Public Opinion Mining for Governmental Decisions

George Stylios, Dimitris Christodoulakis, Jeries Besharat, Maria-Alexandra Vonitsanou, Ioanis Kotrotsos, Athanasia Koumpouri and Sofia Stamou
Patras University, 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: eGovernment refers to the use of information and communications technologies (ICTs) to improve the quality of services and information offered to citizens, to make government more accountable to citizens and advance public sector transparency. As already pointed out by other researchers, one of the most important issues for making eGovernment effective is to enable citizens to participate in the decision-making process. Nowadays, topics related to governmental decisions are among the most widely discussed ones within 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 propose the exploration of text and data mining techniques towards capturing the public's opinion communicated online and concerning governmental decisions. The objective of our study is twofold and focuses on understanding the citizen opinions about eGovernment issues and on the exploitation of these opinions in subsequent governmental actions. 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 against governmental decisions and thus be able to infer how the citizens' viewpoints may affect subsequent government actions. To demonstrate the usability and added value of our proposed approach we have designed an interactive eGovernment infrastructure, the architecture of which we will present and discuss in our paper. Moreover, we will elaborate on the system details, its adaptation capacity and we will discuss its usage benefits for both citizens and public sector bodies.

Keywords: opinion mining, opinion classification, knowledge extraction, linguistic analysis

1. Introduction

In the present day, the challenge for governments is how to move on from focusing on service delivery to providing people-centered applications. In other words, government's success relies on effectively communicate their messages to citizens and build strong alliances with them by empowering their participation in the decision-making process. The Internet has the potential to radically change the face of government by fostering communication between citizens and public officials. Nowadays, governments around the globe have launched ambitious plans for building electronic government (eGovernment) applications and services. The quest in establishing eGovernment applications is multidimensional; facilitate public sector regulations reach society in an simple, instant and cost-effective manner, increase the accountability of government to its citizens, reduce bureaucracy and corruption within the governments' interactions with stakeholders (i.e., citizens, public and private bodies). Despite the resources that have been allocated in realizing eGovernment, it is mainly perceived as a service system to support the activities of governments and sets aside issues dealing with the socio-political impact of those activities.

In this paper, we try to fill this void by proposing a novel eGovernment 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 propose the exploitation of data mining techniques towards firstly capturing the public's opinions (communicated online) about governmental decisions and secondly analysing the polarity of the mined opinions so that they are considered in subsequent governmental decisions. Specifically, we introduce a method for decomposing citizens' opinions and comments that are posted in online fora and blogs, in order to evaluate how governmental decisions are perceived by the public and thereafter how the public's implicit feedback should be interpreted by governmental bodies in their subsequent actions. What motivates our study is that up-to-date gov-

Reference this paper as: Stylios, G, Christodoulakis, D, Besharat, J, Vonitsanou, M, Kotrotsos, I, Koumpouri, A and Stamou, S. "Public Opinion Mining for Governmental Decisions" *Electronic Journal of e-Government* Volume 8 Issue 2 2010, (pp203-214), available online at www.ejeg.com

ernmental social web sites are not consistently evaluated in the governmental decision-making process and as such citizens' voices are most of the times heard in a limited audience.

To realize our study objective, we propose a framework that integrates text and data mining methods for modelling the public's opinions and evaluations of the governmental decisions. We then, statistically analyse the mined citizens' feedback in order to derive on the one hand the sentiment orientation of the public opinions and on the other the underlying correlation between mined opinions and the formulation of new governmental decisions on related issues. To demonstrate the functionality of our proposed mechanism, we present as a proof of concept an experiment we carried out in which we mined and automatically organized into polarized clusters citizen opinions published online and discussing governmental regulations. 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. The findings of our experimental study clearly demonstrate that eGovernment services invoke the citizens' active participation in the decision making process and indicate that by putting together inter-disciplinary methods and tools we can transform eGovernment from a technological infrastructure to a powerful interactive manifestation of e-inclusion and e-participation.

The remainder of the chapter is organized as follows. We start our discussion with a detailed overview on existing studies addressing the current trends on eGovernment as well as the encapsulation of data mining applications into available eGovernment applications. In Section 3, we introduce the utilization of text and data mining techniques for identifying and deciphering the citizen opinions about governmental issues that are communicated online. In particular, we describe how we can exploit the user-generated online content via the use of natural language processing and text mining tools in order to firstly mine user opinions from their posts and then annotate the mined opinions with a suitable polarity label depending on the orientation of the latent user opinions. In Section 4, we present preliminary experiments we carried out in which we relied on real user comments about governmental decisions and tried to organize the mined user opinions into polarized clusters of citizen comments. We present our experimental results and discuss the implications of our findings. We conclude the chapter in Section 5 where we also sketch our plans for future work.

2. Background

A fully inclusive information society exploiting the recent advances in information and communication technologies (ICTs) is the means of sustainable growth. The key aspect of an inclusive information society is the establishment of fully functional Electronic government (or eGovernment) capabilities. It is expected that EGovernment will be accelerated with a view to face major challenges such as ageing, climate change or terrorism, to assure better services, better security and better democracy, to provide seamless public services across borders and to increase citizens' opportunities for mobility and for business

A few years back eGovernment involved two stages Reddick (2004): The first stage involved the information dissemination phase in which governments catalogued information for public use while the second phase involved transaction-based eGovernment (e-service delivery) such as paying taxes online. Thus, it was regarded as a way for governments to use the most innovative information and communication technologies to offer citizens, business and employees with efficient access to information and services (Hayat 2009). In fact a wide range of services is encompassed by the term eGovernment as dissemination of information commerce with the private sector services to individual citizens and businesses, and participatory democracy (Irani et al, 2005). EGovernment indicates that management services and functions are transferred onto the internet enhancing the efficiency of the public sector and developing more personal, customized relationships between citizens and their government. The reduction of administrative and operational costs, as well as the enhancement of the services they offer to businesses, citizens and the general community at large, has been a driving force for the development and implementation of an eGovernment infrastructure.

During more than ten years governments around the world are implementing or launching plans for developing electronic government projects. Among the pioneers in that area the Australian government began to develop its own eGovernment strategies since the early 1990s. By 2004, over 1,600 Commonwealth, Public and Stakeholder Opinion, and e-Democracy government services are available online (Macintsoh, and Whyte, 2006). In 2005, the European Commission developed a strategic framework, "i2010 – A European Information Society for growth and employment", promoting an open and competitive digital economy and emphasizing ICT as a driver of inclusion and quality of life. The importance of eGovernment development in EU forced the elaboration and the adoption of a specific "i2010 eGovernment Action Plan" as an integral part of the i2010 initiative focusing on five major objectives for eGovernment: 1) no citizen left behind, 2) making efficiency and effectiveness a reality, 3) implementing high-impact key services for citizens and businesses, 4) putting key enablers in place and 5) strengthening participation and democratic decision-making.

Many researchers and business organizations have studied the means via which public administrations and citizens can benefit from the availability of eGovernment applications and services. In "I 2010" (2005) is mentioned that: There is a list of 20 basic services, development of which is monitored by the European Commission in all its phases. Twelve of them are "citizen" services: Income taxes, Job search services, Social security benefits, Personal Documents, Car registration, Application for a building permission, Declaration to police, Public libraries (catalogues, search tools), Certificates (birth and marriage), Enrolment in higher education, Announcement of moving and Health-related services. Eight of the services are "business" ones: Social contributions for employees, Corporate tax, VAT, Registration of a new company, Submission of data to statistical offices, Customs declaration, Environment-related permits and Public procurement.

However, even though the internet may be viewed as being a vehicle for government and citizen interaction and a new participatory democracy (Steyaert, 2000), eGovernment is much more than getting information and services online: It is transforming government administration, information provision and service delivery by the application of new technologies, delivering government services in ways that are most convenient to the citizen, and realizing efficiency gains and streamlining government processes (Rimmer, 2002). In order to develop 'citizen-centred' services that provide participants with accessible, relevant information and quality services that are more expedient than traditional 'brick and mortar' transactions, government agencies must first understand the factors that influence citizen adoption of this innovation.

Along this path, lately the two stage approach has been expanded by the insertion of a third stage that consists of feedback and e-participation (Lenova 2009, Osmo, 2008) and that should be the main focus of eGovernment development. According to that the traditional consumerists approach-based model of eGovernment development, which states e-services as the ultimate goal of eGovernment, should be replaced by e- participation model and as a result the key benchmarking indicators should now be related to user – centricity. (Lenova, 2009). Therefore it is obvious that the approach to eGovernment continually adapts to face the emerging challenges (Reem 2009).

The first step towards e-participation is to "listen" to public's opinion. This lately emerged need will be covered by the development of efficient and reliable means to "mine" public opinion through out the World Wide Web.

"Mining someone's opinion" is synonymous to find out what he thinks. The process of finding out what other people think is in most cases a valuable piece of information for most of us during the decision making process. Long before awareness of the World Wide Web became widespread, people were asking their friends who they were planning to vote for in local elections or requested reference letters regarding job applicants from colleagues, or consulted Consumer Reports to decide what product to buy. The internet and the Web have now (among other things) made it possible to find out about the opinions and the experiences of those in the vast pool that are neither our personal acquaintances nor well known professional critics – that is, people we have never heard of. On the other hand companies for decades try to capture public trends and opinions for specific products and services through the use of Gallup polls and questionnaires.

Needless to say that opinions matter a great deal in politics and laws and legislation can change under the (negative) pressure of public opinion. Politicians almost always try to understand what voters are thinking about pending policy or government-regulation proposals.

It is therefore obvious that negative and positive opinions can be used as guidelines for companies to change their strategies toward specific target groups, customers to decide on the purchase of a product or destination place for their holidays and lately for governments to improve services, launch campaigns etc. (Ku et al, 2007).

Here we should note that more and more people are making their opinions available to strangers via the internet (Pang and Lee, 2008). With the explosion of the Web 2.0 platforms such as blogs, discussion forums, peer to peer networks, and various other types of social media citizens have at their disposal a soapbox of unprecedented reach and power by which to share their experiences and opinions positive or negative, regarding any product or service (Zabin and Jefferies, 2008).

Wikis, social networking and folksonomies are often focused on personal life, and many on professional life. In the professional or business environment, both private and public sectors are very interested in offered the best services to the users (Decman, 2009). Web opinion mining aims to extract, summarize, and track various aspects of subjective information on the Web. The latest trend in opinion mining is to extract public opinion expressed in text documents in the web since this information might be more objective since it is expressed without any "pressure". On the other hand the tendency of a person for or against an argument, a product etc is not as easily extracted as in the case of specific questionnaires. It is therefore posing an extra difficulty/challenging in the analysis of this information.

Over the past years a number of research efforts have come up with various proposals for specific or more generic opinion mining tasks (Wiebe, Bruce and O' Hara, 1999; Hatzivassiloglou and Wiebe, 2000; Wiebe, 2000; Wiebe et al., 2002; Riloff, Wiebe and Wilson, 2003; Yu and Hatzivassiloglou, 2003) and assigning them to subcategories such as positive and negative opinions (Pang, Lee and Vaithyanathan, 2002; Turney, 2002; Yu and Hatzivassiloglou, 2003). A variety of machine learning techniques have been employed for this purpose and are generally based on lexical cues associated with opinions.

Liu and his co-workers developed Opinion Observe to compare consumer opinions of different products based on online customer's reviews (Liu et al. 2005). BlogHarvest a blog mining and search framework, conducts opinion comparisons among a set of topics, extracting the interests of the blogger, finding and recommending blogs with similar topics and providing blog oriented search functionality (Sun et al. 2006)..Ku et al. (2007) dealt with mining techniques to deduce positive and negative sentiment words and their weights on the basis of Chinese word structures using web mining techniques. Furuse et al (2007) developed a search engine that can extract opinion sentences relevant to an open-domain query -based not only on positive or negative measurements but also on neutral opinions, requests, advice, and thoughts- from Japanese blog pages. In a more resent study Xu and Ramnath (2009) proposed a system for opinion mining using poll results on the web dealing with opinion answering question, opinion mining on a single object and opinion mining on multiple objects.

AMAZING is a sentiment mining and retrieval system which mines knowledge from consumer product reviews using data mining and information retrieval technology based on a ranking mechanism taking temporal opinion quality and relevance into account to meet customers' information needs (Miao et al. 2009) An opinion utility named Jodange was built in the Leveraging Cornell University. Jodgane identifies opinion holders on issues, organizations, or people of interest. It can track the impact of an issue via publication, region, opinion holder, tonality or any other measurement, uncover important sentiment trends on key issues and correlate opinions against specific outcomes. The VIStology's IBlogs (International Blogs) project, funded by the Air Force Office of Scientific Research's Distributed Intelligence provides blog analysts a tool for monitoring, evaluating, and anticipating the impact of blogs by clustering posts by news event and ranking their significance by relevance, timeliness, specificity and credibility, as measured by novel metrics. This technology allows analysts to discover, from the bottom up, the issues that are important in a local blogosphere, by providing measurements particular to that locale alone.

A common element of current approaches is their focus on either an entire document (Pang, Lee and Vaithyanathan, 2002; Turney, 2002) or on full sentences (Wiebe, Bruce and O'Hara, 1999; Hatzivassiloglou and Wiebe, 2000; Wiebe, 2000; Wiebe et al., 2002; Yu and Hatzivassilogloy, 2003) (Bethard, 2004). Choi et al., (2005), focuses on another aspect of opinion analysis: automatically identifying the sources of the opinions. Identifying opinions sources is especially critical for opinion-oriented question-answering systems (e.g., systems that answer question of the form "How does [X] feel about [Y]?") and opinion-oriented summarization systems, both of which need to distinguish the opinions of one source from those of another. The goal of their research is to identify direct and indirect sources of opinions, emotions, sentiments, and other private states that are expressed in text.

Published news articles often contain factual information along with opinions, either as the outcome of analysis or quoted directly from primary sources. Text materials from many other sources (e.g., the web) also mix facts and opinions. Automatically determining which part of these documents is fact and which is opinion would help in selecting the appropriate type of information given an application and in organizing and presenting that information. EGovernment webs are among the largest webs in existence, based on the size, number of users and number of information providers. Thus, creating a Semantic Web infrastructure to meaningfully organize eGovernment webs is highly desirable.

In any case Semantic Web plays a crucial role in automatic delivery of customized eGovernment services. It extends the existing Web by providing a framework for technologies that give meaning to data and applications for automatic processing (Gribble, et al., 2000). Ontologies[1] are integral to the Semantic Web in facilitating knowledge sharing and reuse. Web services constitute a related technology that has recently emerged to deal with the glut of Web applications. Semantic description of Web services is the key to automating customized service delivery (Medjahed et al, 2003). At the same time, the complexity of the existing eGovernment implementations also challenges the feasibility of Semantic Web creation (Wagner 2006).

Summing up the successful delivery of public policy is of paramount importance and significantly dependent upon the effective technology deployment. Till now the deployment of eGovernment projects has faced various obstacles and in many cases has failed to satisfy the expectations of both the government and citizens in delivering government services (Haya, 2009). In the rest of this paper we will present an approach to leverage the positioning of public opinion in the centre of governmental decisions in an attempt to increase e-inclusion and e-participation.

Opinion mining has recently become a topic of interest trying to combine statistics, Artificial Intelligence and Data Mining technologies in a unified framework (Pang and Lee, 2008).

3. Mining user-generated content expressing opinions about governmental regulations

The aim of our research is to design and implement a method able to automatically detect and analyse the public's stance towards governmental decisions. The availability of such service will help public administrations capture the common understanding of eGovernment. In this respect, considering that people verbalize their opinion in natural language we need to rely on the textual data of their comments (real web content) on governmental issues. Such content can be easily harvested either manually by domain experts who indicate the data sources that need to be collected, or automatically via a trained focused crawler application to eGovernment social media sites, e.g., blogs, fora, etc. Having downloaded the postings, we need to process them 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' viewpoints 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 attempts to mine the citizens' viewpoints towards 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.

The main issue of existing eGovernment services and applications is their failure to capitalize on societal factors. To fill in this void, our method tries to understand how public sector regulations affect the citizens' stance against governmental actions. Thus, we conduct a two-level analysis. The first part of our approach, relying on user online comments, concerns a technique that detects and extracts phrases containing user opinions from their posts. Then, at the second part 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 de-

[1] A formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts.

scribe 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 Mining public opinions on governmental decisions

The first module that our approach integrates is the opinion phrase extraction process. To that end, we need to rely on real web content, as already said, and process it in order to detect and extract both the decision aspects mentioned in the citizens' online postings and the respective citizen evaluations of those decisions. 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 firstly apply HTML parsing to them, which means that we automatically break a data block into smaller chunks by following a set of rules, so that it can be more easily managed, and we eliminate from it non-textual elements. Non-text elements include images, graphical representations of text (including symbols), frames, scripts and video, objects from which we cannot extract useful information. Then we apply tokenization to the posting's body in order to extract the lexical elements of the user generated texts and finally we pass the raw text through a Part-of-Speech tagger, which identifies the word tokens in the text and annotates every token with an appropriate grammatical category. Thereafter, to identify the user opinions in the morphologically annotated text, we use a syntactic dependency parser via which we extract the nouns appearing in the text and their referring adjectives. The reason for relying solely on adjective-noun pairs is based on (i) the findings of Gliozzo et al. (2004) that from all grammatical categories in a text noun represent most accurately the text semantics and (ii) the observation of Turney & Littman (2003) that people use adjectives to evaluate an item or verbalize an opinion. 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 to derive the opinion phrases communicated in the user postings.

Extracted adjective-noun pairs constitute the text chunks that contain implicit user opinions about the issues being discussed in the collected postings. To unravel both the topic of the discussion and people's stance toward this topic, we assume that every noun in the dataset represents a specific aspect of a governmental decision, whereas the adjectives that refer to each of the nouns correspond to the user opinions about the respective decision aspect. Based on this assumption, the problem of identifying how people judge governmental decisions translates into deciphering the sentiment of the adjectives that users select in their online government-related postings. To identify how citizens evaluate governmental decisions, we rely on the notion of words' semantic orientation and we try to discriminate between words of positive and negative sentiments as follows.

We examine every adjective extracted from the harvested postings against the FrameNet ontology (FrameNet) in order to obtain the semantic frame of the adjective. FrameNet project is an on-line lexical resource for English, based on frame semantics and supported by corpus evidence. FrameNet contains more than 10,000 lexical units, more than 6,000 of which are fully annotated, in nearly 800 hierarchically-related semantic frames. Based on the identified frames, we exclude from our data the adjectives that have been assigned the frames: *Increment*, *Relative Time* or *Similarity*, based on the intuition that these are not indicative of the citizens' opinions about the governmental decisions that they discuss in their online postings. Following on, we proceed with the second part of our approach which concerns the sentiment analysis of the user opinions. Sentiment analysis of user opinions refers to labelling opinion phrases with a suitable polarity tag either positive (+) or negative (-), to each of the remaining adjectives in our dataset. In this respect, we rely on the notion of word's semantic orientation, which implies that words and especially adjectives are generally perceived to carry a positive or negative meaning within specific contents. Note however that a number of terms carry a neutral polarity but when it comes to adjectives, neutral ones are less than polarized ones. Therefore, given

the observed polarized nature of adjectives we label their sentiment orientation by manually assigning polarity tags to them. Polarity tags indicate positive or negative orientation. Adjectives lacking a polarity label are considered neutral and are no longer accounted in our method.
The criterion under which labelling takes place is that a positive adjective is one that gives praise to the decision under evaluation, whereas a negative adjective is one that criticizes some or all aspects of the evaluated decision. Note that the manual labelling of the adjectives bears a satisfactory level of objectivity, since several studies (e.g. Hatzivassiloglou and McKeown, 1997) have demonstrated high levels of inter-annotation agreement for the semantic orientation of adjectives. Nevertheless, periodical user studies can always be performed by any interested governmental body before actually putting into action our proposed opinion mining framework. At this point we would like to emphasize that our study objective is not to develop new techniques for text annotation but rather we are interesting in supporting the citizens' participation in eGovernment issues via the availability of an interactive framework that mines the citizens' opinions about governmental decisions and feeds them back to the decision making process.

The result of the above process is a set of positive and negative adjectives that correspond to the opinions of citizens who discuss a governmental decision in their online postings. We refer to those adjective-noun pairs as opinion phrases and we further explore them in order to deduce their strength and their potential encapsulation in future governmental decisions. The overall opinion mining process we suggest is schematically illustrated in Figure 1. Being able to automatically identify and interpret citizen opinions on governmental issues entails a significant benefit toward transforming public sector administrations from mere government service providers into the public's apprentice for confronting bureaucracy. This is because eGovernment cannot be reformed into a significant part of our daily lives unless citizen opinions are taken into account and e-inclusion is encouraged. E-inclusion is a term used to encompass activities related to the achievement of an inclusive information society.

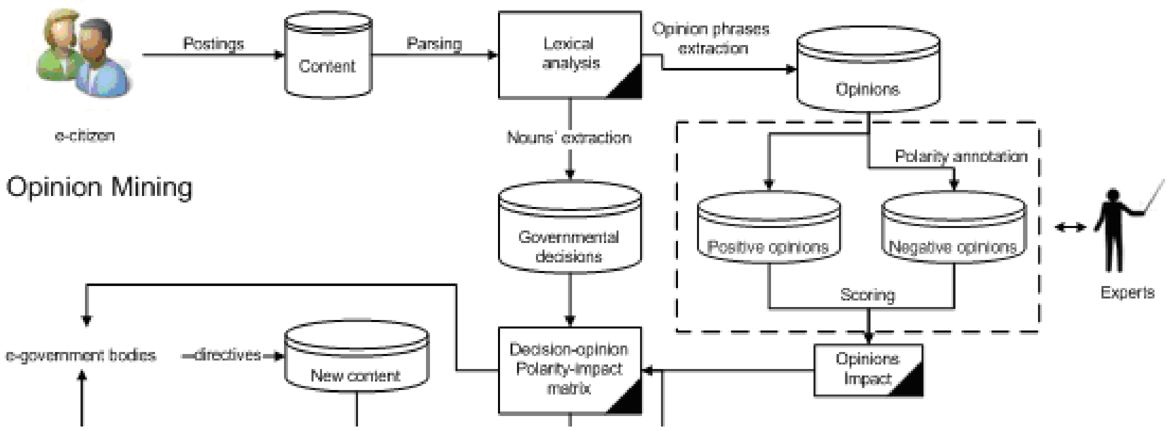

Figure 1: The process of mining user opinions on governmental decisions

Having presented our method for mining user opinions as well as annotating mined opinions with a suitable polarity label, our next step is to utilize the mined user data in order to train a classifier for automatically organizing opinion phrases into bimodal clusters of supportive (i.e., positive) and unsupportive (i.e., negative) citizen views on governmental regulations. The details of our opinion classification approach are discussed next.

3.2 Classifying mined user opinions according to their polarity labels

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 the extracted opinion phrases. The learning accuracy of the classifier is evaluated against the test set and by running several classification iterations. In the experimental Section, we present the details of a classification evaluation we conducted in which we assessed 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. Before presenting our experimental evaluation we discuss how our opinion mining method can serve as a tool for measuring the influence of citizen opinions on governmental issues and therefore how can governmental bodies account for influential opinions into their subsequent regulations and decisions.

3.3 Measuring the strength of citizen opinions for governmental issues

As a final step in our opinion mining approach, we introduce a metric that can statistically analyze the mined user data in order to firstly estimate the overall impact of every opinion mined and then quantify the influence of citizens' opinions on subsequent governmental decisions. By opinion impact, we mean the degree to which an opinion expressed about a governmental directive is commonly shared across citizens, while by opinion influence we denote the probability that an opinion globally shared by the citizens will affect subsequent governmental decisions on similar issues.

To estimate the overall impact of the mined citizen opinions, we will work as follows. At first, we associate every governmental decision discussed in the citizens' online postings that we have previously extracted with the corresponding opinions that have been mined about this decision. Such decision-opinion associations can be easily performed via the use of a relational database management system for the storage of the harvested web data. Then, for every decision for which there are some user opinions expressed (via the use of adjectives) we formulate two data clusters. In the first cluster, we store the adjectives that refer to that decision and which have been assigned a positive polarity label, i.e. the supportive opinions about the decision. In the second cluster, we store the adjectives that refer to the decision and which have been assigned a negative polarity label, i.e. the unsupportive opinions about the decision. This way, we maintain for every governmental decision that is being discussed online a set of positive and negative user opinions. Thereafter, we estimate the impact of every opinion, denoted as Impact (O), as:

$$\mathrm{Impact}\ O = \frac{\sum_{i \in} |O_i\ D}{|O\ D|} \quad \textbf{(1)}$$

Where Oi (D) is the number of times an opinion i appears in the collected e-citizen postings that refer to a decision (D) and O(D) is the number of all opinions contained in the e-citizen postings about D. Impact scores are normalized taking values between 1 and 0; with values close to one indicating that a given opinion is globally shared by the citizens who write a posting about a given decision and values close to zero indicating that the underlying opinion is only representing a few individuals' stance towards the referred governmental decision. Having quantified the impact of every opinion, we associate every mined opinion with its respective polarity label, either positive or negative.

Afterwards, we rely on the above formula and we compute for all the positive opinions about a decision their average impact value in order to quantify the degree to which the decision is generally endorsed by the public. Similarly, we also compute the average impact value of all the negative opinions about a decision in order to deduce the degree to which the latter is rejected by the public. At the end of this process, we annotate every examined governmental decision with the sentiment orientation of the highest average impact in order to indicate how and how much the public values the contribution of the examined governmental decisions.

As a last step, we can estimate the probability that the evaluation result of a governmental decision will affect future decisions on relevant issues. This estimation could provide governmental officials with useful insight on the expected outcome of their decisions on a societal level. To deduce whether an opinion will influence subsequent decisions, we rely on the impact values estimated for that opinion and apply the following criterion:

$$\text{Influence}(O)=\begin{cases} \text{ant} & \text{if avg. Impact}(D) > F \\ \text{insignificant} & \text{if otherwise} \end{cases} \qquad \textbf{(2)}$$

Where the value of *F* can be experimentally set to some threshold based on the statistical analysis of the demographic data collected from the e-citizens whose opinions have been mined. Such data could concern age, popularity, education level, etc. Note though that running experiments to set the value of *F* goes beyond the objective of our work, which focuses on providing a method for mining the public's opinions about governmental issues in an automated manner.

4. Evaluating classification accuracy of polarity-labeled user opinions

So far we have described our method for automatically identifying citizen opinions about governmental decisions in an attempt to assist both the public and governments successfully interact with each other as well as a metric for quantifying the impact of the mined opinions in formulating the public's stance against eGovernment issues. The core of our method is an opinion mining framework, which manages to automatically identify and validate citizen opinions. We have also proposed the architecture of an interactive eGovernment framework that collects the mined opinions, processes them and feeds them back to governmental bodies so that they account for them in their decision making duties. In this section we will focus on the novel method for capturing and assessing citizen opinions on governmental issues trying to experimentally evaluate the performance of our technique in automatically organizing mined opinions in terms of their polarity, i.e., to group positive and negative user opinions on governmental issues separately so that governments can exploit and account for them in their subsequent regulations. Unless we evaluate the accuracy of our model in capturing the public's stance, we will not be able to estimate its impact into transforming eGovernment from a technological infrastructure into an e-democracy application.

4.1 Experimental setup

To assess our study objective, we collected a set of real citizen opinions, downloaded from a Greek forum. The dataset was downloaded from the forum antheorisi[2], which is a Greek forum focusing on policy issues that 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, in order to reduce variability, performs a 10-fold cross validation and generates training and test sets using different partitions and the validation results are averaged over the rounds. Classification training is vital in order to learn the classifiers discriminate between positive and genitive citizen opinions.

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

4.2 Experimental results

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 for each dataset. In particular, the figure depicts the fraction of opinion phrases that were correctly and incorrectly identified by the classifiers as positive or negative.

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. Note that Support Vector Machine is more suitable for text attributes in contrast to Naïve Bayes which has better performance for numerical attributes. 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 organizing opinion phrases in terms of their polarity.

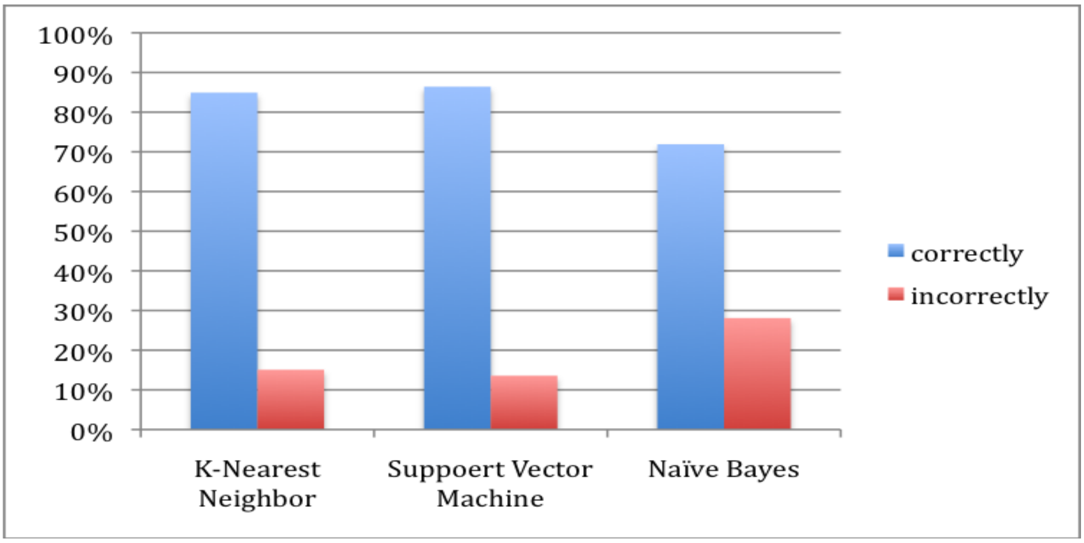

Figure 2: Comparative evaluation of opinion classification accuracy

Table 1: Evaluation of opinion classification accuracy

	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%

Based on our experimental findings, we may deduce the following. First, that our proposed opinion mining and evaluation technique is quite effective in automatically identifying the public's stance towards governmental decisions. Moreover, results demonstrate that our method can be easily integrated into existing classification modules in order for the latter to automatically organize mined user opinions according to the positive or negative orientation. Above all, our experimental study shows that with today's technological advancements it is feasible to deploy existing mechanisms into novel applications such ones related to eGovernment.

5. Concluding remarks and future research directions

In this chapter 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. In addition, we have proposed the architecture of an interactive eGovernment platform that encapsulates the mined user opinions and explores them in subsequent governmental actions. A metric for quantifying the impact of citizen opinions on governmental decisions is also proposed so that the former can be fruitfully employed in subsequent governmental regulations. The application of our proposed method over a set of real user content reveals that properly processed and analyzed opinion phrases can serve as useful 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. Today, all EU Member States have ICT policies and consider them a key contributor to national growth and jobs under the renewed Lisbon agenda. eParticipation is the strongest growing area of eGovernment Action Plan. “eParticipation” is about reconnecting ordinary people with politics and policy-making and making the decision-making processes easier to understand and follow trough the use of new ICT.

As already pointed out by other researchers, one of the most important issues for making eGovernment effective is to enable citizens participate in the decision-making process. Via our proposed approach we ensure that citizen opinions and comments are properly received by public bodies and that they are accounted for in subsequent governmental actions as well as we provide both citizens and governments with the means to effectively interact with each other and actively participate into common actions from which both would benefit. Although the work presented in this chapter is still in early stages and only gives a general notion with respect to how opinion mining techniques can be successfully explored in the course of eGovernment and e-inclusion approaches we believe that it will pave the ground for more initiatives in this respect. As a matter of fact, we are currently working on the incorporation of additional societal aspects in the opinion mining process as well as towards the employment of additional metrics that would evaluate the trustworthiness of the citizen comments and opinions on governmental decisions. Another aspect of future work would be to rely on the mined and polarity-annotated user opinions in order to build and train effective prediction models that would be able to approximate the potential impact of planned governmental decisions on citizens’ stance. Finally, it would be interesting to apply our opinion mining technique towards a wide variety of user opinions on governmental decisions and identify the regulations that interest citizens the most and thus offer them the infrastructure to interact with governmental bodies.

References

Accenture, 2004. “EGovernment Leadership: High performance, Maximum Value.” New York: Accenture.

“i2010 2005– “i 2010 A European Information Society for growth and employment”, Communication from the Commission to the European Parliament, the Council, the EESC and the Committee of the Regions, Brussels, COM(2005) 229 final, 01.06.2005.

Bethard, S.; Yu, Hong.,Thornton, Ashley., Hatzivassiloglou, Vasileios., and Jurafsky, Dan. 2004. “Automatic Extraction of Opinion Propositions and their Holders”. (AAAI 2004),Springer

Choi, Yejin., Cardie, Claire., Riloff, Ellen., and Patwardhan, Siddharth. 2005. “Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns”. In proceedings of Human language technology Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pp 355-362, Vancouver. Association for Computational Linguistics.

Decman. Mitja, 2009. “Web 2.0 in eGovernment: The challenges and opportunities of Wiki in Legal Matters”. Proceedings of the 9th European Conference on eGovernment, pp 229-236.

Fang X and Olivia R. Liu Sheng, 2005 “Designing A Better Web Portal for Digital Government: A Web-mining Based Approach” Proceedings of the 2005 national conference on Digital government research Vol. 89 pp:277 - 278 : National Science Foundation, Digital Government Society of North America, Atlanta, Georgia.

Fellbaum Ch. (ed.). 1998. WordNet: An Online Electronic Lexical Database. The MIT Press.

FrameNet. Available at http://framenet.icsi.berkeley.edu

Gliozzo, A., Strapparava, C. and Dagan, I. 2004. Unsupervised and Supervised Exploitation of Semantic Domains in Lexical Disambiguation. In Computer Speech and Language, vol. 18, no.3, pp. 275-299.

Gribble D. S.et al., 2000 “Scalable, Distributed Data Structures for Internet Service Construction,” Proc. 4th Symp. Operating Systems Design and Implementation, Usenix Assoc., pp. 319–332.

Hatzivassiloglou, V., and Wiebe, J. 2000. “Effects of adjective orientation and gradability on sentence subjectivity.” In proceedings of the Conference on Computational Linguistics (COLING-2000).

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 on European Chapter of the ACL, pp. 174-181.

Hayat. Ali, Linda Macaulay and Liping Zhao. 2009. “A Collaboration Pattern Language for e-Participation”:A Strategy for Reuse”. Proceedings of the 9th European Conference on eGovernment, pp 29-39.

Haya. Alimagwashi and Steve Mcintosh. 2009. “Understanding the Government to eGovernment Transition using a Soft Systems Approach: What is eGovernment Supposed to do?”. Proceedings of the 9th European Conference on eGovernment, pp 45-55.

Irani, Z., Love, P. E. D., Elliman, T., Jones, S. & Themistocleous, M. 2005 "Evaluating eGovernment: learning from the experiences of two UK local authorities". Information Systems Journal, Vol 15 No 1, 61 82

Jian-Tao Sun, Xuanhui Wang, Dou Shen, Hua-Jun Zeng, and Zheng Chen. 2006 “Cws: A comparative web search system.” In International Conference on World Wide Web (WWW), 2006

Lenova. M 2009 “New Index for Measuring Feedback and e-Participation effectiveness of eGovernment in Russia”. Proceedings of the 9th European Conference on eGovernment, pp 445-451.

LW Ku, HH Chen 2007 “Mining opinions from the Web: Beyond relevance retrieval Source”, Journal of the American Society for Information Science and Technology Volume 58, Issue 12,October 2007

Macintsoh, Ann and Angus Whyte. 2006. "Evaluating how e-Participation changes local democracy." in Proceedings of the eGovernment Workshop 2006, edited by Z. Inai and A. Ghoneim. London: Brunel University.

Medjahed B. et al., 2003, “Business-to-Business Interactions: Issues and Enabling Technologies,” VLDB J., vol. 12, no. 1, pp. 59–85

Osamu Furuse, Nobuaki Hiroshima, Setsuo Yamada, and Ryoji Kataoka, 2007 “Opinion sentence search engine on open-domain blog.” In International Joint Conference on Artificial Intelligence (IJCAI).

Osmo. D. 2008. “Benchmarking Government in the web 2.0 era: what to measure, and how”. European journal of e-Practice, V 4, August 2008. http://www.epracticejournal.eu. ISSN:1988-625X.

Pang Bo and Lee Lillian, 2008. “Opinion Mining and Sentiment Analysis.” Foundations and Trends in information Retrieval. Vol 2 Nos. 1-2, pp1-135. DOI:10.1561/1500000001.

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)

Qingliang Miao, Qiudan Li and Ruwei Dai, 2009 “AMAZING: A sentiment mining and retrieval system”, Expert Systems with Applications, in print

Reddick C, 2004, “Citizen interaction with eGovernment: form the streets to servers?”. Government Information Quarterly 22 pp 38 -57, Elsevier.

Reem AL Kaabi and Ezz Hattab. 2009 “eGovernment Success Factors: A Survey”. Proceedings of the 9th

Sacco Maria Giovanni, 2006, “User-Centric Access to eGovernment Information: e-Citizen Discovery of e-Services” Published in Proceeding of the 8th annual international conference on Digital government research: bridging disciplines & domains, American Association Artificial Intelligent (AAAI, www.aaai.org) ISBN:1-59593-599-1

Steyaert ,Jo., 2000. “Local Governments Online and the Role of the Resident: Government Shop Versus Electronic Community”. Social Science Computer Review; vol. 18, 1: pp. 3-16

Turney, P. 2002. “Thumps up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting of the association for Computational linguistics.

Wagner. C, Karen S.K. Cheung, Rachael K.F. Ip, Stefan Bottche. 2006. “Building Semantic Webs for eGovernment with Wiki technology”. Inderscience Publishers. Electronic Government, an International Journal. Issue: Volume 3, Number 1 . pp36-55

Wiebe, J.; Bruce, R.; and O’Hara, T. 1999. Development and use f a gold standard data set for subjectivy classifications. In Proceedings of the 37th Annual Meeting of the Association for Computational Lingyistics (ACL-99), 246-253.

Wiebe, J.2000. “Learning subjective adjectives from corpora”. In Proceedings of the 17th National Conference on Artificial Intelligence (AAAI -2000).

Wiebe, J., Wilson, T.; Bruce, R.; Bell, M.; and Martin, M 2002. “Learning subjective language.” Technical Report TR – 02-100, Department of Computer Science, university of Pittsburgh, Pittsburgh, Pennsylvania.

Yu, H., and Hatzivassiloglou, V 2003. “Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences”. In Proceedings of the Conference on empirical Methods in Natural language Processing (EMNLP-03)

Zabin J and Jefferies A, (2008), “Social media monitoring and analysis: Generating consumer insights from online conversation”, Aberdeen group Benchmark Report.

Z. Xu and R. Ramnath, 2009 “Mining Opinion from Poll Results in Web Pages,” WWW2009, April 20-24, 2009, Madrid Spain