FOR ENGINEERING DESIGN

The findings of this article, while preliminary, highlights the potentialities that data-driven methodologies may have in quality management and engineering design. Understanding the quality determinants of a product, service or product-service system provides some support for their engineering design. In particular, several steps of the design process may be influenced by the results gathered by the application of proposed data-driven methodology, including:

(i) *Identification of design challenges*: the study of latent quality determinants can help designers in identifying new design challenges for the development of innovative products, services or product-service systems.

(ii) *Comparison of existing approaches*: the examination of the prevalence of quality determinants (see for example section "Topics vs. Car sharing scheme") can help designers to group different types of existing products, services or product-service systems into homogeneous families. In this way, the comparison of different existing approaches can be performed by directly considering customers' perceptions. In addition, the analysis of the evolution of the prevalence of quality determinants may also allow to identify potential emerging markets.

(iii) *Identification of customer needs*: as can be seen from the discussion of the proposed results, a comprehensive and detailed overview of the quality determinants of a product, service or product-service system can be obtained by analysing UGCs. In this consideration, UGCs may serve as primary source for the identification of customer needs. Moreover, using a large amount of UGCs produced over several years, it is possible to analyse the temporal dynamics of the quality determinants, eventually predicting patterns and anticipating customer needs.

CONCLUSIONS

This paper reports an analysis of User-Generated Contents related to different car-sharing services. The analysis unveils 20 determinants of car-sharing quality: *customer service (physical office); accident & damages management; registration process; charges & fees; parking areas; app reliability; end trip issues; car condition; convenience; use rates; car proximity; car availability; efficacy; sharing benefits; customer service responsiveness; intermodal transportation; car start-up issues; customer service courtesy; billing and membership; car reservation.*

This analysis presents a number of novel aspects, including: (i) it is one of the first attempts to identify quality determinants by analyzing UGCs, (ii) the study of the quality determinants of a car-sharing service is a scarcely discussed field of research although the car-sharing sector is an increasingly important part of the transport economy.

Results of the proposed approach may have significant implications in engineering design, with particular reference to: (i) the identification of design challenges; (ii) the comparison of existing approaches and (iii) the identification of customer needs.

Despite producing multi-faceted insights, the adopted method uses UGCs, i.e. textual information produced spontaneously by users that can potentially be distorted as generated by an uncontrolled sample of individuals. Moreover, the analysis introduces elements of subjectivity (e.g. in the labelling operation). Future developments will be directed to the solution of the above-mentioned limitations as well as to the use of similar approaches for the study of the quality in different contexts.

ACKNOWLEDGEMENTS

This work has been partially supported by "Ministero dell'Istruzione, dell'Università e della Ricerca"

Award "TESUN-83486178370409 finanziamento dipartimenti di eccellenza CAP. 1694 TIT. 232 ART. 6".

REFERENCES

Aggarwal, C. C. and Zhai, C. (2012) Mining text data. Springer-Verlag New York.

Blei, D. M. (2012) ‘Probabilistic topic models’, Communications of the ACM, Vol. 55 No. 4, pp. 77–

84.

Blei, D. M. and Lafferty, J. D. (2007) 'A correlated topic model of science', The Annals of Applied Statistics, Vol. 1 No. 1, pp. 17–35.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003) 'Latent dirichlet allocation', Journal of machine Learning research, Vol. 3 , pp. 993–1022.

Boyaci, B., Zografos, K. G. and Geroliminis, N. (2015) 'An optimization framework for the development of efficient one-way car-sharing systems', European Journal of Operational Research, Vol. 240 No. 3, pp. 718–733.

Costa, E., Lorena, A., Carvalho, A. and Freitas, A. (2007) 'A review of performance evaluation measures for hierarchical classifiers', in Evaluation Methods for machine Learning II: papers from the AAAI-2007 Workshop, pp. 1–6.

Frost & Sullivan (2016) 'Future of Carsharing Market to 2025'. Available at: https://www.statista.com/statistics/415636/car-sharing-number-of-users-worldwide/.

Guglielmetti Mugion, R., Toni, M., Di Pietro, L., Pasca, M. G. and Renzi, M. F. (2019) 'Understanding the antecedents of car sharing usage. An empirical study in Italy', International Journal of Quality and Service Sciences, Vol. 11 No. 4, pp. 523–541.

Guo, Y., Barnes, S. J. and Jia, Q. (2017) 'Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation', Tourism Management, Vol. 59 , pp. 467–483.

Illgen, S. and Höck, M. (2019) 'Literature review of the vehicle relocation problem in one-way car sharing networks', Transportation Research Part B: Methodological, Vol. 120 , pp. 193–204.

Jeong, Y., Yang, Y., Suk, J. and Kim, K. (2019) 'A text-mining analysis of online reviews on car-sharing services', in International Conferences on Internet Technologies & Society 2019, pp. 167–169.

Jivani, A. G. (2011) 'A comparative study of stemming algorithms', International Journal of Computer Applications in Technology, Vol. 2 No. 6, pp. 1930–1938.

Mastrogiacomo, L., Barravecchia, F. and Franceschini, F. (2019) 'A worldwide survey on manufacturing servitization', The International Journal of Advanced Manufacturing Technology, Vol. 103 , pp. 3927–3942.

Mastrogiacomo, Luca, Barravecchia, F. and Franceschini, F. (2020) 'Definition of a conceptual scale of servitization: Proposal and preliminary results', CIRP Journal of Manufacturing Science and Technology, Vol. 29 No. Part B, pp. 141–156.

Mastrogiacomo, L., Barravecchia, F. and Franceschini, F. (2020) ‘Enabling factors of manufacturing servitization: empirical analysis and implications for strategic positioning’, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, Vol. 234 No. 9, pp. 1258–1270.

Meyer, D., Hornik, K. and Feinerer, I. (2008) ‘Text mining infrastructure in R’, Journal of statistical software, Vol. 25 No. 5, pp. 1–54.

Möhlmann, M. (2015) ‘Collaborative consumption: Determinants of satisfaction and the likelihood of using a sharing economy option again’, Journal of Consumer Behaviour, Vol. 14 No. 3, pp. 193–207.

Müller, O., Junglas, I., vom Brocke, J. and Debortoli, S. (2016) ‘Utilizing big data analytics for information systems research: challenges, promises and guidelines’, European Journal of Information Systems, Vol. 25 No. 4, pp. 289–302.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y. and Ngo, D. C. L. (2014) ‘Text mining for market prediction: A systematic review’, Expert Systems with Applications, Vol. 41 No. 16, pp. 7653–7670.

Prescient & Strategic Intelligence (2019) Carsharing Market by Car, by Fuel Type, by Business Model, by Application, by Geography Global Market Size, Share, Development, Growth, and Demand Forecast, 2014-2025. Available at: https://www.psmarketresearch.com/market-analysis/car-sharing-market.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B. and Rand, D. G. (2014) ‘Structural topic models for open-ended survey responses’, American Journal of Political Science, Vol. 58 No. 4, pp. 1064–1082.

Roberts, M. E., Stewart, B. M. and Tingley, D. (2019) ‘STM: R package for structural topic models’, Journal of Statistical Software, Vol. 91 No. 2, pp. 1–40.

Scott, J. and Baldridge, J. (2013) ‘A recursive estimate for the predictive likelihood in a topic model’, Journal of Machine Learning Research, Vol. 31 , pp. 527–535.

Shaheen, S. A. and Cohen, A. P. (2007) ‘Growth in worldwide carsharing: An international comparison’, Transportation Research Record, Vol. 1992 No. 1, pp. 81–89.

Shaheen, S. A. and Cohen, A. P. (2013) ‘Carsharing and personal vehicle services: worldwide market developments and emerging trends’, International Journal of Sustainable Transportation, Vol. 7 No. 1, pp. 5–34.

Wallach, H. M., Mimno, D. M. and McCallum, A. (2009) ‘Rethinking LDA: Why priors matter’, Advances in neural information processing systems, Vol. 23 , pp. 1973–1981.

Zaki, M. and McColl-Kennedy, J. R. (2020) ‘Text Mining Analysis Roadmap (TMAR) for Service’, Journal of Services Marketing, Vol. 34 No. 1, pp. 30–47.