Deciphering the Public's Stance towards Governmental Decisions

George Stylios; Dimitris Christodoulakis; Jeries Besharat; Maria-Alexandra Vonitsanou; Ioannis Kotrotsos; Athanasia Koumpouri;, Sofia Stamou

Computer Engineering and Informatics Department
Patras University, 26500 GREECE

gstylios@yahoo.gr; dxri@upatras.gr; besarat@ceid.upatras.gr ; bonitsan@ceid.upatras.gr; kotrotso@ceid.upatras.gr; koumpour@ceid.upatras.gr; stamou@ceid.upatras.gr

Abstract. The evolution of the Internet and the proliferation of the computer-literate population have enabled people publish online content through which they express their personal views on a variety of issues. As a result, the web is not any longer perceived as a mere information sharing environment but has also turned into a social network where users articulate their opinions on different topics. Nowadays, topics related to governmental decisions are among the most widely discussed ones within both physical and digital societies. This is not only because web 2.0 has empowered people with the ability to communicate remotely but also because governments all around the globe publish a great volume of their decisions and regulations online. In this paper, we examine several features in the user-generated content discussing governmental decisions in an attempt to automatically extract the citizen opinions from online posts dealing with public sector regulations and thereafter be able to organize the extracted opinions into polarized clusters. Our goal is to be able to automatically identify the public's stance (i.e. positive/negative) against governmental decisions and thus be able to infer how the citizens' viewpoints may affect subsequent government actions. To carry out our study objective, we run a preliminary experiment against a small set of real citizen opinions referring to government decisions on the reform of the Greek educational system. The application of our opinion extraction and classification technique against our dataset reveals that user-generated content entails clearly polarized opinions with respect to governmental decisions.

1. Introduction

A critical factor of success in today's governments is to effectively communicate their messages to citizens and build strong alliances with them by empowering their participation in the decision-making process. With the advent of the web and the technological advancements, governments around the globe have launched ambitious plans for building electronic government (e-government) applications and services. The quest in establishing e-government applications is on the one hand to facilitate public sector regulations reach society in an instant and cost-effective manner and on the other to get around any obstacles imposed by bureaucracy within the governments' interactions with stakeholders (i.e., citizens, public and private bodies).

Despite the resources that have been allocated in realizing e-government, still is deployment is confined to the technological aspects of implementing public services and fails to capitalize on the societal factors involved in the determination of information policies via technological means. In other words, in its current form e-government is mainly perceived as a service system to support the activities of governments and lets aside issues dealing with the socio-political impact of those activities.

In this paper, we try to fill this void by proposing a novel e-government mechanism that captures the societal impact of public sector regulations in an attempt to decipher the public's stance towards governmental decisions. In particular, we introduce a model that harvests user-generated content from government-related web 2.0 sites and applies lexico-semantic

processing to the collected data in order to firstly identify the citizen opinions latent in their posts and then cluster identified opinions in terms of their referential governmental decisions and sentiment orientations. The novelty of our model is that it goes beyond processing user content relevant to governmental issues and addresses ways of clustering and evaluating user opinions.

To demonstrate the functionality of our proposed mechanism, we present as a proof of concept an experiment we carried out with real data pertaining to the recent governmental decisions on the reform of the Greek educational system. Specifically, we describe how we mined and automatically clustered citizen opinions published online and discussing the Greek educational reform in order to derive the societal impact of the reform regulations across citizens. Moreover, we investigate how public opinions were reflected in remedial actions taken by the government while fine-tuning the regulations' details. The objective of our study is to demonstrate that e-government services invoke the citizens' active participation in the decision making process and stress that by putting together inter-disciplinary methods and tools we can transform e-government from a technological infrastructure to a powerful interactive manifestation of e-inclusion and e-participation.

The remainder of the paper is organized as follows. We start our discussion with a brief overview on existing studies addressing the current trends in e-government issues. In Section 3 we introduce our proposed opinion extraction and classification technique. In particular, we discuss the details of our method which relies on the exploitation of the actual text of the user-generated content and applies a number of lexico-semantic text processing steps in order to firstly identify the citizen opinions about public sector regulations and then group mined opinions according to their underlying polarity. In Section 4, we present a preliminary experimental study we carried out with using real web content in which we evaluated the performance of our opinion classification method and we discuss obtained results. We conclude the paper In Section 5.

2. Related Work

E-government refers to the use of information and communication technology in order to improve government services and interactions with citizens (Chan et al., 2003) (West, 2004). Many researchers and business organizations have studied the means via which public administrations and citizens can benefit from the availability of e-government applications and services (cf. UN E-Government Survey, 2008). According to the survey of (Beierle and Cahill, 2000) the electronic democracy in state-level decision making is in an early and experimental phase. Moreover, it has been recently attested that despite the deployment of e-government projects, the latter have failed to satisfy the expectations of both the government and their citizens in developing functional government services (Alimagwashi and Mcintosh, 2009). To overcome this shortcoming, the authors suggest providing citizens with the means to comment on governmental rules online, so that they become an active part of the democratic process. In a latter study (Okot-Uma, 2002) claimed that the objective of e-government is to create the conditions for the empowerment of individuals, communities or the civil society either by creating robust infrastructures for the information society or by enabling participation in the e-society. Moreover, Gant et al. (2004) conducted a comparative evaluation of six e-government services and found that although technological factors have important consequences for e-government development, nevertheless social networks and performance-based incentive systems mediate the effectiveness of e-government. In addition to the above (Usero et al., 2009) examined methodological and technical guidelines to foster the usage of new procedures for information integration in the e-government era. Specifically, they suggest the exploitation of information interoperability technologies, the design of a model to evaluate interoperability in public websites and the empirical analysis of key factors in information management, in order to provide citizens with better e-government services. On the other hand, (Sabucedo and Rifon, 2006) proposed the use of a semantic platform that integrates different e-government solutions so as to expand accessibility of services in a broad sense. A pilot application of their platform reveals that there are several advantages associated with the application of semantics in e-government tools. Motivated by the above studies and bearing in mind that e-government aims primarily at fostering the participation of citizens in the decision making process, in our current work we address e-government from the perceptive of the end users. Specifically, we introduce a novel framework that not only enables govern-

mental bodies inform citizens about their decisions and directives but it also assists citizens actively participate in the decision making process in a seamless yet powerful manner. The details of our proposed framework are discussed in the following section. Finally, we should note that we deem our study to be complementary to existing works that examine the amelioration of e-government services.

3. Extracting Citizen Opinions from Web 2.0 Content

The objective of our work is to design a method that can automatically capture the polarity of the online user-generated content that is published to verbalize citizen opinions about governmental decisions. In this direction, we rely on real web content harvested from e-government social media sites, e.g., blogs, fora, etc., and we analyze the associated user posts in order to firstly identify and extract the citizen opinions from the posts' textual data, then mine the opinions' polarity and eventually train a classifier to automatically organize mined opinions according to their sentiment orientation into positive and negative ones. Unlike traditional approaches that attempt to infer the users' stance against the issues they discuss online by examining their ratings (usually denoted on a numeric/star scale), our approach focuses on the actual text of the user posts and tries to mine the citizens' stance against particular aspects adhering to governmental decisions. This is because we believe that numerical ratings do not convey much information about which decision aspects citizens value positively or negatively and as such they cannot be fruitfully utilized in subsequent governmental regulations. Conversely, the textual content of user posts, if properly processed and analyzed, can be much more revealing about the impact that governmental decisions have on citizens' perception.

To understand better how public sector regulations affect the citizens' stance against governmental actions, we conduct a two-level analysis. Firstly, we rely on user online comments, which we process as describe below in order to identify the latent user opinions and then be able to extract such opinions, which we call opinion phrases. Then, at the second level of analysis we annotate the sentiment orientation of the identified opinion phrases in order to assign them with a positive or negative polarity label depending on their publishers' underlying stance against the issues they discuss. Based on the output of the above analysis, we may not only capture the citizen's viewpoints on governmental issues but with the proper tools and techniques we may also be able to build predictive models about how citizens value public sector regulations. In the following paragraphs, we describe in detail how we process the user postings to identify and evaluate opinion phrases as well as how to utilize the mined opinions along with their polarity labels in order to train an opinion classifier.

3.1. Opinion Extraction and Sentiment Analysis from User Postings

To identify citizen opinions within their postings pertaining to governmental issues, we firstly download the content of the user posts and we process it in order to identify within their textual data the text extracts that contain user opinions. To process the data, we apply tokenization to the posting's body in order to extract the lexical elements of the user generated texts and then we apply Part-of-Speech tagging to the identified elements in order to obtain their grammatical categories. Thereafter, we rely on postings containing adjectives and we further analyze them to derive opinion phrases from their body. The reason for concentrating on postings containing adjectives is because according to the study of Gliozzo et al. (2004) people evaluate an item or verbalize an opinion via adjectives. Therefore, user posts containing adjectives are highly probable of indicating implicit user opinions as opposed to posts that contain no adjectives at all. But, relying entirely on the study of adjectives is not sufficient for characterizing the user opinions given that we must firstly identify the topic to which every adjective (i.e., opinion) refers. In other words, we need to detect within a user's post the topic(s) being discussed as well as the user's opinion on those topic(s). Only then we will be able to successfully extract the phrases that communicate user opinions within their online postings. Unlike current approaches that focus on either the entire body of the posting (Pang et al., 2002) or on full sentences (Bethard, 2004), in our work a user opinion is communicated via the use of adjectives, while the topic to which the opinion refers is communicated via the nouns or proper nouns to which the corresponding adjectives refer. To be able to detect such references within the user posts, we employ a syntactic dependency parser, which given as input a piece of text containing adjectives it identifies the noun (or proper noun) to which

every adjective refers, i.e., characterizes. Then, we rely on the syntactically dependent adjective-noun pairs[1] to derive the opinion phrases communicated in the user postings.

Having extracted opinion phrases from the users' online content, our next step is to examine how users evaluate the topics for which they express their opinions. To enable that, we rely on the notion of the words' semantic orientation and we try to discriminate between words or positive and negative sentiment. For such discrimination, we manually annotate the identified adjectives with a suitable polarity label, i.e., positive (+) or negative (-) depending on whether the adjective signifies approval or disapproval for its referring topic. The criterion under which annotation takes place is that a positive adjective is one that gives praise to the topic under evaluation, whereas a negative adjective is one that criticizes some or all aspects of the topic being discussed. Note that the manual labelling of adjectives bears a satisfactory level of objectivity as previously reported in the studies of Hatzivasiloglou and McKeown (1997) (Turney and Littman, 2003).

The result of the above process is as set of positive and negative opinion phrases that correspond to citizen views on governmental decisions discussed in their posts. Based on the annotated user opinions, we employ a classification module, which we train in order to be able to automatically identify the orientation of user comments as we will describe next.

3.2 Classifying User Opinions in terms of their Polarity

Having extracted user opinions about governmental decisions from the citizens' online postings and having also deduced the polarity orientation of every identified opinion, our next step is to rely on our processed data in order to train a classification module to automatically organize opinion phrases into bimodal clusters of supportive (i.e., positive) and unsupportive (i.e., negative) citizens' views on governmental regulations. In this respect, we employ a decision tree-based classification module and proceed as follows. We rely on the annotated opinion phrases previously extracted from our dataset and we expand them with semantic information harvested from WordNet (Fellbaum, 1998). WordNet is a lexical hierarchy, which organizes concepts into synonym sets and links them together depending on the underlying semantic relations that connect concepts together. In our approach, we explore the adjectives encoded in WordNet and utilize them for expanding our collected opinion phrases. Opinion phrases' expansion follows a two-level expansion: (i) expansion with semantically equivalent adjectives, which we call synonymy expansion and (ii) expansion with antonymous adjectives, which we call antonymy expansion. In the first expansion level, we append to all the adjectives appearing in our collected opinion phrases their synonyms recorded in WordNet, while in the second expansion level we append to all the adjectives appearing in our collected opinion phrases their antonyms recorded in WordNet.

Following opinion phrases' expansion, we proceed with the sentiment annotation of the expanded phrases as follows. Opinion phrases formulated after applying synonymy expansion take the same polarity label with that of their originating opinion phrases, i.e., before expansion, while opinion phrases formulated after applying antonymy expansion take the opposite polarity label from that of their originating phrases. Then, based on the expanded set of polarized opinion phrases, we split them into training and testing examples and use the training set to learn the classifier automatically identify the polarity of opinion phrases. The learning accuracy of the classifier is evaluated against the test set and by running several classification iterations. In the following section, we present the details of a classification experiment we conducted in which we evaluated the classification accuracy of our method when relying on the extracted polarized opinion phrases. Obtained results give useful insights with respect to how citizens comment governmental decisions as well as with respect to the usefulness of opinion phrases into revealing the public's stance against public sector regulations.

4. Experimental Evaluation

To assess our study objective, we collected a set of real citizen opinions. This dataset was downloaded from the forum antheorisi[2], which is a Greek forum focusing on policy issues that

---

[1] Note that there might apply different patterns to the adjective-noun syntactic structure, e.g. noun-adjective, adjective-noun, determiner-adjective-noun, etc.

are being developed by several users, and consists of 124 comments. After processing these comments as previously described we extracted from their body 652 distinct opinion phrases. Then, we trained three different classifiers incorporated into the Weka platform[3]. The three classifiers we use are Support Vector Machine, K-Nearest Neighbor and Naïve Bayes.

Before the training phase, our software prepares the data by randomizing the full dataset and then stratify it because the classification class is nominal. Then, performs a 10-fold cross validation and generates training and test sets. After the training, each classifier returns a summary of the results. The following chart shows the average of cases that were correctly and incorrectly predicted for the three classifiers.
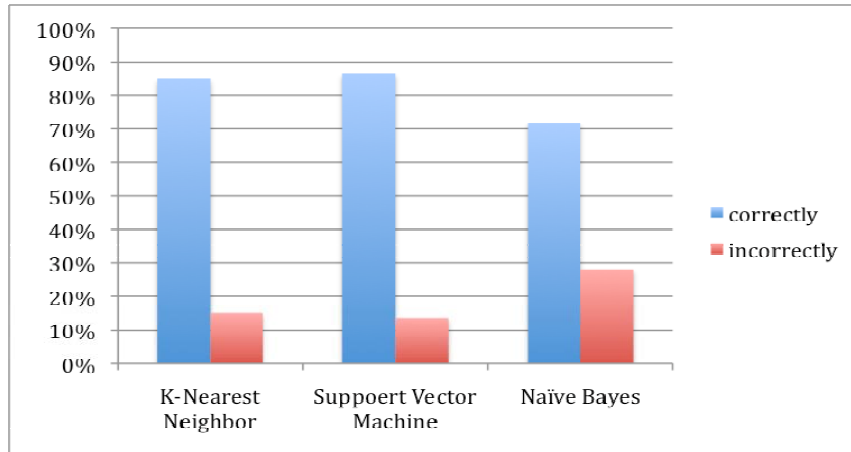
Figure 1. Comparative evaluation of opinion classification accuracy

As the figure shows, the algorithm with the best classification performance is the Support Vector Machine where the average accuracy is about 86% while the worst performing classification algorithm is Naïve Bayes with 72% accuracy. Table 1 summarizes the performance details of the three classification modules we employed in our study and as results suggest the proposed method is quite effective into automatically organized opinion phrases in terms of their polarity.

	K-Nearest Neighbor	Support Vector Machine	Naive Bayes
True Positive Rate	67.58%	67.74%	12.07%
True Negative Rate	96.09%	97.71%	99.37%
False Positive Rate	3.905%	2.28%	0.62%
False Negative Rate	32.41%	32.25%	87.92%

Table 1. Evaluation of opinion classification accuracy

5. Concluding Remarks

In this paper we have presented a method for extracting citizen opinions about governmental decisions from social media sites, as well as a technique for classifying opinion phrases in terms of their sentiment orientation. The application of our proposed method over a set of real user content reveals that properly processed and analyzed opinion phrases can serve as use-

---

[2] Anatheorisi.org

[3] http://www.cs.waikato.ac.nz/ml/weka/

ful indicators for the perception of governmental decisions by the public. Our method relies on the intuition that there is plentiful data available on social web sites that communicates implicit information about how citizens perceive governmental regulations and directives. Being able to collect, process and mine such data can provide decision-makers with valuable information about how the recipients of their actions evaluate the latter and it can also empower citizens with the ability to actively participate in governmental decision making aspects.

References

Alimagwashi, H. and Mcintosh, S. (2009). "Understanding the Government to e-Government Transition using a Soft Systems Approach: What is e-Government Supposed to do?" In Proceedings of the 9th European Conference on e-Government, pp 45-55.

Beierle Th., Cahill S. (2000). "Electronic Democracy and Environmental Governance: a survey of the States". http://www.rff.org/rff/documents/rff-dp-00-42.pdf.

Bethard S., Yu H., Thornton A., Hatzivassiloglou V. and Jurafsky D. (2004). "Automatic Extraction of Opinion Propositions and their Holders". In Proceedings of the AAAI Conference, Springer.

Chan, C.M.L., Pan, S.L. and Tan, C.W. (2003) "Managing Stakeholder Relationships in an e-Government Project". In Proceedings of the 9th AMCIS Conference, pp.783-791

Fellbaum Ch. (Ed.). (1998). WordNet: An electronic lexical database. The MIT Press.

Gant J., Nam W.J., Bretschneidetr S., Kim S., Choi H. (2004). "The Transformation of E-Governence in Local Government: Comparison of Critial Success Factors in Gangnam-Gu, Seoul South Korea and Washington D.D. USA" In Proceedings of the National Conference on Digital Covernance Research.

Gliozzo, A., Strapparava, C. and Dagan, I. (2004) "Unsupervised and Supervised Exploitation of Semantic Domains in Lexical Disambiguation". In Computer Speech and Language, 18(3), pp. 275-299.

Hatzivassiloglou, V. and McKeown, K.R. (1997) "Predicting the Semantic Orientation of Adjectives". In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL. pp. 174-181.

Okot-Uma R. (2002). "Electronic Governance: Re-inventing Good Governance". Available at: http://webworld.unesco.org/publications/it/EGov/wordbank%20okot-uma.pdf

Pang B., Lee L. and Vaithyanathan S. (2002). "Thumps up? Sentiment classification using machine learning techniques". In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-02)

Sabucedo L.A., Rifon L.A. (2006). "Semantic Service Oriented Architectures for eGovernment Platforms". In Proceedings of the AAAI Spring Symposium: the Semantic Web meets eGovernment, Stanford University, CA, USA.

Turney, P.D. and Littman, M.L. (2003) "Measuring Praise and Criticism: Inference of Semantic Orientation from Association". In ACM Transactions on Information Systems 21 (4): 315-346.

UN E-Government Survey: Form E-Government to Connected Governance, 2008. Available at: http://unpan1.un.org/intradoc/groups/public/documents/UN/UNPAN028607.pdf

Usero M., Orenes P.B., Comeche M., Segundo R. (2009). "Model for Interoperability Evaluation in E-Government Services". In Proceedings of the International Conference in Multimedia and ICT, pp. 105-109.

West, D.M. (2004) "E-Government and the Transformation of Service Delivery and Citizen Attitudes". In Public Administration Review, vol.64, no.1, pp. 15-27.