A Method for Supporting Customer Model Construction: Using a Topic Model for Public Service Design

Satoshi Mizoguchi, Takatoshi Ishii, Yutaro Nemoto, Maiko Kaneda, Atsuko Bando, Toshiyuki Nakamura, and Yoshiki Shimomura

Abstract

For the design of public services, it is important to clarify service customers. For this purpose, various methods of customer modeling were proposed. Before constructing customer models, it is required to group customers and to characterize each customer group. However, the customer grouping based on some statistical barometers (e.g. age, sex, and job categories) may not reflect actual customer requirements for the service. This paper aims to propose a method for supporting customer grouping and characterizing without such statistical barometers. Finally, the proposed method is applied to an urban development case to demonstrate the effectiveness.

Keywords

Public service design • User modeling • Natural language processing • Latent Dirichlet allocation

1 Introduction

Recently, services have been attracting much attention from both academic and industrial sides. In general, public service design involves various customers who have different requirements for the service. For example, in a regional environment improvement project, some customers want to enhance the safety of their children, and other customers want to widen the road to solve traffic jam. Thus, in order to realize high value in public service, designers need to accommodate various customer requirements

To analyze requirements for public services, it is important to clarify various customers. For this purpose, there are a lot of customer modeling methods (e.g., persona method [1], Delphi method [2], and conjoint analysis [3]). In particular, persona method [1] is often used for the design of public services. Watanabe et al. construct stakeholder models for the next generation of agriculture by using persona method [4]. By sharing personas of the stakeholders, they designed social services which can solve local problems.

Before creating personas, it is usually required to generate customer clusters and characterize each cluster. For the customer clustering, existing methods generally use statistical barometers which can be easily acquired (e.g., age, sex, and job categories). However, the customer clustering based on such statistical barometers does not necessarily reflect their requirements for the service. Customer clusters should be constructed based on the similarity of customers' requirements. For example, there are two customers who are both 35 years old, female and homemaker. One customer has a baby and wants to remove bicycle parking on a road for pushing her baby carriage. The other customer has a 5-year-old child and wants to park bicycle on a road for going a children's garden by bicycle. If the statistic barometers are only used to cluster these two customers, they may be classified into same cluster even though

S. Mizoguchi (✉) • T. Ishii • Y. Nemoto • Y. Shimomura
Department of System Design, Tokyo Metropolitan University, Asahigaoka 6-6 Hino-shi, Tokyo 191-0065, Japan
e-mail: mizoguchi-satoshi@ed.tmu.ac.jp

M. Kaneda • A. Bando • T. Nakamura
Design Division, Hitachi Ltd., Akasaka Biz Tower, 3-1, Akasaka 5-chome Minato-ku, Tokyo 107-6323, Japan

Y. Sawatani et al. (eds.), *Serviceology for Smart Service System*, DOI 10.1007/978-4-431-56074-6_3

they have the opposite requirements concerning bicycle parking

To solve this problem, this paper proposes a method for supporting customer clustering and characterizing without such statistical barometers. The proposed method aims to cluster customers by using free description concerning customer requirements. In this paper, to demonstrate the usability of the method, we show an application to an urban development case.

2 Literature Review

In order to clarify issues in customer modeling, this chapter summarizes a review on existing methods for customer model construction.

2.1 Persona Method

Cooper proposed persona method in which customers are modeled as "persona" [1]. The "persona" is a fictional character that is described by some barometers such as statistical barometers (e.g., age, sex, and job categories) and scenarios customer are using (e.g., what customer uses, what product the customer uses, when the customer uses it, how the customer uses it). The persona is useful for clarifying and sharing images of customer and is used in both product and service design.

2.2 User Modeling Based on ID-POS

Motomura proposed a user modeling method based on ID-POS data [5]. This method estimates customer categories from purchase behaviors. For example, some people who often buy a high-price beer are characterized as high-end customers, and other people who buy a low-malt beer are characterized as a budget-minded customer. These customer categories indirectly appear on statistical data (ID-POS data). For estimating the customer categories, this method employs a topic model. Topic models estimate "topics" (theme or subject) of documents from the words that frequently appear in the documents. By using a topic model, this method estimates user category from the purchase log: how many times and what items the user purchases. In similar way, this method also estimates product category. From these results, this method estimates customer purchase behaviors by using Bayesian network [6]. The analysis results by using the method reveal "what product category does a customer often buy?" and "what customer category is a product probably bought by?"

2.3 Topic Analysis of Web User Behavior on Proxy Logs

Fujimoto et al. proposed a topic model for web user profiling and clustering [7]. This method estimates the abstract purpose (abstracted user intentions or tasks) of web page access. For example, people who access a site about hotel have an abstract purpose, they want to seek hotel costs, and people who access a site about a local city have an abstract purpose, they want to get information of the local city. These abstract purposes of web access indirectly appear on statistical data (web access log data). This method employs a topic model by regarding a user's web access log as a document and a web address as a word. By using this method, the authors found out 24 abstract purposes of web access from the data obtained from students in Osaka University. A part of the abstract purposes is as follows: "YouTube user," "Wikipedia user," "job hunter," "programming," and "how to make a report."

2.4 Scope of This Study

For the design of public services, it is important to cluster customers for requirement analysis. The cluster should be constructed based on customer requirements (or its similarity). However, customer clustering based on only statistical barometers (e.g., "persona method" [1]) probably does not reflect customer requirements, because it is difficult to find what barometers point out difference of customer requirements.

In addition, there are some methods that estimates the barometer that point out difference of customer requirements (e.g., Motomura [5] and Fujimoto [7]). However, these methods require a large quantity of operating data (e.g., ID-POS or web access log). Thus, it is difficult to apply these methods to the analysis for some cases such as new public service design. Additionally, it is difficult to apply these methods for public service, because of the following two reasons:

1. Owing to a variety of customers, observing data of public service requires huge costs.
2. It is hard to know what data is related to various requirements.

To solve these problems, this paper proposes a method for requirement analysis without using the statistical barometers.

3 Proposed Method

This chapter introduces the proposed method. The proposed method supports service designers to clarify customers by providing the requirement similarity of each customer. The method calculates the requirement similarity from free descriptions about customers' requirements. For calculating the similarity, we employ latent Dirichlet allocation (LDA) that is a method for analyzing natural language.

3.1 Latent Dirichlet Allocation

There are some topic models: latent semantic analysis (LSA) [8], probabilistic LSA (pLSA) [9], latent Dirichlet allocation (LDA) [10], and more. From these topic models, this study employs LDA for estimating customer requirements. LDA assumes and models that the document includes some abstract "topics" that are subject of documents. LDA estimates topic of each word from bias of word frequency in the documents. By aggregating the topics of words in the document, LDA estimates the topic allocation of each document.

For example, there is an example document: "when I went to a local city for a business trip, I want to go to the local tourist attraction and eat a local food." In this example, the result of LDA indicates that the word "tourist attraction" has a topic of "sightseer" and the other word "business" has a topic of "salesman." Thus, this document has two main topics: "sightseer" and "salesman."

LDA assumes that each document has multiple topics. On the other hand, LSA and pLSA assume that the document consist of one abstract topic. Thus, the assumption of LDA is fitter actual document and has better accuracy than LSA and pLSA.

In our study, we assume that free descriptions concerning customer requirements include some topics of requirement. For example, for analyzing city users, a description written by an office worker from other city will include two types of requirement "as a sightseer" and "as a salesman." LDA is able to model such case that various topics are included in a document. Therefore, in this study, LDA is employed for estimating topics of customer requirement.

3.2 A Method for Supporting Customer Model Construction

An overview of the proposed method is shown in Fig. 1.

This method assumes that free descriptions concerning customer requirement have several topics that may be represented as customer's lifestyle (e.g., as a sightseer and/or as a salesman). By using LDA, the proposed method estimates the topics of requirement from the free descriptions. The questionnaire for customers can be used as these descriptions. Then, this method clusters the questionnaires based on estimated topics (and this clustering result is represented as customer clustering). Moreover, this method characterizes customer clusters based on estimated topics. Thus, our method is expected to cluster the customers based on customer requirements.

This method includes four steps;

Step1: *Obtain data about customer requirements*
In Step1, data about customer requirement are obtained. To be more precise, this method requires free descriptions from each customer. One way in getting those data is conducting survey by giving out questionnaires to actual customers. In this case, items in the questionnaire are needed to elicit customer requirements, for example, "what is the important point in the service use?" "why do you emphasize this point?" and so on. In the following steps, a customer is represented by the information obtained in the questionnaire.

Step2: *Estimate topics and a topic allocation of each document*
Step2 estimates topics and a topic rate of each document. In this step, it applies LDA analysis to the free description data obtained in Step1. LDA outputs estimated topic allocation (rate) of each document and representative words of each topic.

Step3: *Cluster customers based on topic rate*
Step3 classifies the free descriptions (i.e., the customers) based on topic rates estimated in Step2. This method regards this classification result as clusters of the customers.

Step4: *Construct customer models based on these results*
In Step4, customer models are constructed based on results of Step2 and Step3. To be more precise, each cluster is characterized based on the representative words for each topic estimated in Step2. For example, if a customer cluster has a topic that includes "business trip," "work," "convenience," and "drink" as representative words, this cluster can be characterized as a "salesman" cluster.

This method is expected to cluster the customer based on customer requirements and to support the analysis of customer requirements. The advantage of our method is able to apply for some cases such as public service design. The advantage of this method is that the free description data is easier to obtain than the operating data.

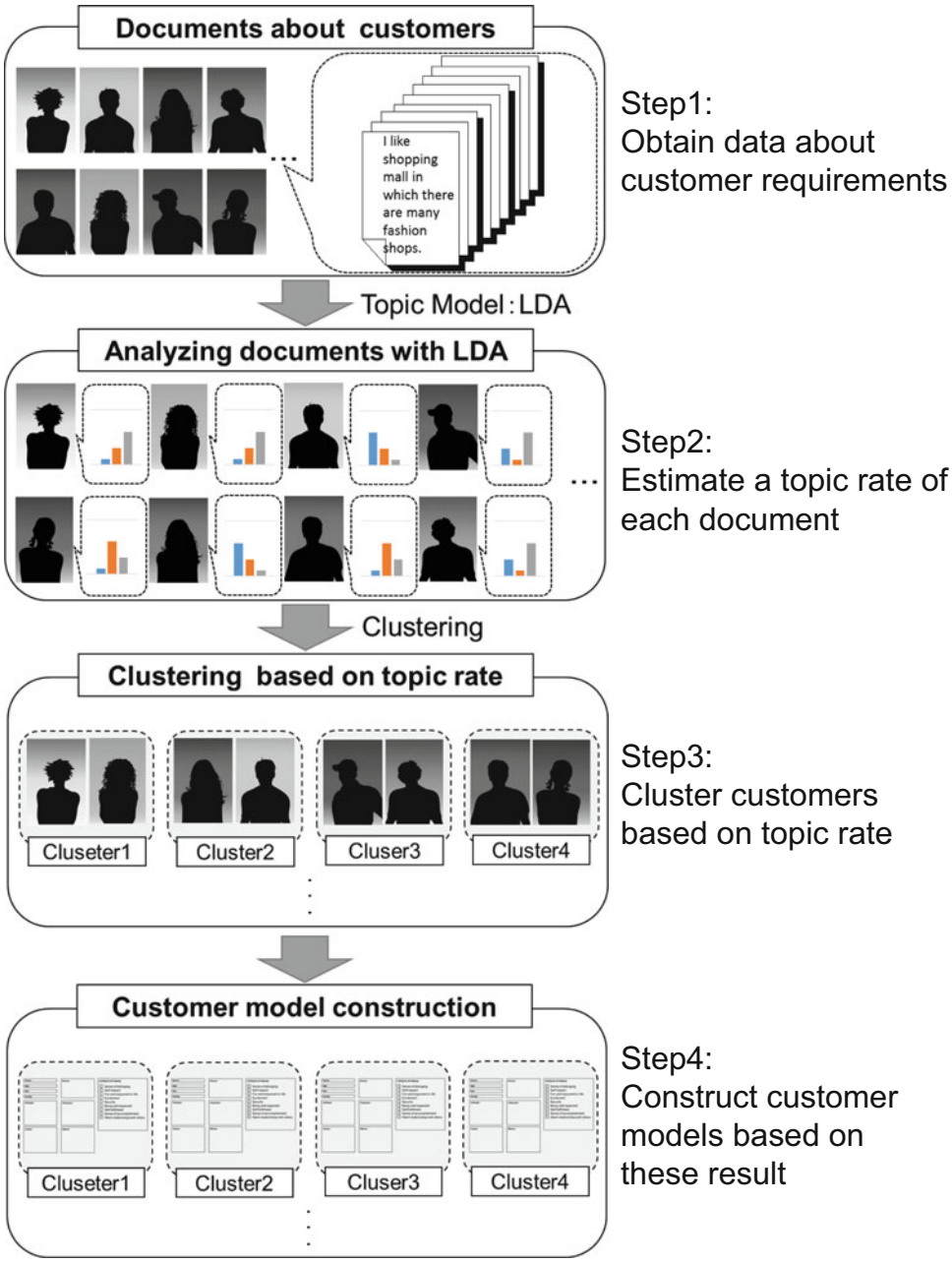

Fig. 1 Overview of the proposed method

4 Application

For checking the advantage of the proposed method, this chapter shows an application result. In this application, the proposed method was applied to an urban development case which can be regarded as a public service design.

4.1 Application to Urban Development

This case is about the development of Tenjin area, a downtown in Fukuoka, Japan. In Tenjin, there are various customers who have different requirements. Therefore, we tried to cluster the customers based on their requirements by using the proposed method.

First, in order to get data about customer requirements for Tenjin, we conducted a survey by giving out questionnaire to actual customers. In the questionnaire, the customers described free descriptions about their requirements. For instance, the customers described their requirements for Tenjin by answering the following questions: "what is an appeal of Tenjin area?" "what should be improved in Tenjin area?" and so on. In addition, we asked customers about their sense of values (e.g.. "what is important for you and why?" "what media you usually use and why you use this?"). We

got those data form web-questionnaire survey. The number of answers was 1122. The total number of words was 360,343, and the average number of words in a document was 321 (a standard deviation was 315).

From the analysis by using the proposed method, 25 topics were found out from the documents obtained in the questionnaire. Table 1 shows a part of result in Step2 that lists the representative five words of five topics (translated to English).

For example, Table 1 shows that the descriptions in Topic2 included the words, e.g., "eat," "restaurant," and "delicious," more frequently than the other topics. From these words, Topic2 could be explained as a topic about "gourmet." The meanings of each topic are used to characterize clusters in this study. Figure 2 shows a part of the result of clustering. In this case, the proposed method classified the documents (i.e., the customers) to 13 clusters by k-means clustering [11]. Figure 2 shows the average topic rate for each cluster.

For example, Cluster11 had approximately 20 % of Topic2 "gourmet," and this rate is larger than other clusters. Thus, this result shows that the customers in cluster11 are interested in "gourmet" and/or have requirements related to "gourmet." In this application, the proposed method could characterize each cluster based on the topic rate of each cluster and the meaning of each topic.

Table 1 The representative seven words of five topics

Topic0	Topic1	Topic2	Topic3	Topic4
Baseball	Parking	Eat	Shop	Bus
Sport	Place	Restaurant	Crowd	Get on
Relay	Car	Delicious	Friend	Traffic
Soccer	Fee	Ramen	Associate	Subway
Spectate	Go	Much	Tenzin	Convenience
Fan	Traffic jam	Cook	Shopping	Hakata
Pro	High	Stall	Go	Station

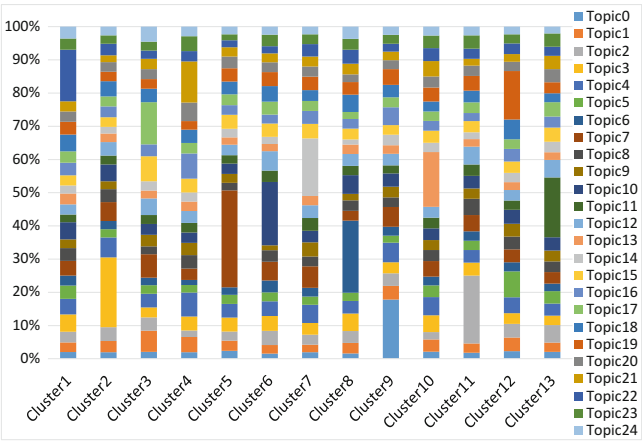

Fig. 2 Topic rates of each cluster

4.2 Cluster Validation

For validating the result of the application, we carried out depth interview to some actual customers who answered the questionnaire. In this result, we focused attention to list of value (LOV) [12] of the customers. LOV represents the importance of interpersonal relations, as well as personal factors, and a personal factor in value fulfillment. We assumed that customers who have similar LOV have similar requirements. Thus, in this validation, LOVs of each customer (#1~#12 in Table 2) were compared.

First, we selected 4 clusters from estimated 13 clusters: Cluster2, Cluster7, Cluster8, and Cluster13. Then, we got interviews for three customers from each four clusters. The selected customers had wide distribution of statistical barometers as shown in Table 2.

Table 2 also shows LOV of each customer, which is selected by each customer in the interview. As shown in Table 2, the customers in the same cluster had similar LOV regardless of the statistical barometers. For example, the customers in Cluster2 (i.e., customers #1, #2, and #3) have similar LOV (i.e., self-respect, fun and enjoyment in life, security, and warm relationships with others). In these results, the proposed method possibly clustered customers who emphasize similar LOV.

However, in Cluster13 shown in Table 2, #11's LOV is different to the LOV of other customers (especially of #12) in Cluster13. Customer #11 has the following LOV: sense of belongingness, fun and enjoyment in life, and security. However, #12 has the following LOV: being well respected, sense of fulfillment, and sense of accomplishment. From the depth interview, it could be understood that the customers in Cluster13 want to develop their ability. The ultimate goal of #11 was different from the ultimate goal of #10 and #12. Customers #10 and #12 want to get the accomplishment by developing their ability. On the other hand, #11 wants to contribute to belonging society by developing their ability. In Cluster13, the customers had common values similar to other clusters. However, there are some cases that the common values cannot be described on LOV [12].

4.3 Discussion

In Sect. 4.1, the proposed method clustered customers from the free descriptions for customer requirement. The free description is more easily to obtain than the operating data. Therefore, this method can be used in the case of such as public service design.

Table 2 LOV of each customer

Characteristics		Cluster2			Cluster7			Cluster8			Cluster13		
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Statistical barometers	Age	35	63	29	46	62	64	36	50	46	23	35	29
	Sex	F	F	M	M	F	F	F	M	F	F	M	M
	Job	Part-time job	Homemakers	Un employed	Office worker	Part-time job	Homemaker	Part-time job	Office worker	Homemaker	Un employed	Office worker	Office worker
LOV													
Sense of belongingness												✓	
Self-respect		✓	✓	✓	✓			✓	✓				
Fun and enjoyment in life			✓	✓	✓		✓					✓	
Excitement						✓							
Security		✓		✓		✓	✓	✓		✓	✓	✓	
Being well respected											✓		✓
Sense of fulfillment									✓	✓			✓
Sense of accomplishment											✓		✓
Warm relationships with others		✓	✓		✓	✓	✓	✓	✓	✓			

In Sect. 4.2, the customers in the same clusters had similar LOV and requirement regardless of the statistical barometers. Therefore, this proposed method is expected to cluster customers based on customer requirements without statistical barometers.

However, this application is insufficient to verify that the proposed method clustered all customers based on their requirements, because of the number of customers who are targeted in the depth interview. Therefore, we need to analyze another data. In addition, in order to cluster customers based on their abstracted requirements, we need to clarify what abstracted level of customer requirements is.

5 Conclusion

To cluster customers based on customer requirements for public service design, this paper proposed a method for clustering and characterizing customers without statistical barometers. To achieve this, the method applies LDA for calculating the requirement similarity. To demonstrate the advantage of the proposed method, an application to an urban development case was conducted. The result of this application showed that the proposed method was expected to support to construct customer models which reflect requirements of each customer. Future works should include clarifying the suitable abstracted level of customer clustering for public service design and developing a method to utilize a classification result for public service design.

Acknowledgment This research is supported by JSPS KAKENHI Grant Number 26280114, Research Institute of Science and Technology for Society (RISTEX) and We Love Tenjin Council.

References

1. A. Cooper, 1999, "The Inmates are Running the Asylum", Sams.
2. Norman C. Dalkey and Olaf Helmer, 1963, "An Experimental Application of the Delphi Method, to the Use of Experts, Management Science", Vol. 9, No. 3, April
3. Paul E. Green and V. Srinivasan,1978, "Conjoint Analysis in consumer research: Issues and outlook", journal of Consumer Research, vol 5, September 1978, pp 103–123
4. S. Watanabe, N. Sashida, A. Nakamura, T. Ugai, K. Ishigaki, 2011, "Atrial application study of persona method to the cooperative social design by the people living in a region: in case of Suzaka City of Nagano Prefecture", The Japan of the Regional Science Association International (JSRSAI), CD-ROM, 48th, in Japanese
5. Y. Motomura, 2011, "User modeling in Service Engineering", The Institute of Electronics, Information and Communication Engineers, Vol. 94, No. 9, pp. 783–787, in Japanese
6. J. Pearl, 1985, "Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning", the 7th Conference of Cognitive Science Society, 329–334
7. H. Fukimoto, M. Etoh, A. Kinno, Y. Akinaga, 2011, "Topic Analysis of web User Behavior Using LDA Model on Proxy Logs", Advances in Knowledge Discovery and Data Mining Lecture Notes in Computer Science Volume 6634, pp 525–536
8. S. Deerwester, S. Dumais, Y. Landauer, G. Furnas, and R. Harshuman, 1990, "Indexing by latent sematic analysis", Journal of the American Society of Information Science, vol. 41(6), pp. 391-407
9. T. Hofmann, 1999, "Probabilistic latent semantic indexing", SIGIR '99 Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 50-57
10. D. M. Blei, Andrew Ng, and Michael Jordan, 2003, "Latent Dirichlet Allocation", Journal of Machine Learning Research, 3-993-1022
11. J. MacQueen, 1967, "Some Methods for Classification and Analysis of Multivariate Observations", Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1, University of California Press, pp. 291–297
12. L. R. Kahle, 1983, "Social Values and Social Change: Adaptation to Life in America", Praeger