Applying text and data mining techniques to forecasting the trend of petitions filed to e-People

Jong Hwan Suh [a,*], Chung Hoon Park [b], Si Hyun Jeon [c]

[a] Information and Communication Technology Research and Development Center, SAMSUNG SDS, 159-1 Samseong-dong, Gangnam-gu, Seoul 135-798, Republic of Korea
[b] Consulting Division, SAMSUNG SDS, 707-19 Yeoksam-dong, Gangnam-gu, Seoul 135-918, Republic of Korea
[c] Anti-corruption and Civil Rights Commission (ACRC), Uijuro 81, Seodaemun-gu, Seoul 125-705, Republic of Korea

ARTICLE INFO

ABSTRACT

As the Internet has been the virtual place where citizens are united and their opinions are promptly shifted into the action, two way communications between the government sector and the citizen have been more important among activities of e-Government. Hence, Anti-corruption and Civil Rights Commission (ACRC) in the Republic of Korea has constructed the online petition portal system named e-People. In addition, the nation's Open Innovation through e-People has gained increasing attention. That is because e-People can be applied for the virtual space where citizens participate in improving the national law and policy by simply filing petitions to e-People as the voice of the nation. However, currently there are problems and challenging issues to be solved until e-People can function as the virtual space for the nation's Open Innovation based on petitions collected from citizens. First, there is no objective and systematic method for analyzing a large number of petitions filed to e-People without a lot of manual works of petition inspectors. Second, e-People is required to forecast the trend of petitions filed to e-People more accurately and quickly than petition inspectors for making a better decision on the national law and policy strategy. Therefore, in this paper, we propose the framework of applying text and data mining techniques not only to analyze a large number of petitions filed to e-People but also to predict the trend of petitions. In detail, we apply text mining techniques to unstructured data of petitions to elicit keywords from petitions and identify groups of petitions with the elicited keywords. Moreover, we apply data mining techniques to structured data of the identified petition groups on purpose to forecast the trend of petitions. Our approach based on applying text and data mining techniques decreases time-consuming manual works on reading and classifying a large number of petitions, and contributes to increasing accuracy in evaluating the trend of petitions. Eventually, it helps petition inspectors to give more attention on detecting and tracking important groups of petitions that possibly grow as nationwide problems. Further, the petitions ordered by their petition groups' trend values can be used as the baseline for making a better decision on the national law and policy strategy.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Owing to the development of information and communication technology (ICT), particularly the Internet, the government sectors around the world have tried to progress themselves into the electronic government (e-Government), a.k.a. digital government. The construction of e-Government aims at providing citizens with services quickly and accurately, effectiveness of government work, innovation by redesign work process, and raising national competitiveness by improving productivity (Lee & Jung, 2004). Therefore, the anticipated benefits of e-Government can be more efficiency, greater convenience, improved services, better accessibility of public services, less corruption, more transparency, revenue growth, and cost reductions (Atkinson & Castro, 2008). According to Palvia and Sharma (2007), various activities of e-Government can be summarized into four categories with respect to interaction domains. The first type is to push information over the Internet, e.g. regulatory services, issue briefs, notifications, etc. Secondly, some models aim at improving two way communications between the government agency and the citizen, a business, or another government agency. In the second type of models, users can engage in dialogue with government agencies and post problems, comments, or requests to the government agencies. Third, e-Government helps conduct transactions such as lodging tax returns and applying for services and grants. The fourth type is based on governance, e.g. online polling, voting, and campaigning.

Among these four types in activities of e-Government, the second one has recently been more important, especially in the Republic of Korea. This is because the Internet has started to be used as the virtual space where citizens are united and their opinions are promptly shifted into the action such as a mass rally. Therefore, Anti-corruption and Civil Rights Commission (ACRC), one of the government sectors in the Republic of Korea, has decided to strengthen its activities of collecting public opinions by building up the online petition portal system. At last, in June 2006, ACRC constructed the online petition portal system called e-People[1] on purpose to hear the voice of the nation by merging scattered online channels which had collected civil complaints and petitions. As a result, ACRC could achieve fame by developing e-People with its winning the Best Demonstration Stand Award at the e-Challenge Conference and Exhibition 2008 held in Stockholm, Sweden (ACRC, 2008).

Nevertheless, we expect that e-People will be faced with another challenging need as Open Innovation gains increasing attention as a new paradigm for innovation in the company's business. Shortly, the concept behind Open Innovation is that companies cannot rely entirely on their own research but they should instead buy or license processes or inventions from other companies (Chesbrough, 2003). Similarly, the government sectors will also need to consider putting the concept of Open Innovation into action in the way like the innovation of the national law and policy not only from the inside, i.e. themselves, but also from the outside, i.e. citizens. Thus, we predict that the concept of Open Innovation is going to affect on the activities of e-Government sooner or later. As a result, we reached an agreement that we should evolve e-People into the virtual space where boundaries between citizens and government sectors are melted down so that citizens can participate in innovating on the government system conveniently by filing petitions to e-People. However, currently there are problems and challenging issues until e-People realize the concept of Open Innovation in the government sector with raising e-Government of the Republic of Korea to the next level. And they can be summarized into two matters as follows.

First, there's no objective and systematic method to analyze a large number of petitions filed to e-People without a lot of manual works of petition inspectors. As you may expect, petitions in e-People are collected from citizens all around the places in the Republic of Korea. Therefore, the number of petitions to be read is beyond the man's ability. Besides, more than half data that a petition contains is text-based, and thereby it is difficult to understand petitions at a glance by manual works of petition inspectors. This makes it hard to conceive the voice of the nation on the basis of petitions filed to e-People. Therefore, we need to take advantage of text mining techniques. If keywords are elicited from text in petitions and petitions are clustered into petition groups with the elicited keywords, petition inspectors will be able to focus on analyzing the trend of petitions and consequently conceiving the voice of the nation with their manual works being reduced much in reading a great number of petitions. However, the result from our literature survey showed there are few researches done about applying text mining techniques to petitions to overcome these problems and we need to perform the related research.

As the second matter, it is required to forecast the trend of petitions more accurately and quickly by using e-People rather than petition inspectors for the better national law and policy strategy. e-People currently perceives the importance of petitions after they get serious actually because they are evaluated by manual analyses of petition inspectors. So it takes a lot of time to find important petitions that might grow as nationwide problems while delaying

planning and practicing their related national law and policy strategy. However, if the prediction models are built up by applying data mining techniques to petitions, we are going to be able to predict the trend of petitions more accurately and quickly by using e-People rather than petition inspectors. In the end, the predicted trend of petitions will contribute to making it possible for the government sectors to make a better decision on the national law and policy strategy. Therefore, to applying data mining techniques for forecasting the trend of petitions can be the challenging issue to be solved by us.

Hence, we propose the framework of applying text and data mining techniques to petitions filed to e-People to solve those problems and challenging issues that are stated previously. In other words, we apply text mining techniques to unstructured data to elicit keywords from petitions and identify groups of petitions with the elicited keywords. Moreover, we apply data mining techniques to structured data of the identified petition groups to forecast the trend of petitions. To sum up contributions of our applying to text and data mining techniques to petitions filed to e-People, we provide an objective and systematic method for analyzing a large number of petitions filed to e-People and predicting the trend of petitions with manual works of petition inspectors being reduced. And we consequently help the government sectors to make a better decision on the national law and policy strategy on the basis of the trend of petitions forecasted by our approach.

The rest of the paper is structured as follows. In Section 2, we introduce the taxonomy of e-Government, and we take a look at related works on keyword extracting and document clustering in text mining, and forecasting models with data mining techniques. In Section 3, we explain the framework of our methodology through three subsections: eliciting keywords from petitions; identifying petition groups; forecasting the trend of petition. In Section 4, we apply the methodology suggested in Section 3 to 8 groups of petitions filed to e-People, i.e. the online petition portal system constructed by ACRC of the Republic of Korea. In detail, we perform the 8 fold validation on the prediction models based on RBFNs and C5.0 after dividing the 8 petition groups of petitions repeatedly 8 times into 7 petition groups for training sets and the rest petition group for the test set. And we discuss the implication of results in the performance validation. In Section 5, finally we conclude the paper with discussion of contributions and further researches.

2. Literature reviews

2.1. e-Government and its cases in the Republic of Korea

e-Government refers to the delivery of national or local government information and services via the Internet or other digital means to citizens, businesses or other governmental agencies (UN & ASPA, 2002; Evans & Yen, 2006). And its cases can be categorized into three groups according to Marchionini, Sanan, and Brabdt (2003).

First, government to citizen (G2C) service of e-Government provides one-stop and on-line access to information and services to citizens. And G2C applications enable citizens to ask questions and receive answers about filing income taxes, paying taxes, renewing driver licenses, petition, and so on. In the case of the Republic of Korea, the online civil service portal site named as government for citizen (G4C[2]) enables integrated one-site delivery of a wide array of information provided by various administrative agencies. Through the G4C portal site, citizens are able to obtain support related to over 5,000 types of public services. In particular, G4C does

---

[1] www.epeople.go.kr.

[2] www.egov.go.kr.