Citymis OpTree: Intelligent Citizen Management Using Sentiment Analysis and Opinion Trees

Florencia Gamboa Marcos Malamud
Mismatica Management
Martín y Omar 129. Oficina 206, CP 1642, San Isidro, Buenos Aires, Argentina
Tel: +54-11-60939386
E-mail: florenciagamboa212@gmail.com, malamud@mismatica.com

Ana Maguitman Carlos Chesñevar
ICIC-CONICET
Dep. Of Computer Science and Eng.
Universidad Nacional del Sur
Alem 1253 - B8000CPB B.Blanca, Argentina
Tel. : +54-291-459-5135
E-mail: {agm,cic}@cs.uns.edu.ar

ABSTRACT

Citymis is a web and mobile application, designed for the operational management of reports on municipal utilities. It is very easy to use both for residents and for municipal officials, and was designed especially for small and medium municipalities. Citymis is an eGov app through which municipalities listen and respond in real time in a collaborative network, significantly reducing the response time and improving the quality of the interaction with citizens. Recent advances in intelligent systems for eGov have led to the integration of sentiment analysis and argumentation technologies. In this paper we discuss work in progress on an extension of Citymis, called Citymis OpTree, which extends the natural capabilities of Citymis by adding intelligent opinion mining.

CCS Concepts

• Information systems~Decision support systems • Computing methodologies~Artificial intelligence • Applied computing~E-government

Keywords

Government 2.0; e-Governance; Participation; Intelligent Tools

1. INTRODUCTION AND MOTIVATIONS

Several citizen management platforms have been developed in the last years (e.g., [1], [2]), providing suitable channels for efficient electronic communication and coordination connecting the involved stakeholders (citizens, government officials, etc.). In the particular context of intelligent city management, the extensive deployment of mobile phones makes it possible for many citizens to easily provide input concerning opinion, needs, complaints, etc. related to different aspects of everyday life. Mobile and web apps allow citizen to provide information in real time, which can be compiled by government officials in charge of providing solutions. Nevertheless, existing eGov platforms do not model and process emerging collective thinking patterns in communities. Collective thinking patterns could correspond to ideas, proposals, criticisms or viewpoints, which decision makers can identify and confront based on atomic, individual inputs from citizens and users, such as tweets, Facebook posts, web-based product reviews, etc. Such patterns can take place in different contexts associated with social innovation and change, e.g. crowdfunding initiatives, opinion mining, citizen journalism, cyberactivism, etc.

Electronic Empowerment Participation (E2P) [3,4] captures a radically new perspective on e-Participation, where collective thinking patterns can be identified under the generic form of “arguments” [5,6], being contrasted automatically, enhancing thus the abilities of the different stakeholders to engage in creative participatory processes. The underlying machinery that makes E2P possible is given by agreement technologies [7], a new metaphor that integrates several aspects from database theory, artificial intelligence, multi-agent systems and social infrastructures.

Citymis is an eGov application developed by Mismatica Management (http://www.mismatica.management) through which municipalities listen and respond in real time in a collaborative network, significantly reducing the response time and improving the quality of the interaction with citizens. Recent advances in intelligent systems for eGov have led to the integration of sentiment analysis and argumentation technologies. This ongoing-research paper summarizes some of the main advances on applying intelligent opinion mining to provide an argumentation characterization for a collection of comments collected through an eGov platform. In particular, it presents Citymis OpTree, a prototype tool that results from applying the proposed approach to the analysis of reports that have been submitted through the Citymis application.

2. CITYMIS - OVERVIEW

The Citymis application (http://www.citymis.info - Citizen Centric E-Gov), allows users to report issues online. To report an issue, the user has to specify the category of the issue that is being reported. The possible categories include street conditions, transit and traffic signals, public lighting system, urban hygiene, running water, urban trees, parks and squares, lost pets and pets for adoption. In the meantime, each of these categories has a list of associated issues. For instance, the public lighting system category includes issues such as streetlight off, blinking streetlight, new light request, and streetlight repair request, among others. The user has to type in the street and number where the problem has been noticed. Alternatively, a map can be used to enter the location of the issue, providing a mechanisms that not only allows to specify the exact geographic position but also to find and examine other issues reported in the same or surrounding area. In addition the user can fill in a comment field and provide a photo to better describe the reported issue. The application also requests the contact information of the user.

From an operational perspective, a municipality is divided into geographic sub-regions, usually referred to as "areas of assignment''. This allows to allocate reports to specific employees (inspector, maintenance team, etc.) to solve or verify the reported issues. Each report is firstly associated with a "management unit", which is basically a geographic location (latitude, longitude, street and number), and it initially enters the "pending moderation" state. At this point, the report can be accepted, in which case it is assigned the "in hand" status, or it can be dismissed, receiving the "rejected" status. Once the report has been accepted, the remaining step is to solve the reported problem in order to attain the "fulfilled" status and complete the process. Each time there is a change in the status of the report, the user that submitted the report receives an email notification and is invited to add new comments, to provide additional data that can help solve the problem, or simply to give an opinion of the service.

Currently, Citymis is actively being used by 16 municipalities in Argentina, Mexico and Chile. In the meantime, approximately 50 municipalities are testing the demo version of the application, covering more than 15,000,000 citizens.

3. EXTENDING CITYMIS WITH SENTIMETN ANALYSIS AND OPINION TREES

Identifying collective thinking patterns or prevailing problems in a community from the data collected through the Citymis application involves addressing a number of important issues. On the one hand, as the magnitude of the comments entered by the users increases, it becomes necessary to rely on text-mining techniques to filter noise and detect meaningful patterns in the data. On the other hand, such data are usually incomplete, or potentially inconsistent, as citizens (users) might have different views on certain issue, resulting in different pro and con arguments which have to be assessed and confronted against by authorities in order to make a decision. Such decisions, on their turn, are to be backed by arguments when informed back to citizens in order to gain acceptance.

In the light of these issues, we have developed Citymis OpTree, a tool that helps to impose an initial structure on the collected material, leading to the construction of opinion trees. Given a query representing a specific topic (e.g., "streetlight system"), an opinion tree associated with the query is a tree-like structure with different sentiments associated with the topic as a whole, as well as with the subtopics derived from the general topic. Citymis OpTree offers the possibility of getting a structured picture of how different comments and opinions interrelate. The tool facilitates decision making by government officials as well as analysis by the general public by mining comments and opinions and organizing them in a structured form

Building an opinion tree involves the following steps: *Step 1:* Given a topic of interest represented by the query Q and some selection criteria *C* (such as location, date range, type of service, or type of problem), all comments matching *Q* and *C* are collected and associated with the root of the tree.

Step 2: A small set of descriptors $\{d_1, d_2,.., d_n\}$ is strategically selected from the terms occurring in the collected comments to extend the query *Q*. This results in a sequence of new queries extending *Q*, namely $Q \cup \{d_1\}$, $Q \cup \{d_2\}$,…, $Q \cup \{d_n\}$.

Step 3: For each new query $Q \cup \{d_i\}$, all comments matching the query are identified and used to build an immediate descendant of *Q*. This allows to incrementally construct a tree-like structure (rooted in *Q*), where each node is associated with a query and its matching comments.

Step 4: Repeat steps 2 and 3 each time a new node is added to the tree, creating additional levels of the tree, until no alternative ways of constructing more specific queries are possible.

Step 5: For each node in the tree, the prevailing sentiment (positive, negative or neutral) is computed based on the set of associated comments.

Various techniques are being analyzed to strategically select a good set of descriptors from the terms occurring in the collected comments (*Step 2*). The analysis of these techniques lies outside the scope of this paper. For a description of one of these strategies, which relies on the identification of descriptors for sets of opinions grouped into equivalence classes, see [8].

A naïve Bayes classifier was used to assign a prevailing sentiment to each set of comments associated with the nodes of the tree (*Step 5*). The machine learning library *scikit-learn* [9] was used to implement this classifier. To generate the necessary examples to train the classifier we used a Spanish-language corpus of pre-classified movie reviews. The resulting sentiment analysis classifier was enhanced by incorporating an ad-hoc list of Spanish language terms labeled as positive or negative.

4. A CASE STUDY: ASSESSING THE QUALITY OF THE PUBLIC LIGHTING SYSTEM

This section illustrates the Citymis OpTree tool by showing how an opinion tree for a specific query *Q* and criteria *C* can be constructed and visualized. The example has been created using reports from a medium size municipality (approx. 350,000 citizens). The dataset employed by the tool is in Spanish and contains approximately 90,000 comments coming from citizens that used a mobile app, the Citymis web system or a toll-free phone number to report a problem. The dataset also contains nearly 310,000 comments that are part of conversations associated with existing reports (for instance comments coming from registered users that replied to reports submitted by citizens, which in their turn can contain further answers and comments). In addition, the dataset contains roughly 25,000 reports that were generated by users that submitted a comment directly through the iOS, Android or WinPhone Citymis apps.

In order to develop opinion trees, the Citymis OpTree algorithm has access to various attributes associated with each report, such as the report date, the type of service, the type of problem and the location of the problem (latitude, longitude, street name and number). In addition, the tool also has access to the user's comments (description of the claim), photo (if available) and the solution to the problem.

Based on the data available through the Citymis application, a number of approaches can be taken to identify and analyze problems with a specific service. By adjusting the criteria *C* in the OpTree algorithm, different perspectives can be obtained on the problem under analysis. In particular, the date of the claim can be used to restrict the construction of an opinion tree to reports submitted during a specific date range. The date can also be used to monitor how the service has improved or deteriorated through time in the whole municipality or in specific areas, if location

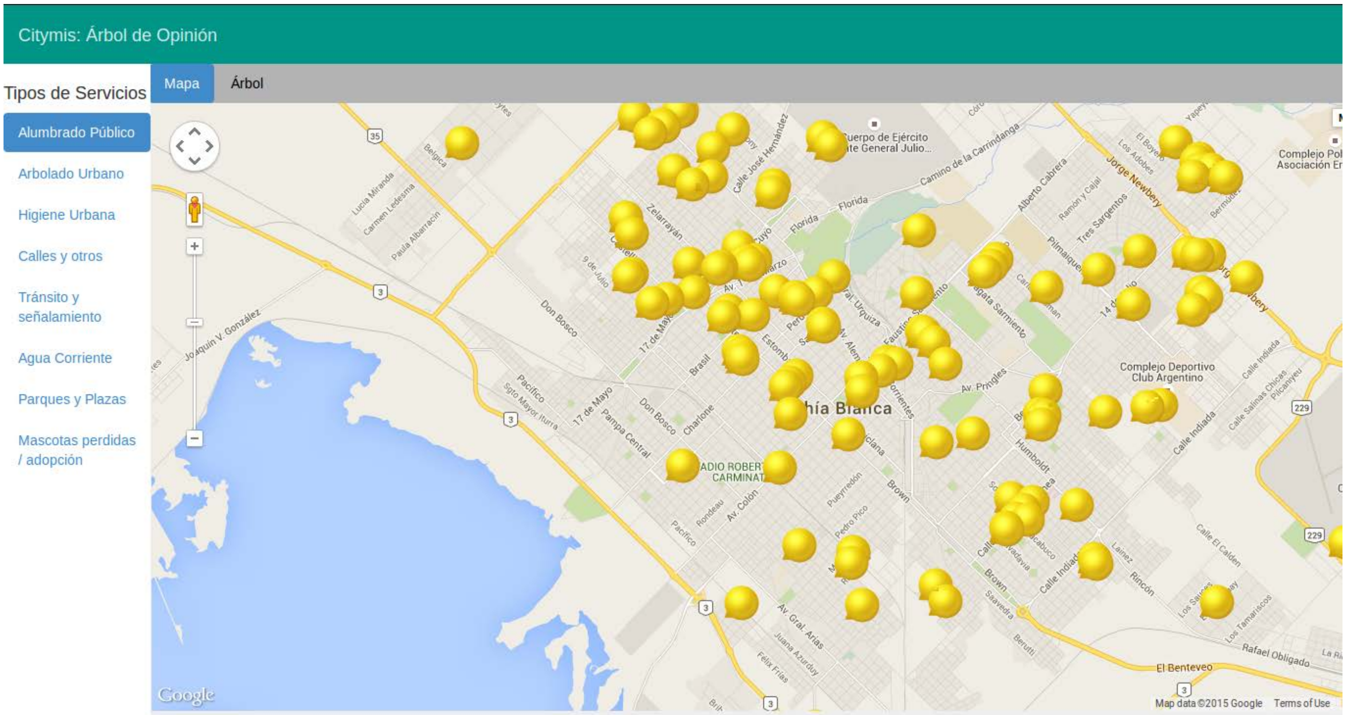

Figure 1: The geographic location of reports associated with the public lighting system.

information is also considered. Visualizing reports about certain types of services on a map can also help identify prevailing problems and patterns related to those services. Figure 1 presents a map showing the location of various reports associated with the public lighting system in a municipality.

Augmenting this analysis with the construction of opinion trees can enhance the decision making process, and can provide users with access to this process by offering a structured and coherent view of the comments and solutions.

To illustrate this analysis, Figure 2 presents an opinion tree derived from the query “luminaria” (Spanish word for “streetlight system”). The opinion tree is represented by a tree-like structure with sentiments associated with different nodes. Each node is labeled with a word, and is characterized by a color and a size. The root node is labeled with the initial query used to derive the entire opinion tree. The other nodes are labeled with words that represent more specific queries, which were incrementally constructed by the algorithm starting from the initial query by adding one word at a time. For instance, the word “barrio” (“neighborhood”) that appears at the third level of the tree indicates that the corresponding node is associated with the query “luminaria luz barrio” (“streetlight system light neighborhood”).

The color red is used to denote that the predominant sentiment associated with the set of comments obtained for the corresponding query is negative, while the color green is used to indicate a positive sentiment. For instance, the node labeled with the word “robo” (“robbery”) is associated with a negative predominant sentiment while the node labeled with the word “obras” (“construction work”) has a positive predominant sentiment. To aid visualization in Figure 2, the nodes associated with a positive sentiment were marked with

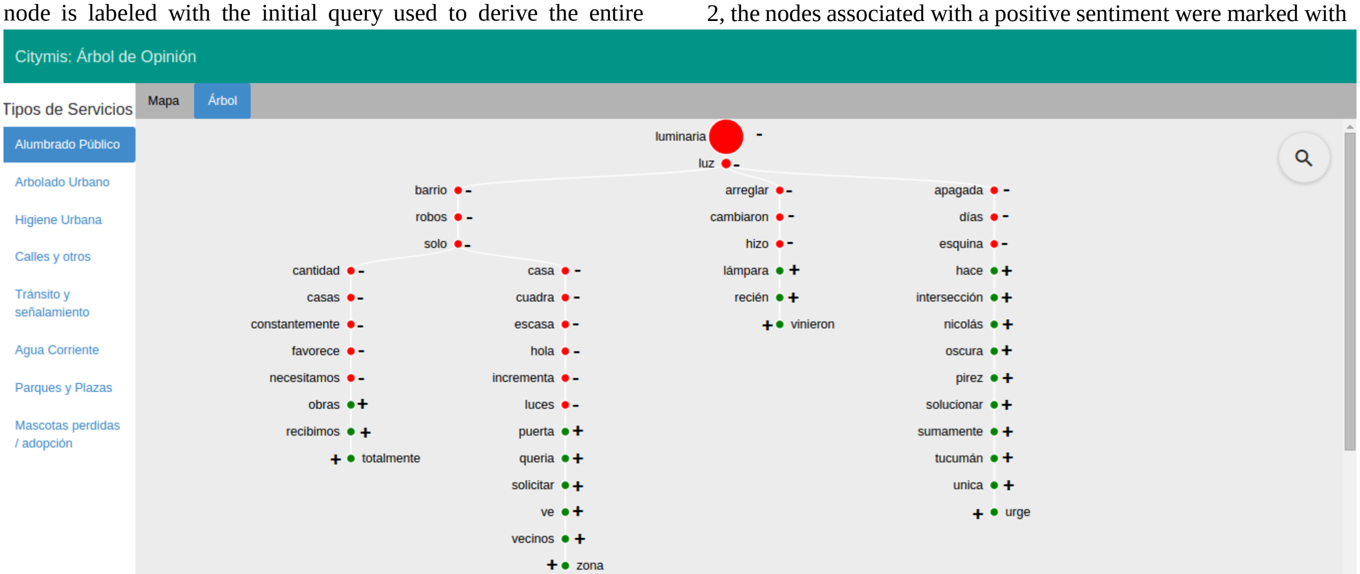

Figure 2: Opinion tree obtained for the query “luminaria” (“streetlight system”).

"+", while those associated with a negative sentiment were marked with "-". The size of each node is proportional to the number of comments associated with the corresponding query. Note that as we descend towards the leaves, the queries become more specific and therefore the size of the node will decrease as the number of comments will naturally be smaller.

Given an opinion tree we might be interested in finding a minimal structure that reflects all existing conflicts between comments in the tree. In other words, we might want to build a minimal tree such that arguments and counter-arguments are easy to visualize. To accomplish this, it is possible to apply a partitioning algorithm to generate a minimal structure that preserves the conflicts between comments existing in the original opinion tree. The application of this algorithm results in a natural grouping of comments that are related to and no conflicting with each other, forming equivalence classes of comments represented by arguments. For a detailed description of this algorithm see [8], where the algorithm is applied in the context of opinion mining from Twitter.

5. RELATED WORK

The developed tool is related to recent research were argumentation and social networks are integrated to enhance e-democracy. A tool closely related to our proposal is the Parmenides system [10], which applies argumentation technologies in an e-government context. This tool allows the government to present policy proposals to the public so that users can submit their opinions on the justification presented for the particular policy. In contrast with our approach, this research work is intended to the analysis of proposals rather than to the analysis of users' reports in a municipality. In addition, the Parmenides tool assumes that argument schemas are established beforehand, and are not detected automatically.

Another related proposal is presented in [11], where the authors show how the theory of argumentation schemes can provide valuable support to formalize and structure online discussions and user opinions. Differently from our proposal, this approach aims at applications different from e-government (e.g., business) and does not address the issue of aggregating comments and associating them with high-level arguments.

The Collaboratorium system [12] provides a platform that enables collaborative deliberation where users can create networks of posts organized as an argument map. In this sense, this system resembles our proposal in that it adopts knowledge sharing technologies to facilitate logic-based knowledge organization. However, differently from our proposal, it is not intended to mine reports to automatically identify conflicting opinions but to support large-scale argumentation, where users are allowed to enter arguments and a moderator takes a key role.

Another related proposal is presented in [13], where the authors investigate how governments can make sense of large amounts of policy inputs received from citizens with the aid of online consultation forums. Similarly to our proposal, this work explores the use of technology in visualizing argumentation data. However, it does not attempt to automatically construct arguments or to apply opinion mining.

6. CONCLUSION

In this paper we have presented Citymis OpTree, a prototype that applies intelligent opinion mining to provide an argument-based characterization of the reports and comments collected through the Citymis eGov platform. Citymis OpTree supports the process of identifying pervasive problems and conflicting views associated with issues in a community.

As part of our future work we plan to develop robust methods to identify good descriptors that will help construct better opinion trees. We also plan to apply a partitioning algorithm as the one described in [8] to generate minimal structures that facilitate the identification of conflicts among opinion. Finally, we plan to conduct experiments and user studies to assess the effectiveness of the proposed tool.

7. ACKNOWLEDGMENTS

This research was partially supported by LACCIR (Microsoft Research Latinamerica) through Project R1211lAC004, PICT-ANPCyT 2014-0624, PIP-CONICET 112-2012010-0487, PGI-UNS 24/N039, and PGI-UNS 24/N040.

8. REFERENCES

[1] P.-L. Sun, C.-Y. Ku, D.-H. Shih: An implementation framework for E-Government 2.0, Telematics and Informatics, Volume 32, Issue 3, August 2015, 504-520.

[2] O. Fedotova, L. Teixeira, H. Alvelos: E-participation in Portugal: evaluation of government electronic platforms, Procedia Technology, Volume 5, 2012, 152-161.

[3] C. Chesñevar, A. Maguitman, E. Estevez, N. Osman, C. Sierra: E2 participation: electronically empowering citizens for social innovation through agreement technologies. DG.O 2013: 279-280.

[4] C. Chesñevar, A. Maguitman, M. P. González: Empowering citizens through opinion mining from twitter-based arguments. In Proc. ICEGOV 2014, Guimaraes, Portugal, October 2014, pp 275-278.

[5] S. Toulmin: The uses of argument. Cambridge: Cambridge University Press, 1959.

[6] I.Rahwan, G.Simari (Eds.): Argumentation in Artificial Intelligence. Springer, 2009.

[7] S. Ossowski (ed.): Handbook of Agreement Technologies, volume 8 of Law, Governance and Technology, Springer, New York, January 2013.

[8] K. Grosse, M. P. González, C. Chesñevar, A. Maguitman: Integrating Argumentation and Sentiment Analysis for Mining Opinions from Twitter. In AI Communications, IOS Press (in press), 2015.

[9] The Scikit-Learn Machine Learning Library. http://scikit-learn.org/stable/. Accessed June 2015.

[10] D. Cartwright, K. Atkinson: Using Computational Argumentation to Support E-participation. IEEE Intelligent Systems, 2009, 24(5): 42-52.

[11] S. Heras, K. Atkinson, V. Botti, F. Grasso, V. Julian, P. McBurney: How argumentation can enhance dialogues in social networks. COMMA, 2010. IOS Press, 267–274.

[12] M. Klein, L. Iandoli: Supporting collaborative deliberation using a large-scale argumentation system: The MIT Collaboratorium, 2008. Technical Report 4691-08, MIT Sloan.

[13] F. Steibel, E. Estevez: Designing Web 2.0 tools for online public consultation. Impact of Information Society Research in the Global South, 2015, pp 243-263, Springer.