



Identifying User Experiences for Decision-Making in Service Science

Silvana Aciar^{1,2(✉)}, Mayela Coto³, and Gabriela Aciar²

¹ CONICET, Buenos Aires, Argentina
saciar@unsj-cuim.edu.ar

² Universidad Nacional de San Juan, San Juan, Argentina
gaby_aciar@yahoo.com.ar

³ Universidad Nacional Costa Rica, Heredia, Costa Rica
mcoto@una.ac.cr

Abstract. The purpose of this paper is to propose a method based on text mining techniques that employs user comments on the Web to clarify users' experiences of a service. The method classifies user experiences into the categories of instrumental qualities, non-instrumental qualities, and emotional reactions. The Text-Miner Software Kit (TMSK) and Text Rule Induction Kit (RIKTEXT) were used to obtain the classification rule sets. A case study on the tourist domain was conducted to assess the rules' precision in classifying sentences into the three user experience categories. One thousand hotel reviews from the website www.tripadvisor.com were used in the evaluation process. Two datasets were built: one for training and one for testing. For validation, 100 new opinions were used to assess the rules' precision. The results are considered good because there was only a 10% error in automatic classification versus manual classification. Based on this result, we argue that the automatic identification of information about user experience of a service can be done accurately using text mining techniques, such as those presented in this work.

Keywords: User experience · Opinion mining · Ontology

1 Introduction

Service providers must have the capacity to adapt to the dynamic, globalized market with greater access to technology [1, 2], and services should be oriented to customers and their expectations. Tools that help convert data into useful information for decision-making play an important role in knowing the value of use that customers give to services [2, 3]. Analyzing this value provides important indicators for evaluating and improving the services. In this sense, it is relevant to identify the sources containing user information that can be used to obtain useful indicators for decision-making.

Currently, one of the main sources of information is the Web, where most information is in the form of unstructured text, emails, personal or institutional webpages, comments on social networks, digital books, newspapers, and databases. The process for extracting this information was developed approximately 80 years ago [4–6]. Given the amount of information that exists, it is almost impossible to retrieve it, index it, and

interpret it manually. Therefore, text mining techniques are used to identify patterns, predict usage trends, identify semantic structures, and classify instances and entities automatically [5, 7]. Text mining allows service providers to obtain relevant information for decision-making, for example, by analyzing customers' comments and opinions, which are available on the Web, and adapting their services according to the customers' needs and preferences [8–10].

The purpose of this paper is to propose a method based on text mining techniques that will allow for understanding users' experiences of a service on the basis of their written comments on the Web. For this, user experiences were conceptualized, recovered, and quantified. The user's experience of a service was classified into the categories of instrumental qualities, non-instrumental qualities, and emotional reactions [11]. A case study on the tourist domain was conducted to assess the accuracy of the rules in the classification of sentences in the three categories.

This study contributes to the growing literature on service science, and especially to the field of user experience, by introducing an automatic method for obtaining and presenting the key factors that influence the success of a service focused on the experiences of its users. The remainder of this paper will proceed as follows. Sections 2 and 3 present the main theoretical concepts related to the study, while Sect. 4 provides an overview of the methods for establishing the sets of rules. Section 5 presents the controlled experiment that was conducted to assess the validity of the rules. The paper ends with concluding remarks and future work in Sect. 6.

2 Science of Services and User Experience

In general terms, a service has an economic value but lacks material consistency [2, 3]. The science of services is the application of scientific disciplines to add value to services while satisfying the needs of their users [11]. Through the application of models and theories, the science of services fosters innovation, competition, and quality through the joint creation of value by providers and users [12].

One of the most important challenges in the science of services is to obtain information to feed the analysis process, which allows the decision-making associated with continuous improvement of the service design [2, 3, 11]. The more user information that can be obtained or more variables that can be measured, the greater the opportunity to improve the service. For example, a variable such as the user's previous behavior can contribute to improving new versions of a service. Continually innovating services benefits both providers and users, who will continue to use a service if they are satisfied with it.

The design of services has migrated from a design specified by specialists to one that is focused on the user. Previously, most companies and entities providing services focused on analyzing the economic benefits or quality of products and overlooked the internal processes that produced those results [13]. A user-centered design refers to the

use of methods that ensure that services have high levels of usability [13, 14]. A service's degree of usability is influenced, to a large extent, by the user's experience of interacting with the service or previous interactions with similar services.

The user experience is defined as an integrating concept of the interaction between the end-user and the company, its services, and products [15, 16]. User satisfaction is the result of the actual user experience and user expectations of the service. As mentioned earlier, one of the most important challenges in designing services is how to obtain information about users' expectations and experiences of a service to optimize its design.

3 Evaluation of User Experience and Opinion Mining

Numerous methods have been developed to obtain information on user experience, including laboratory studies with users, surveys, and evaluation by experts [17–19]. Laboratory methods involve inviting participants to use prototype versions of products or services. Psychophysical variables are measured, and experts observe all user actions to detect usability problems as well as the emotions associated with the product or service in a possible context of use. Surveys are generally used to obtain information from users after they have interacted with a product or service. Online surveys facilitate obtaining comments from a large number of users in a short time. Another method of evaluation is using expert knowledge. Usability experts are recruited to inspect a product or service design according to usability heuristics. This method is typically used in the early design stages, before the product or service is evaluated by users, and can be more expensive in both time and money.

Of the aforementioned methods, the most popular are surveys [17, 18], which consist of a group of questions specially designed to a certain how users use a product or service. The main advantage of surveys is that they elicit concrete answers, allowing for a statistical treatment to analyze the results. However, there are some drawbacks. For example, (1) there are few opportunities for the user to express themselves about other aspects not considered in the survey and (2) people are generally reluctant to answer surveys, resulting in few or incomplete answers. Given these problems, alternative means of obtaining information about user experience are social networks, forums, and opinion blogs [20, 21].

Extracting user experience from these sources allows the providers to make good decisions, improve versions of the product/service, or correct any problems detected from customer experiences. The essential difference in the information obtained from networks compared to that obtained by surveys is the immediacy. This information is spontaneous and unstructured and has great value for companies and their business strategies, customer service, and trend detection [5–7].

However, manually analyzing all these opinions would take a long time given their volume and variety. Therefore, opinion mining arises with the purpose of automating the analysis of user opinion information. Opinion mining is an extension of text mining. While text mining entails automatically analyzing any text, opinion mining automatically analyzes user opinions about a product, person, entity, or service [6]. There are numerous applications of opinion mining [5], such as the analysis of opinions

on a product or service in marketing, analysis of opinions on a political candidate, or the compilation of opinions considered as “potential threats” to society, as in the case of the fight against terrorism or in the defense industry.

The objective of opinion mining is to classify positive, negative, and neutral opinions on the basis of certain criteria indicating to which of these three categories a given expression belongs. For this, it is necessary to identify what is most likely to indicate that an opinion is considered positive or negative. The methods used for opinion processing can be divided into two categories—lexical-based and learning-based [3, 4]—although both are currently used together to improve the performance of opinion mining algorithms.

Some of the techniques used in these methods are machine learning (ML), natural language processing (NLP), and information retrieval (IR). These three techniques are not mutually exclusive. The most common strategy is applying several of them to obtain stronger results than those obtained by applying only one technique [3, 5, 6]. My Opinion Tools is a tool that can process large volumes of data in a short time, with high accuracy, at a low cost.

4 Text Mining to Extract User Experience Information

The objective of this study was to identify user experience automatically based on the opinions written by users. To achieve this goal, it was necessary to obtain association rules that would allow for discovering the words in a text associated with user experience.

4.1 User Experience Categories

To classify text relating to user experience, categories or classes of user experience had to be identified. In this paper, the dimensions defined in [22] are taken as categories of user experience: instrumental qualities, non-instrumental qualities, and emotional reactions.

The instrumental qualities are related to the technical characteristics and effectiveness of the function provided by the product, system, or service; self-description; and the ability to control its operation. Non-instrumental qualities refer to design features, such as materials, shapes, colors, and location. Emotional reactions include emotions, which consist of motor expressions and subjective feelings, which can be positive or negative. Figure 1 shows the text mining process for obtaining rules enabling the classification of user opinions into categories of user experience.

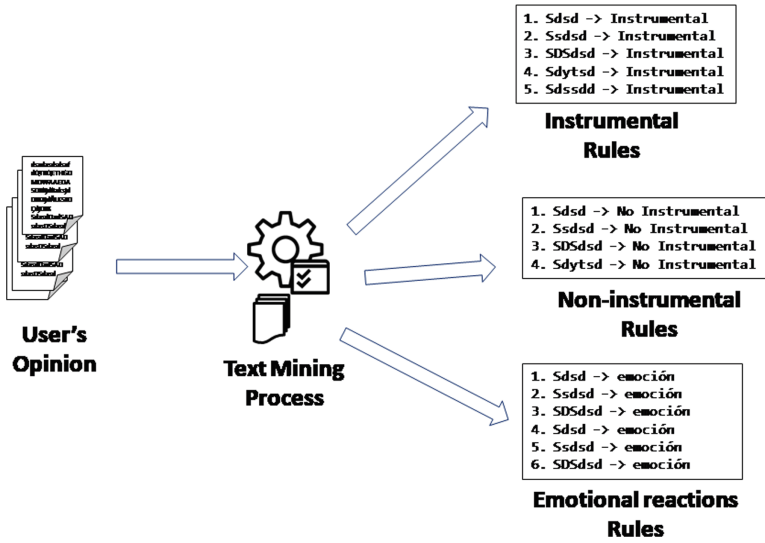


Fig. 1. Text mining to obtain information about user experience from opinions written on the Web.

4.2 Text Mining Process

Once the user experience categories were defined, the next step was the classification process, which followed the text extraction process described in [23]. Each sentence of the user's opinion was classified by means of rules-based classification techniques into either the "Instrumental", "Non-instrumental", or "Emotional Reaction" category. The Instrumental category included all sentences containing information on objective aspects of the service, such as if it fulfilled the objectives for which it was created. The Non-instrumental category grouped together all sentences containing information about the physical characteristics of the service, such as location, access to it, and payment method. Finally, the Emotional Reaction category contained all sentences describing the feelings caused by the user's interaction with the service.

The Text-Miner Software Kit (TMSK) and Text Rule Induction Kit (RIKTEXT) were used to obtain the classification rule sets [24]. TMSK generates a dictionary from a set of documents (opinions in our case) and converts a set of sentences into vectors, which creates a dictionary of relevant words. Preprocessing must be done to place user opinions in an XML file. This format is required by TMSK. The "Vectorize" function of TMSK creates the vectors from the XML documents. Documents are converted into an array format where each row corresponds to a document and each column corresponds to a word in the dictionary. RIKTEXT uses the dictionary and vectors representing each category to train a classifier, consequently returning the set of rules for each category.

The best set of rules was selected on the basis of the complexity and error rate of each rule. RIKTEXT found the set of rules with the minimum error rate and then identified a less complex set of rules whose error rate was reasonably close to the

minimum error rate. The concept of “reasonably close” was governed by the set of properties specified by the standard error number. By default, this was set to 1 such that “reasonably close” meant “within a standard error.”

4.3 Obtaining the Classification Rules

To obtain the classification rules, a case study in the tourism domain was analyzed—specifically, the opinions of hotel users. The data included 1,000 reviews of 10 hotels in the province of San Juan, Argentina, obtained from [www.tripadvisor](http://www.tripadvisor.com). Opinions were randomly selected from a set of more than 2,000 opinions. As the opinions were not in XML format, the use of a special processing program was necessary to transform the data. Each sentence of each opinion was treated as a document. Volunteers were asked to manually label sentences in one of the experience categories: Instrumental, Non-instrumental, and Emotional Reaction. Each sentence was labeled by three people. To resolve classification conflicts in cases where the sentences were labeled differently by the three people, they were labeled again by three other people to reach a consensus through the voting system.

Once the data were labeled and in XML format, they were ready to be processed by TMSK to generate the dictionary and a set of tagged vectors. A 950-word dictionary was generated, which was used to generate the vectors. The vectors were divided into portions of training and tests. Test cases were randomly selected in RIKTEXT to determine how many cases should be used for testing. Two thirds of the cases were defined for training and the rest for validation tests. Figures 2, 3, and 4 present the sets of rules obtained to classify sentences within the categories of Instrumental, Non-instrumental, and Emotional Reaction, respectively.

Rule Set	Number of Rules	Variables	Training error	Test error	Test standard deviation	Error/variable
1	23	26	0,1696	0,3650	0,0449	0,00
2	23	25	0,1736	0,3650	0,0449	1,01
3	22	23	0,1795	0,3450	0,0435	2,34
4	18	18	0,2091	0,3250	0,0421	3,34
5	12	12	0,2249	0,3050	0,0397	4,34
6	10	10	0,2308	0,2850	0,0393	5,04
7	8	8	0,2409	0,2700	0,0321	5,74
8*	6	6	0,3222	0,2567	0,0265	6,44
9	5	5	0,3201	0,2799	0,0378	7,14
10	1	1	0,3470	0,6319	0,0423	1,33

*best rule set according to error rate and simplicity

Selected Rule Set

1. fácil & comunicarse → Instrumental
2. eficientes → Instrumental
3. novedoso → Instrumental
4. seguro & no → Instrumental
5. tradicional → Instrumental
6. amistoso → Instrumental

Fig. 2. Set of rules for classifying sentences in the Instrumental category of user experience

Rule Set	Number of Rules	Variables	Training error	Test error	Test standard deviation	Error/variable
1	23	24	0,1096	0,2571	0,0360	0,00
2*	13	13	0,1136	0,2571	0,0360	0,61
3	11	11	0,1195	0,2733	0,0365	1,94
4	10	10	0,1491	0,3059	0,0374	2,94
5	9	9	0,1649	0,2896	0,0370	3,94
6	8	8	0,1708	0,4115	0,0390	0,18
7	1	1	0,1809	0,4059	0,0404	0,12

*best rule set according to error rate and simplicity

Selected Rule Set

1. limpieza → No instrumental
2. ubicación → No instrumental
3. restaurant → No instrumental
4. habitaciones → No instrumental
5. servicios → No instrumental
6. propietarios → No instrumental
7. ruido → No instrumental
8. baños → No instrumental
9. desayuno → No instrumental
10. transporte & aeropuerto → No instrumental
11. atracciones → No instrumental
12. desayuno → No instrumental
13. recepción → No instrumental

Fig. 3. Set of rules for classifying sentences in the Non-instrumental category of user experience

Rule Set	Number of Rules	Variables	Training error	Test error	Test standard deviation	Error/variable
1	19	33	0,4696	0,3050	0,0549	0,00
2	16	27	0,4736	0,3050	0,0549	1,41
3	12	17	0,4795	0,2850	0,0349	2,74
4*	8	8	0,5091	0,2650	0,0149	3,74
5	6	6	0,5249	0,2950	0,0449	4,74
6	5	5	0,5308	0,3250	0,0749	5,44
7	4	4	0,5409	0,3550	0,1049	6,14
8	3	3	0,6222	0,3850	0,1349	6,84
9	2	2	0,6201	0,4150	0,1649	7,54
10	1	1	0,6470	0,4450	0,1949	8,45

*best rule set according to error rate and simplicity

Selected Rule Set

1. feliz → Instrumental
2. furioso → Instrumental
3. sorpresa → Instrumental
4. contento → Instrumental
5. asco → Instrumental
6. fantástico → Instrumental
7. horrible → Instrumental
8. muy & bueno → Instrumental

Fig. 4. Set of rules for classifying sentences in the Emotional Reaction category of user experience

The chosen rules were those that had a minimum error rate or were very close to the minimum, but perhaps simpler than the minimum (*). The result of the rules validation is shown in Table 1, which depicts the measurements of precision, recall, and F-measure for the training data and test data.

Table 1. Classification process evaluation

Measures	Categories					
	Instrumental		Non-instrumental		Emotional reaction	
	Training data	Test data	Training data	Test data	Training data	Test data
Precision	71.6189	67.5816	73.0299	70.3844	74.0881	60.0140
Recall	89.2448	76.9371	54.1316	44.2000	64.5301	37.5140
F-Measure	79.4661	71.9564	62.1762	54.2997	68.9795	46.1678

5 Validation

Once the set of rules for classifying sentences in the different user experience categories was obtained, a controlled experiment was conducted to assess the validity of the rules. For this, 100 new opinions about San Juan hotels from www.tripadvisor.com were used. A computer algorithm was created to automatically apply the classification rules. The rules were applied to each hotel opinion, and sentences were obtained for each of the user experience categories. An example is shown in Table 2.

Table 2. Sample sentences containing information about the user's experience of the hotel

Category	Hotel reviews sentences
Instrumental	I was satisfied by the presence of a guard, very attentive. Highly recommended
Non-instrumental	The beds made noise. It is very close to the city center
Emotional reaction	Very dirty, disgusting, and depressing

Additionally, the same group of three volunteers made a manual classification of the 100 new opinions to identify whether they contained instrumental, non-instrumental, emotional reaction, or irrelevant information. Table 3 shows a comparison of the results obtained with the automatic classification and the manual classification, measured by the mean absolute error in each category. There was an average automatic classification error of 0.10.

Table 3. Automatic classification Vs manual classification results

Sentences	Automatic classification	Manual classification	Absolute average error
Non-instrumental	355	367	0.12
Instrumental	98	106	0.08
Emotional reaction	124	115	0.09
Irrelevant	123	112	0.11
Total sentences	700	700	0.10

Of the success cases, more than 96.7% of the sentences that were manually classified in the non-instrumental category were also classified in this category by the automatic method using the defined rules. Likewise, 92.4% of the sentences in the instrumental category were classified well by the automatic method, and 100% of the sentences classified manually in the Emotional Reaction category were also classified in this category by the automatic method. As these data demonstrate, in general terms, the defined rules had a high degree of effectiveness in the three categories of user experience.

An analysis of the error cases revealed that they were generated mainly for two reasons. First, there were words in the sentences that were not considered in the rules or in the synonyms dictionary. For example, a sentence classified by the automatic method as irrelevant contained the word “windows,” which was not synonymous with the word “room” in the rules of the Non-instrumental category. Second, some sentences that were automatically classified into a single category were classified into two categories in the manual classification because they were too long and contained information from more than two categories of user experience.

6 Conclusion

This paper presents an automatic classification process of user experience based on text mining techniques. Information about user experience of a service can be obtained using this process, but identifying such information is not an easy task.

A case study on the tourist domain was analyzed to assess the rules' precision for the process of classifying sentences in three categories of user experience defined in the literature: Instrumental, Non-instrumental, and Emotional Reaction. One thousand reviews of hotels in the province of San Juan, Argentina were obtained from the website www.tripadvisor.com. Two data sets were built: one for training and one for testing. To evaluate the rules' precision, 100 new opinions were used. The results obtained are considered good because there was only a 10% error against a classification made manually. On the basis of this result, we can conclude that the automatic identification of information about users' experience of a service can be done accurately using text mining techniques, such as those presented in this work.

That is, the proposed method allows for automatically filtering relevant information regarding user experience of services/products from thousands of opinions written in comments on the Web. As mentioned in this paper, such techniques are less intrusive

and less expensive than conducting surveys of end users. Filtering user experience information using text mining techniques goes beyond just knowing if the opinion is positive or negative; relevant comments should be filtered from user experience information that involve physical, objective, and emotional qualities. The rules obtained through the text mining techniques in this study are a first step toward obtaining that kind of information about user experience.

In future work, the rules will be refined by updating the synonyms dictionary to improve the classification process. A way to deal with sentences that may belong to more than one category will also be analyzed. In addition, an interface will be created to summarize and provide this information in an understandable way for decision-makers of the entities that provide services. This interface must provide relevant information, for decision-making, about the experiences of users who have used a product or service. Once the interface is created, it must be validated with future users (service providers).

References

1. Maglio, P.P., Kieliszewski, C.A., Spohrer, J.C., Lyons, K., Patricio, L., Sawatani, Y. (eds.): Handbook of Service Science, Volume II. SSRISE. Springer, Cham (2019). <https://doi.org/10.1007/978-3-319-98512-1>
2. Peters, C., et al.: Emerging digital frontiers for service innovation. CAIS **1**(39) (2016). <https://doi.org/10.17705/1cais.03908>
3. Netzer, O., Feldman, R., Goldenberg, J., Fresko, M.: Mine your own business: market-structure surveillance through text mining. Mark. Sci. **31**(3), 521–543 (2012)
4. Hsiao, Y.H., Chen, M.C., Liao, W.C.: Logistics service design for cross-border E-commerce using Kansei engineering with text-mining-based online content analysis. Telematics Inform. **34**(4), 284–302 (2017)
5. Aggarwal, C.C., Zhai, C. (eds.): Mining Text Data. Springer, Boston (2012). <https://doi.org/10.1007/978-1-4614-3223-4>
6. Mostafa, M.M.: More than words: social networks' text mining for consumer brand sentiments. Expert Syst. Appl. **40**(10), 4241–4251 (2013)
7. Amado, A., Cortez, P., Rita, P., Moro, S.: Research trends on big data in marketing: a text mining and topic modeling based literature analysis. Eur. Res. Manag. Bus. Econ. **24**(1), 1–7 (2018)
8. Gan, Q., Ferns, B.H., Yu, Y., Jin, L.: A text mining and multidimensional sentiment analysis of online restaurant reviews. J. Qual. Assur. Hosp. Tour. **18**(4), 465–492 (2017)
9. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl.-Based Syst. **89**, 14–46 (2015)
10. Sun, S., Luo, C., Chen, J.: A review of natural language processing techniques for opinion mining systems. Inf. Fusion **36**, 10–25 (2017)
11. Vargo, S.L., Wieland, H., Akaka, M.A.: Innovation in service ecosystems. J. Serviceol. **1**(1), 1–5 (2016)
12. Ostrom, A.L., et al.: Moving forward and making a difference: research priorities for the science of service. J. Serv. Res. **13**(1), 4–36 (2010)
13. Lin, C.J., Cheng, L.Y.: Product attributes and user experience design: how to convey product information through user-centered service. J. Intell. Manuf. **28**(7), 1743–1754 (2017)

14. Chilana, P.K., Ko, A.J., Wobbrock, J.: From user-centered to adoption-centered design: a case study of an HCI research innovation becoming a product. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1749–1758. ACM, New York, April 2015
15. Bilgihan, A.: Gen Y customer loyalty in online shopping: an integrated model of trust, user experience and branding. *Comput. Hum. Behav.* **61**, 103–113 (2016)
16. Lallemand, C., Gronier, G., Koenig, V.: User experience: a concept without consensus? Exploring practitioners' perspectives through an international survey. *Comput. Hum. Behav.* **43**, 35–48 (2015)
17. Trischler, J., Scott, D.R.: Designing public services: the usefulness of three service design methods for identifying user experiences. *Public Manag. Rev.* **18**(5), 718–739 (2016)
18. Weast, J.C., et al.: US Patent Application No. 15/168,084 (2016)
19. Helms, K., Brown, B., Sahlgren, M., Lampinen, A.: Design methods to investigate user experiences of artificial intelligence. In: 2018 AAAI Spring Symposium Series. Stanford University, Palo Alto, California USA, March 2018
20. Milovanovic, S., Bogdanovic, Z., Labus, A., Barac, D., Despotovic-Zratic, M.: An approach to identify user preferences based on social network analysis. *Future Gener. Comput. Syst.* **93**, 121–129 (2019)
21. Khalaj, J., Pedgley, O.: A semantic discontinuity detection (SDD) method for comparing designers' product expressions with users' product impressions. *Des. Stud.* **62**, 36–67 (2019)
22. Van Der Linden, J., Amadiou, F., Vayre, E., Van De Leemput, C.: User experience and social influence: a new perspective for UX theory. In: Marcus, A., Wang, W. (eds.) *HCI 2019*. LNCS, vol. 11583, pp. 98–112. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23570-3_9
23. Aciar, S., Zhang, D., Simoff, S., Debenham, J.: Informed recommender: basing recommendations on consumer product reviews. *IEEE Intell. Syst.* **22**(3) (2007)
24. Weiss, S.M., Indurkha, N., Zhang, T., Damerou, F.: *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer, New York (2004). <https://doi.org/10.1007/978-0-387-34555-0>