A Government Decision Analytics Framework Based on Citizen Opinion

Mohamed Adel Rezk
Department of Information Systems and Computers
Alexandria University
mohamed.adel@insight-centre.org

Ghada A. El Khayat
Department of Information Systems and Computers
Alexandria University
ghadaek@gmail.com

Adegboyega Ojo
Insight Centre for Data Analytics
National University of Ireland, Galway
adegboyega.ojo@insight-centre.org

Safaa Hussein
Department of Information Systems and Computers
Alexandria University
safaa26bedo@yahoo.com

ABSTRACT
This ongoing research aims to develop a Government Decision Support Framework that employs citizen opinions and sentiments to predict the level of acceptance of newly proposed policies. The system relies on a knowledge base of citizen opinions and an Ontological Model comprising aspects and related terms of different policy domains as an input and a Bayesian predictive procedure. The work proceeds in four basic steps. The first step involves developing domain models comprising aspects for different policy domains in government and automatically acquiring semantically related terms for these aspects from associated policy documents. The second step involves computing citizen sentiments and opinions for the different policy aspects. The third involves updating the ontology with the computed sentiments and the last step involves employing a Bayesian Predictive Process to predict likely citizen opinion for a new proposal (policy) based on information available in the ontology. We provide some background to this work, describe our approach in some detail and discuss the progress made.

CCS Concepts
• Applied computing~E-government

Keywords
Government Decision Support; Decision Analytics; Bayesian Policy Acceptance Prediction; Citizen Satisfaction; Policy Aspects; Opinion Mining; Sentiment Analysis; Semantic Relatedness

1. INTRODUCTION
Citizens-to-Government (C2G) interactions have changed significantly over the last decade due largely to the emergence of Web 2.0 and Social Media. Noticeable government efforts are ongoing for utilizing social media platforms in both pushing information to citizens (one-way interaction) and obtaining timely citizen feedbacks and opinions on government actions. With the very large amount of information generated over these social media platforms, filtering off irrelevant information to harnessing social trends in governmental departments' policies for the purpose of increasing citizen satisfaction [14] is challenging. Using citizen satisfaction data to predict future acceptance or citizen satisfaction of future policy proposal is potentially very beneficial to the government. However, there is currently no established approach for this. Manually monitoring social web contents is ineffective and can consume considerable human resources. Hence, a system for automatic acquisition, classification and aggregation of citizen opinions would be of great value for government decision making.

This research work which is grounded in Opinion mining and Bayesian prediction aims to fill this innovation gap. The proposed system enables the automatic processing of social media contents to determine citizen sentiments and opinions on different aspects of government policies to support government decision-making process. While citizen satisfaction is rarely employed in government decision-making even when citizen surveys are available, we argue that past opinions and perception on related policy issues provide an important input into the decision process. This is important, given that the success of policies and services are measured through citizen satisfaction survey. However, acquiring the required citizen satisfaction rate through surveys on an ongoing basis is resource intensive and infeasible for the government. In this regard, employing computational tools which automatically capture citizen satisfaction and opinions such as Natural Language Processing (NLP) and Machine Learning tools make this goal feasible. The use of similar tools is reported in Teufl et al. [22] in which an e-Participation classification method employing Associative Networks, Spread Activation, and Unsupervised Learning was described. Our goal here is to exploit NLP techniques for Citizen Opinion Mining using policy keywords and a Bayesian predictor for predicting citizen opinions and sentiments using the ontological classes and instances defined in a knowledge base.

2. RELATED WORK
This section reviews three streams of research which are foundational to our proposed work – Government Decision Support Systems, e-Participation and Natural Language Processing.

2.1 Government Decision Support Systems
Decision Support Systems (DSS) are a variety of tools that assist humans for making decisions in different domains. These range from tools for sophisticated interactive decision modeling system to simple information delivery tools [20].

Sprague and Carlson in early 80s provided a narrow definition of DSS as "a class of information system that draws on transaction processing systems and interacts with the other parts of the overall

ICEGOV '15-16, March 01 - 03, 2016, Montevideo, Uruguay

ACM 978-1-4503-3640-6/16/03...$15.00
DOI: http://dx.doi.org/10.1145/2910019.2910093

information system to support the decision-making activities of managers and other knowledge workers in organization."[21].

A broader and more abstract definition by Sharda, Ramesh, Steve H Barr, and James C. McDonnell, better describes the wide spectrum of systems sharing the ability to assist humans in decision-making: "1) 'Decision' emphasizes the primary focus on decision making in a problem situation rather than the subordinate activities of simple information retrieval, processing or reporting; 2) 'Support' clarifies the computer's role in aiding rather than replacing the decision maker and 3) 'System' highlights the integrated nature of the overall approach, suggesting the wider context of the user, machine and decision environment." [20].

In 2008, Power, Daniel J. had published a useful Decision Support Systems History overview [16], summarizing the previous four decades work on the DSS area. They presented in their research seven types of DSS identified by Steven Alter in his doctoral dissertation back in 1980. These types are as follows:

a. File drawer systems.

b. Data analysis systems.

c. Analysis information systems.

d. Accounting and financial models.

e. Representational models.

f. Optimization models.

g. Suggestion models [16]."

According to this categorization of Decision Support Systems, the proposed system falls under Data Analysis Systems. The inclusion of citizen input in government decision-making processes are widely studied in the domain of e-Participation explained below.

2.2 e-Participation and Citizen-to-Government Interaction Analysis

This research theme is concerned with the analysis of C2G interactions. E-participation is defined as: "The participation of individuals and legal entities and groups thereof in the decision-making process in the branches of government using information and communication technology (ICT) equipment. In the context of a government's e-government activities, we interpret this as an offer to participate, in a form enhanced by the use of ICT, in plans and decisions at different levels of government.

In the international debate, a distinction is made between the term 'e-Participation' and the term 'e-Democracy' in that the latter also covers elections as the most binding form of citizen participation [6]. Hence, we stress here that study does not deal with e-voting [1]. In general, e-Participation describes efforts to broaden and deepen political participation by enabling citizens to connect with one another and with their elected representatives and governments by using Information and Communication Technologies (ICT) [5]. By "Citizen-to-Government", we mean the interactions between citizens and government through Social Media channels such as Twitter [3] and Facebook. These initiatives are important for government decision effectiveness [19, 12, 9, 8, 15, 4]. To process the contents generated from the C2G interactions, Natural Language Processing tools described in the next section are required.

2.3 Natural Language Processing Tools

Natural Language Processing (NLP) is concerned with the use of computational tools for processing human or natural languages.

Different NLP tools will be applied in different parts of this research. NLP techniques are introduced and developed in the literature for text analysis. For the purpose of this research, NLP tools will be utilized as follows:

a. Named Entity Recognition (NER) tools such as Stanford Named Entity Recognizer [7] and Alchemy Entity Extractor reviewed in [18] are required for extracting entities from social media contents and policy documents.

b. Sentiment Analysis or opinion mining tools such as Sentiwordnet citeesuli2006sentiwordnet and WEKA assisted [23] or a combined approach [17] would be used for computing the sentiments associated with identified entities.

c. Knowledge Bases and ontologies to assist in text analysis e.g. DBpedia [2] will be used as language resources.

d. Semantic Similarity tools to be used to determine related terms and topics in the contents generated from C2G interactions. The Semantic Similarity tool is Extracting DISCO described in [10,11] will be adopted in this work.

3. METHODOLOGY

The proposed research will tackle the following two questions:

a. What are the issues and challenges for Government Support Systems in harnessing the contents generated from C2G interactions on social media platforms?

b. How can we develop an NLP-based tool to harness citizen sentiments and opinions inherent in contents generated from C2G interactions to improve the chances for better citizen acceptance and satisfaction of future government policies and initiatives?

The first question will be answered through a thorough literature study while the second question will be answered by constructing a technology artifact based on the Design Science Research methodology [13].

4. THE PROPOSED DECISION ANALYTICS FRAMEWORK

This section highlights the major elements of our framework which will tackle the second research question above (see Figure 1).

Policy - Policy documents will be fed as text input into the proposed system that will initiate the whole processing cycle. Contents on social media are assumed to be related to one or more public policies.

Keyword Extraction and Recognition - Origin Keyword is a keyword extracted from the original policy text fed to the system. Named Entity Recognition Algorithm will be used to produce the origin keywords.

Semantically Related Keyword Recognition (Branch Keyword): a set of keyword generated by applying semantic relatedness algorithm over origin keywords.

Public Policy Ontology Modeling	Computational Analysis of Citizen Opinions and Sentiments	Knowledgebase Population	Prediction
• Investigated Public Policy. • Origin and Branch Keyword Extraction from Policy Document. • Linking Keywords to Policy Aspects.	• Harvesting Citizens Contents from Twitter. • Opinion mining.	• Populating our ontology with Keywords. • Attaching Citizen Opinions.	• Mining the Accumulated Knowledge to calculate citizen satisfaction rates towards policy aspects. • Producing Citizen Satisfaction Insights.

Figure 1: Citizens' Satisfaction Analysis Model

Policy Aspects Detection - An aspect of the input Policy to be detected by applying semantic relatedness algorithm over origin keywords and branch keywords towards a set of domain aspects gathered previously.

Harvesting Citizens' content related to Policy (Real-time Scenario only) - Using origin and branch keywords, the system will start harvesting contents generated by citizens, starting with the Twitter platform.

Opinion Mining – It involves the use of sentiment analysis algorithm on citizen contents and aggregation of computed sentiments over contents associated with a Policy aspect. The output is used both in the knowledge acquisition phase and in the prediction phase.

Knowledge Base (KB) Construction - Real-time Scenario – Constructed based on the Public Policy Ontology and relates the original and branch keywords with policy aspects. Policy aspects are associated with sentiments value. It provides data for estimating sentiments for new policies.

Citizens' Satisfaction Rates Computation - Element calculates the satisfaction rates on policy aspects using the KB.

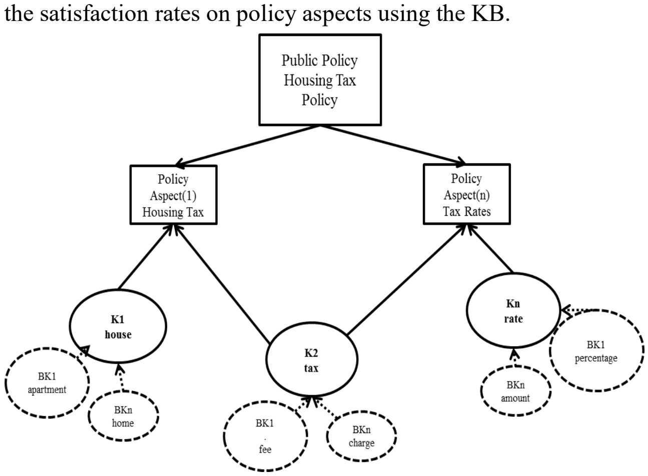

Figure 2: Policy Aspect Detection Model

4.1 Usage Scenarios

This section provides examples of two scenarios in which the proposed system could be used.

First Scenario - Real-time example: A government introduced a Housing Tax policy three months ago and would like to know the level of citizen's satisfaction based on citizen comments and opinions expressed on Twitter. The information will be used to tune or adjust the policy towards better citizen satisfaction. Using our system, in this case, involves first feeding the policy text as input

Citizens' Satisfaction Insights – Gives the user an indication for the required action for policy aspects e.g. needs revising, good.

In order to break down a policy into aspects, we are proposing a semantic relatedness approach, which calculates the distance between extracted keywords from the policy text and a pre-defined aspects in the Policy Domain Model (see Figure 2).

There are two usage scenarios for the system. The first scenario called the Real-time Scenario involves continuous monitoring and generation of sentiments and opinions related to Policy aspects and storing the information in the KB. The second scenario called the "Prediction scenario" involves applying a Bayesian process in predicting the likely opinion (or citizen satisfaction rate) on a new policy based on the opinions and sentiments associated with policy aspects and keywords using the information in the KB. Predicting the likely citizen opinion over a new policy is based on estimating of the estimated opinion on its various policy aspects. The opinion on a policy aspect is computed as an aggregate of the sentiments on related keywords (both origin and branch keywords) as shown in Figure 3.

Thus in a government decision context, our system could be used as-as a real-time citizens' satisfaction rate calculator based on an input policy text.

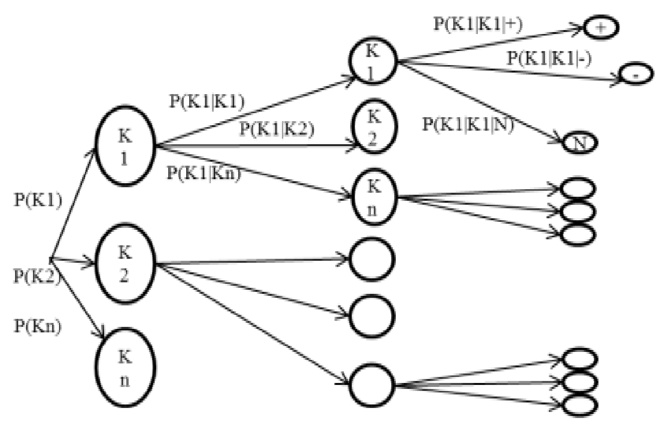

Figure 3: Prediction using Bayes Theorem

into the system charge to extract origin keywords and branch keywords then detect policy aspects (Housing tax and Tax rates) and start harvesting relevant contents from Twitter. Next, the system will apply opinion mining and calculate satisfaction rates towards policy and its detected policy aspects. Finally, it will give insights based on the computed opinions to advise decision makers about what aspects of the policy may need revision.

Second Scenario - Prediction example - In this case a new Investment policy is under analysis before introducing it to citizens, and the decision maker seeks to predict citizen satisfaction rate before introducing it to the public. By using the system in prediction mode, it will extract origin and branch keywords then detect policy aspects using NER and Semantic relatedness algorithms (Tax rates and Duty-free areas). The final step involves the use of the Bayesian prediction to compute the probable citizen satisfaction based on already stored opinion and sentiments information in the information in the KB.

5. CONCLUSION AND FUTURE WORK

Our proposed work addresses the problem of lack of tools to support critical government decision making in which knowledge of citizen opinions expressed on social media constitute a critical input. We have proposed the development of a tool for processing relevant contents generated on social media to assess level of citizen satisfaction based a framework harnessing tools and techniques from the domains of Decision Support and Analytics, e-Participation and NLP. We believe that the availability of such a Government Decision Analytics Framework will significantly contribute to smarter and more agile policy making and tuning.

We have completed the concept phase and are now developing a Public Policy Ontology to support the specification of policy aspects and related keywords (both origin and branch) as well as associated sentiments.

6. REFERENCES

[1] S. Albrecht, N. Kohlrausch, H. Kubicek, B. Lippa, O. M¨arker, M. Tr´enel, V. Vorwerk, H. Westholm, and C. Wiedwald. eparticipation-electronic participation of citizens and the business community in e-government. Study on Behalf of the Federal Ministry of the Interior (Germany), Division IT, 1(1), 2008.

[2] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: A nucleus for a web of open data. Springer, 2007.

[3] E. Bonso´n, S. Royo, and M. Ratkai. Citizens' engagement on local governments' facebook sites. an empirical analysis: The impact of different media and content types in western europe. Government Information Quarterly, 32(1):52–62, 2015.

[4] S. A. Chun, S. Shulman, R. Sandoval, and E. Hovy. Government 2.0: Making connections between citizens, data and government. Information Polity, 15(1):1, 2010.

[5] DEMO-net. (2006b). Introducing eParticipation. DEMO-net booklet series, no, 1., 2006.

[6] DEMO-net. (2007c). Introducing eParticipation. DEMO-net booklet series, no, 1., Jan. 2008.

[7] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pages 363–370. Association Association for Computational Linguistics, 2005.

[8] N. Karacapilidis, E. Loukis, and S. Dimopoulos. Computer supported g2g collaboration for public policy and decision-making. Journal of Enterprise Information Management, 18(5):602–624, 2005.

[9] R. Klinger, P. Senger, S. Madan, and M. Jacovi. Online communities support policy-making: The need for data analysis. In Electronic Participation, pages 132–143. Springer, 2012.

[10] P. Kolb. Disco: A multilingual database of distributionally similar words. Proceedings of KONVENS-2008, Berlin, 2008.

[11] P. Kolb. Experiments on the difference between semantic similarity and relatedness. In Proceedings of the 17th Nordic Conference on Computational Linguistics NODALIDA'09, 2009.

[12] A. Macintosh, T. F. Gordon, and A. Renton. Providing argument support for e-participation. Journal of Information Technology & Politics, 6(1):43–59, 2009.

[13] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. Journal of management information systems, 24(3):45–77, 2007.

[14] J. Peignot, A. Peneranda, S. Amabile, and G. Marcel. Strategic decision support systems for local government: A performance management issue? The use of information systems on the decision making and performance management of local government. International Business Research, 6(2):p92, 2012.

[15] V. Peristeras, G. Mentzas, K. Tarabanis, A. Abecker, et al. Transforming e-government and e-participation through it. Intelligent Systems, IEEE, 24(5):14–19, 2009.

[16] D. J. Power. Decision support systems: a historical overview. In Handbook on Decision Support Systems 1, pages 121–140. Springer, 2008.

[17] R. Prabowo and M. Thelwall. Sentiment analysis: A combined approach. Journal of Informetrics, 3(2):143–157, 2009.

[18] K. J. Rodriquez, M. Bryant, T. Blanke, and M. Luszczynska. Comparison of named entity recognition tools for raw ocr text. In Proceedings of KONVENS, pages 410–414, 2012.

[19] Ø. Sæbø, J. Rose, and L. S. Flak. The shape of eparticipation: Characterizing an emerging research area. Government information quarterly, 25(3):400–428, 2008.

[20] R. Sharda, S. H. Barr, and J. C. MCDonnell. Decision support system effectiveness: a review and an empirical test. Management science, 34(2):139–159, 1988.

[21] R. H. Sprague Jr and E. D. Carlson. Building effective decision support systems. Prentice Hall Professional Technical Reference, 1982.

[22] P. Teufl, U. Payer, and P. Parycek. Automated analysis of e-participation data by utilizing associative networks, spreading activation and unsupervised learning. In Electronic Participation, pages 139–150. Springer, 2009.

[23] V. Umadevi. Sentiment analysis using weka.International Journal of Engineering Trends and Technology (IJETT), 18, Dec. 2014