

User Segmentation in e-Government Services

Leo Iaquinta¹, M. Alessandra Torsello¹, Marco Comerio¹,
Anna Maria Fanelli², and Giovanni Semeraro²

¹ University of Milano-Bicocca - Dep. of Informatics, Systems and Communication
Viale Sarca 336, 20126 Milano (Italy)
{iaquinta, torsello, comerio}@disco.unimib.it

² University of Bari - Dep. of Computer Science
Via Orabona 4, 70125 Bari (Italy)
{fanelli, semeraro}@di.uniba.it

Abstract. E-Government is becoming more attentive towards providing intelligent personalized services to citizens so that they can receive better services with less time and effort. This work presents an approach for inferring user segments that could be properly exploited to offer personalized services that better satisfy user needs and their expectations. User segments are derived starting from data that essentially describe demographic characteristics of users and that are gathered by questionnaires. A clustering process is performed on gathered data in order to derive user segments, i.e. groups of users sharing similar characteristics. Finally, for each derived segment, we define a user profile that summarizes characteristics shared by users belonging to the same segment. The suitability of the proposed approach is shown by providing results obtained on a case study.

Keywords: clustering, e-Government, user segmentation.

1 Introduction

E-Government (e-Gov) concerns the use of innovative systems, information and communication technologies to provide advanced and efficient services to users (citizens and businesses). The acceptance of these powerful tools in this domain has led to a variety of benefits including reduction of costs, revenue growth, transparency and accountability to governments, greater convenience, and increased productivity. Moreover, e-Gov services have a great potential for delivering better governmental services to users, improving the quality of the provided services and the accessibility to information/services.

One key factor that can help to increase the success of e-Gov is represented by the possibility to provide personalized services that are able to meet the actual needs and demands of users. Hence, in e-Gov domain, a crucial activity consists in acquiring extensive knowledge about target users of public services. Research interest is focusing on the development of strategies aimed to endow governments with personalization mechanisms that enable to conduct their communications

and services in a more user-centric way [1, 5]. One of the most employed strategies for collecting and analyzing knowledge about users is segmentation that, in general terms, refers to the process of identifying groups of users (i.e., segments) which share specific needs, characteristics or behaviour. The aim of user segmentation is to gain extensive knowledge about users in order to satisfy their needs, and to build customised relationships with them [6]. Segmentation approaches can be broadly distinguished into two main types, namely user-based segmentation and product/service-based segmentation. Based on the first approach, segments are derived by considering user characteristics involving demographic, geographic and psychographic/lifestyle variables. In the second approach, segments are derived by considering specific features of the products/services that have to be delivered such as their properties, expected benefits and usage motives.

This work focuses on user segmentation in the e-Gov service domain and it presents preliminary results of an experience carried out within the Italian research project SMART (Services & Meta-services for smART e-Government). SMART is aimed to define, develop and validate innovative methodologies, technologies and tools for the design and the delivery of services in the e-Gov context. Specifically, in this paper, we present a user-based approach for deriving segments from data gathered by questionnaires. Firstly, a number of variables related to user characteristics are identified as segmentation basis, i.e. variables used in the effective process of user segment derivation. Then, data are gathered by requiring users to fill a questionnaire prepared to acquire information about users of a set of services. In the next step, a clustering process is performed on the collected data in order to derive a number of groups of users with similar characteristics. Finally, each derived segment is described by a profile that synthesizes characteristics of users belonging to each segment.

The paper is articulated as follows. Section 2 describes our approach to user segmentation. In section 3, we present results obtained by applying the approach to a case study. In section 4 some conclusions and future works are drawn.

2 The User Segmentation Approach

Starting from data about some user characteristics, the aim of the approach is to derive a number of user segments, i.e. groups including users that are similar with respect to the considered characteristics. To achieve this aim, four main steps are performed namely Segmentation Variable Definition, User Data Gathering, User Segment Derivation and Segment Profile Definition (see Fig. 1). In the following, we provide details for each step.

Segmentation Variable Definition. The first step is devoted to select features that allow to effectively characterize users. The main goal of this step is to identify a set of variables related to user characteristics that can be employed as a basis for segmentation, i.e. variables used for grouping together similar users. A wide number of variables could be selected as segmentation bases. Among these, the most employed variables concern different user features related to

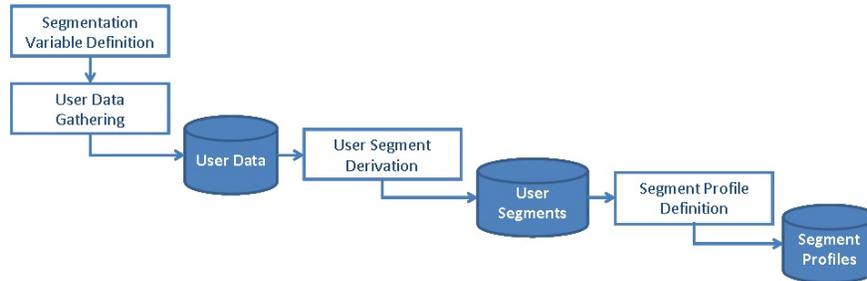


Fig. 1. The user segmentation approach

geographic, demographic, psychological, psychographic and behavioral characteristics [4]. Broadly speaking, some approaches segment users exploiting only one kind of variables, often simple to obtain, such as in geographic segmentation or demographic segmentation. On the other hand, more sophisticated approaches combine different kinds of variables in order to obtain a better user characterization allowing the derivation of more significant segments. For instance, in life-style segmentation, demographic and psychological variables are combined together to characterize users. In this work, demographic and geographic variables are selected as a basis for segmentation. Precisely, selected variables are age, gender, marital status, family, children number, occupation, and residence city features (size, chief town, altimetric zone, tourist vocation). This choice is essentially due to the fact that in the considered domain users of provided services can be properly differentiated with respect to demographic and geographic characteristics.

User Data Gathering. This step has the aim to acquire data about users according to the variables selected in the previous step. User data can be gathered through several methods that may be distinguished into two main classes: methods that require a strong involvement of users and methods that acquire data without asking for the explicit intervention by users [4]. Typically, methods of the first class collect user data by means of questionnaires, telephone interviews, and focus groups. Conversely, the second class methods implicitly gather data from different sources that collect information exchanged by users when they interact to share opinions, perceptions, etc. For example, behavioral and psychographic data can be derived through the application of mining techniques on user generated content shared in blogs, social networks, etc. Other important data sources are informative systems that, collecting information on user interactions during their running, allow to mine data about users and their behavior. In this work, data about users are gathered by means of a questionnaire devised to collect not only demographic data of users but also perceptions about their usage experiences of a set of services on some aspects, such as the time spent to request the service, the time passed to obtain the service, the cost, the service ability to satisfy user needs, the service quality, etc. In addition, starting from the residence of users gathered by questionnaires, geographic data related to residence

city features are obtained by exploiting data of national statistics. To segment users in this work, only demographic and geographic data are considered. Data gathering is performed in some cases by requiring users to autonomously fill online questionnaires and in other cases by interviews (questionnaires are filled by the interviewer).

User Segment Derivation. Once user data are gathered, the next step consists in deriving user segments by grouping together similar users on the basis of segmentation variables. Among the wide variety of segmentation techniques, clustering remains the most popular and widely applied technique. Clustering techniques attempt to identify groups, called *clusters*, of similar users by maximizing the homogeneity within each group and the differences between different groups through the application of suitable distance measures able to evaluate the (dis)similarities existing between each pair of users. The process of clustering can be accomplished in different ways: users may be grouped to derive a partition in which each group represents a cluster (*partitional clustering*) or users may be grouped to form a tree structure in which the intermediate nodes represent clusters of possible partitions (*hierarchical clustering*) [3]. In this work, user segments are derived by a partitional clustering technique, namely the k -means algorithm [3]. Such algorithm is a simple but powerful iterative technique that implicitly partitions the available user data by minimizing the distances between each user into a set of k centroids summarizing user data belonging to each cluster in the partition. The iterative process ends when distance values stabilize or a prefixed maximum number of iterations is reached. As a result, k -means derives k cluster centroids represented as n -dimensional vectors where n represents the number of the considered segmentation variables. A user is matched with all centroids and he is associated to the segment corresponding to the nearest centroid.

Segment Profile Definition. The last step of our approach is devoted to describe user segments derived in the previous step. The aim of this step is to provide extensive knowledge about users belonging to each segment. Such knowledge can be useful in order to select the target segments and to plan the most appropriate strategies that effectively address needs of the targeted users. Specifically, in this step, for each user segment is defined a profile that summarizes the relevant aspects of users within the segment. Typically, information included in a segment profile comprise an identifier that permits to unambiguously refer to the segment, a textual description that synthesizes the most important characteristics of users within the segment and a collection of properties on the set of variables selected in the first step for segmenting users. To derive profiles, the components of each segment centroid are used to describe the characteristics of users belonging to the segment in terms of actual values taken by the corresponding variables.

3 A Case Study

The proposed approach was applied to a case study in order to show its suitability. In particular, the approach was applied in order to derive and characterize

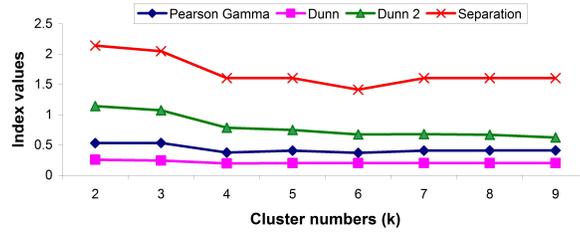


Fig. 2. The obtained index values

segments of entrepreneurs interested in services that are useful to open public businesses in Italy such as café and Bed & Breakfast (B&B).

To gather user data, about 200 entrepreneurs of café and about 250 entrepreneurs of B&B from different Italian regions were contacted and, among these, 38 entrepreneurs of café and 64 entrepreneurs of B&B have accomplished the questionnaire. Before segmenting entrepreneurs, we mapped values of nominal variables of gathered data into numerical values to represent data in a more suitable form for clustering. Successively, k -means was applied on the available data to segment entrepreneurs. We performed several runs with different k ($k = 2, \dots, 9$) and, to establish the proper number of clusters (segments) for the gathered entrepreneur data, we employed some validity indexes (Pearson's gamma, separation and Dunn indexes [2]) that are usually used in literature to this aim. Fig. 2 reports the index values obtained in correspondence of k values. As can be observed, index trends show sharper variations in $k = 3$, thus we selected such value as appropriate number of segments to partition entrepreneurs.

Table 1 shows the centroids obtained by k -means related to the three segments derived from gathered entrepreneur data. Each row contains the component values of a centroid for the corresponding segmentation variables indicated in the column headers. In addition, the last column reports the coverage values on the total number of entrepreneurs of each derived segment.

Finally, each derived segment was described by a profile. A possible interpretation of the defined profiles is as follows:

- S1 Males aged 40-50 years, self-employed. Conjugated with an average family with 4 members including 3 dependents. They live in large municipalities that are not of tourist interest located in coastal hilly area.

Table 1. The three centroids obtained by k -means

	Gender	Age	Marital status	Family	Children number	Occupation	Residence city: chief town	Residence city: allometric zone	Residence city: littoral	Residence city: mountain	Residence city: tourist vocation	Residence city: size	Coverage
S1	1.71	3.57	2.00	4.14	2.57	3.86	1.00	2.14	0.71	0.14	0.00	5.86	19%
S2	1.40	3.00	1.80	3.70	2.00	2.70	0.00	2.30	0.10	0.20	0.10	2.80	28%
S3	1.63	3.37	1.79	2.74	0.21	3.58	0.68	2.84	0.74	0.21	0.16	5.53	53%

- S2** Mostly women aged 36-45 years, involved in employee jobs. Married with an average family of 4 people with 2 dependents. They live in small-medium sized cities that are not of tourist interest located in internal hilly area.
- S3** Mostly male of about 40 years, professionals. Living with an average family of 2 people and no dependents. They live primarily in medium-large sized chief-towns located in plains or coastal areas of some tourist interest.

The knowledge mined from demographic and geographic data about users, embedded in the defined segment profiles, could be conveniently used to adapt services according to the user characteristics and to provide personalized services able to meet their actual needs.

4 Conclusions and Future Works

This work has presented an approach for user segmentation in e-Gov domain based on a partitional clustering technique exploiting demographic and geographic variables. The suitability of the approach was shown by considering data about entrepreneurs interested in opening public businesses. The obtained preliminary results encourage the application of the approach to wider scenarios involving e-Gov services. As future work, the integration of other segmentation variables that describe behaviours and attitudes of users could be fruitful to infer segments able to better capture needs and expectations of users in order to predict their future behaviours.

Acknowledgment

The work presented in this paper has been partially supported by the Italian PON project PON01 00861 SMART (Services and Meta-services for smART eGovernment). The authors would like to thank Ancitel S.p.A. and Halley Sud-Est, industrial partners of SMART, for the provision activity of questionnaires.

References

1. X. Guo and J. Lu. Intelligent e-government services with personalized recommendation techniques. *Int. J. Intell. Syst.*, 22(5):401–417, 2007.
2. M. Halkidi, Y. Batistakis, and M. Vazirgiannis. On clustering validation techniques. *J. Intell. Inf. Syst.*, 17(2-3):107–145, 2001.
3. A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323, 1999.
4. C.W. Lamb, J.F. Hair, and C. McDaniel. *Essentials of Marketing*. South-Western College Publ., 2011.
5. J. Lu, Q. Shambour, Y. Xu, Q. Lin, and G. Zhang. BizSeeker: A hybrid semantic recommendation system for personalized government-to-business e-services. *Internet Research*, 20(3):342–365, 2010.
6. M. Wedel and W.A. Kamakura. *Market Segmentation: Conceptual and Methodological Foundations*. Kluwer Acad. Publ., 2000.