Opinion Mining and Sentiment Analysis in Policy Formulation Initiatives: The EU-Community Approach

Yannis Charalabidis[*], Manolis Maragoudakis[¤] and Euripides Loukis[¤]

[*]National Technical University Athens-University of the Aegean
`yannisx@aegean.gr`
[¤]University of the Aegean
2 Palama Str., Karlovasi, Samos
83200, Greece
`{mmarag,eloukis}@aegean.gr`

Abstract. In the last decade there is extensive and continuously growing creation of political content in the Internet, and especially in the Web 2.0 social media, which can be quite useful for government agencies in order to understand the needs and problems of societies and formulate effective public policies for addressing them. So a variety of ICT-based methods have been developed for the exploitation of this political content by governments ('citizensourcing'), initially simpler and later more sophisticated ones. These ICT-based methods are increasingly based on the use of opinion mining (OM) and sentiment analysis (SA) techniques, in order to process the extensive political content collected from numerous sources. This paper describes a novel approach to OM and SA use, created as part of an advanced ICT-based method of exploiting political content created in the Internet, and especially in social media, by experts ('expertsourcing'), aiming to leverage the extensive policy community of the European Union, which is developed in the European EU-Community project. Furthermore, some first experimental results of it are presented.

Keywords: Opinion Mining, Sentiment Analysis, Public Policy, Social Media

1   Introduction

In the last decade there is extensive and continuously growing creation of political content in the Internet, and especially in the Web 2.0 social media, which has the form of numerous political postings, comments, articles and debates in various electronic spaces (e.g. web sites and social media accounts of political institutions and traditional or electronic media, political blogs, etc.). Citizens are increasingly using the above electronic media as efficient channels for the creation and exchange of extensive political content, and also for the quick organization of collective political action with large numbers of participants (Chadwick, 2009; Shirky, 2011; Bekkers, Edwards, Moody and Beunders, 2011). This political content can be quite useful for government agencies, as it can significantly assist them to understand the needs and problems of society, and the perceptions and feelings of the citizens, and to formulate

effective public policies. The 'wicked nature' (Kunz and Rittel 1979; Conklin and Begeman 1989; Conclin 2003) and the continuously increasing complexity of social problems and needs makes this political content even more valuable, as it contains extensive knowledge of citizens concerning social problems and needs, which can be quite useful for understanding and managing their complexity.

For the above reasons a variety of ICT-based methods have been developed for the exploitation of this political content by government agencies (aiming at conducting various forms of 'citizensourcing'), initially simpler and later more sophisticated ones. In this evolution we can distinguish four discrete 'generations'. The first generation includes the creation of websites and social media accounts of government agencies, which provide information about existing and planned policies of the latter and allow citizens to create political content on them (e.g. comments, suggestions, etc.). As this citizens-generated political content becomes massive, it becomes difficult to process it manually, and this gives rise to the development of a second generation of ICT based methods, which automatically retrieve this political content from various sources using (e.g. various social media accounts and websites of the particular government agency) their APIs, and then process it in order to generate various kind of 'analytics' (Kokkinakos et al., 2012; Wandhöfer et al., 2012; Charalabidis and Loukis, 2012; Ferro et al., 2013). However, gradually it was realized that the most useful political content is generated beyond the websites and social media accounts, in numerous 'external' political forums, blogs, news websites, and also in various Twitter, Facebook, etc. accounts, without any stimulation from government; this gave rise to the development of a third generation of ICT based methods, which are oriented towards the automatic retrieval of this 'external' content using the APIs of the sources, and then its advanced processing (Bekkers et al, 2013; Loukis and Charalabidis, 2014; Charalabidis et al., 2014a). The above three first generations of ICT tools were oriented towards the political content generated by the general public. This can provide valuable insights into the perceptions of the general public concerning important social problems and existing or planned public policies, which are definitely quite important for the formulation of effective policies. However, in order to collect higher quality content concerning social problems and public policies it is necessary to target specific communities having strong interest and knowledge on them. This leads to the development of a fourth generation of such ICT based methods, which focus on the retrieval and processing of high quality content about social problems and public policies created by experts ('expertsourcing'); some first research in this direction is conducted in the European project EU-Community (partially funded by the 'ICT for Governance and Policy Modelling' research initiative of the European Union (EU) - see http://project.eucommuni-ty.eu/), which aims to exploit and leverage the extensive policy community of EU (Charalabidis et al., 2014b; Charalabidis et al., 2014c).

These ICT-based methods are increasingly based on the use of opinion mining (OM) and sentiment analysis (SA) techniques (Maragoudakis et al., 2011; Wandhöfer et al., 2013), in order to automatically process the extensive political content collected from numerous sources, and extract from it useful elements (e.g. main topics discussed, sentiments (positive or negative), etc.), which enable better insights into social

problems and needs, and provide substantial assistance for formulating public policies. However, the evolution of the ICT-based methods for the exploitation of the political content created in various web sites and social media accounts, creates new requirements with respect to OM/SA use, and necessitates the development of new ways of using OM/SA for meeting the particular needs and serving the objectives of each new method. This paper makes a contribution in this direction, by describing a novel approach to OM and SA use, created as part of the abovementioned fourth generation ICT-based method of exploiting political content created in the Internet, and especially in social media, by experts ('expertsourcing'), aiming to leverage the extensive policy community of the European Union. Furthermore, some first experimental results of it are presented.

In the following section 2 the EU-Community project is outlined. In section 3 the state of the art with respect to OM/SA use in similar project is discussed, together with the innovative features of the proposed OM/SA approach for the EU-Community project, which is described in section 4. Some first experimental results of it are presented in section 5, while the final section 6 summarizes the conclusions.

2　Context: the EU-Community project

As mentioned in the introduction, though the first three generations of ICT based methods for the exploitation of the political content created in the Internet, and especially in the Web 2.0 social media, were focusing on the general public, it was realized that in order to collect higher quality political content about the social problems and the public policies (existing and planned) it is necessary to adopt a more selective 'expertsourcing' oriented approach: to focus on content created by experts, who are highly knowledgeable on the specific topic/policy we are interested in and widely recognized. This is in line with previous research on policy networks (Skogstad, 2005; Rhodes, 2006 and 2007), which has concluded that due to the growing complexity and of social problems and policy making, and also their dynamic nature (frequent changes), governments cannot design and implement public policies based only on their own information and knowledge resources; for this reason they are increasingly making use of 'external' information and knowledge resources of various nonstate actors (initially economic actors and later other social and scientific actors as well), and cooperate with them, in order to design and implement effective public policies, and this has resulted in the generation of public policy networks.

The above lead to the development of a fourth generation of ICT based methods, which are focusing on the most knowledgeable and credible people on each topic/policy we are interested in. Such a method is developed in the European project EU-Community, which aims to exploit and leverage the extensive policy community of EU in order to support its policy making. An overview of this method is shown below in Figure 1. It consists of three main processes: the first two of them crawl at regular time intervals the most relevant external sources of EU policies knowledgeable and credible people, and also of relevant documents of various types, update the corresponding databases, and also process the retrieved data, and assess their reputa-

tion/credibility of the former and the relevance of the latter. These databases are then used by the third process, which processes users' queries (e.g. concerning the most reputable/credible people, or the most relevant documents on a specific topic, etc.) and presents the results, making use of visualization/visual analytics techniques. More information about this ICT based expertsourcing method are provided by Charalabidis et al. (2014b) and Charalabidis et al (2014c). The research presented in this paper concerns the second of the above processes (shown in the right part of Figure 1), and focuses on the advanced processing of the retrieved documents through an innovative use of OM/SA techniques.
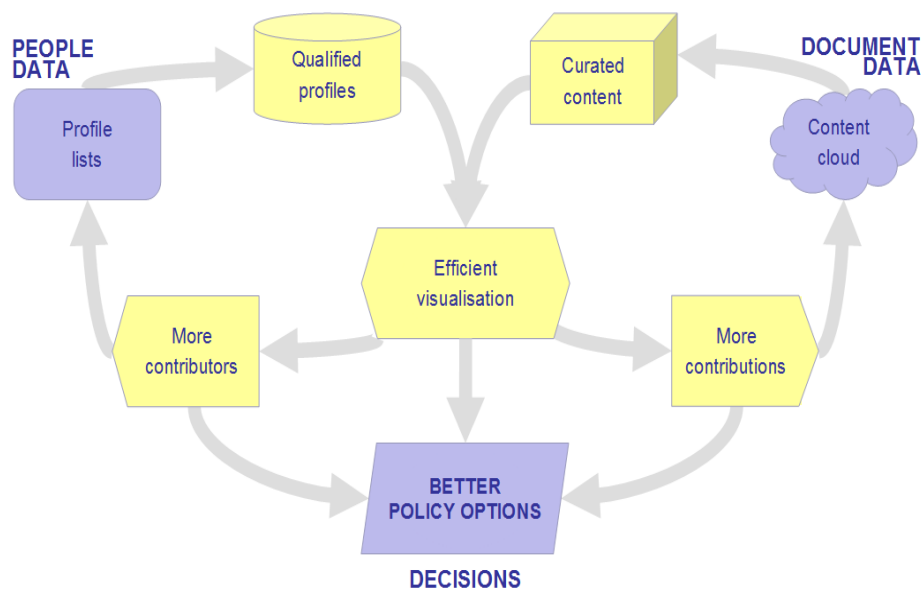

Figure 1: An overview of the ICT-based expertsourcing method developed in the EU-Community project

3   State of the Art

In order to design the OM/SA approach in the EU Community, we have studied the use of these technologies in related recent EU projects, which all aim at the exploitation of various types of political content created by citizens in the Internet, with main emphasis on the social media (in order to develop knowledge models, recommendations, intelligent services, etc.). A review of this OM/SA use approaches is initially provided in this section, and then are described the innovative features of the proposed OM/SA approach for the EU-Community project with respect to these recent related projects.

3.1 Related Projects

+spaces (2010-2012)

+Spaces ("Positive Spaces") was a research project research aiming at supporting the formulation of effective public policies by assessing the impacts of prospective policies, using political content created in various Facebook and Twitter, alongside virtual spaces (VS), like Open Wonderland. Modern VSs can be viewed as micro-societies, with dynamics resembling those of real world societies, the most evolved ones having virtual economies as well as regulations analogous to real-life legislative frameworks. Moreover, VSs are controlled environments in which all parameters of users' reactions and interactions can be tracked, so they can be very useful for assessing impacts of various policy options. In this project, information retrieval mechanisms and ON/SA were applied in order to collect data from VS and process them. Structured data from polls and petitions as well as unstructured data from VS blogs and debate logs were incorporated, together with relational information like social networks from user tracing.

E-policy (2011-2014)

The E-Policy ("Engineering the POlicy-making LIfe CYcle") project aimed at supporting policy makers for 'engineering' the policy making life-cycle, integrating both global and individual perspectives. Its objectives included assessment of social impacts through opinion mining on e-participation data from various thematic web sites and Web 2.0 platforms that allowed users to express their opinions on energy related topics through textual messages. In this project opinion mining identified social impacts that should be considered at both global and individual levels. At the global level, opinion mining aggregated individual opinions as trend line in order to conduct policy evaluation. Finally regression analysis was performed, in which text sentiment, estimated as a numeric score in the interval of $[-2; 2]$, with negative (positive) values indicating negative (positive) sentiment, was used as one of many independent variables in impact simulation modules.

Render (2010-2013)

The Render ("Reflecting Knowledge Diversity") project targeted at leveraging diversity (viewed as a valuable asset and crucial source of innovation and adaptability) in information management, in order to allow for better communication and collaboration. Under this objective, they addressed the problem of sentiment analysis in multiple domains and several languages, such as English, French, German, Italian, and Spanish. They exploited domain knowledge in the form of different sentiment lexicons, as well as the influence of various lexical surface features. Experimental results showed that the improvement resulting from using a two-layer model, sentiment lexicons, surface features and feature scaling is quite important, especially in social media textual datasets. Also, in this project, a tool was developed that performs sentiment classification and visualization of Twitter short texts, which enables the analysis and visualization of diversity in tweets.

Arcomem (2011-2014)

This project aims to enable memory institutions like archives, museums, and libraries to use and incorporate relevant social media content. A series of initial applications have been developed for opinion mining from social media using GATE, a freely available toolkit for language processing. Based on the work described in Maynard and Funk (2011), which focused on sentiments identification in tweets about political parties, their methodologies were extended to a more generic analysis of sentiment about any kind of entity or event mentioned, focusing on two specific domains: the current Greek financial crisis and the Rock am Ring rock festival in Germany in 2010. For both cases, first a basic sentiment analysis was performed, by associating a positive, negative or neutral sentiment to each relevant opinion target, together with a polarity score. As a next step, entity or event extraction was performed. A modified version of ANNIE, the default Named Entity (NE) recognition system in GATE, was used in order to identify mentions of persons, locations, organizations, dates, times and financial concepts. Sentiment analysis is performed by using a rule-based approach.

TrendMiner (2011-2014)

TrendMiner dealt with large-scale, cross-lingual trend mining and summarization of real-time social media streams. This project was very similar to the abovementioned project ArcoMem, involving more or less the same OM/SA methodology and tools. Again, the use of sentiment lexicons, special purpose gazetteers and rule-based approaches, under GATE platform, comprised the overall strategy for performing OM/SA on social media.

Padgets (2010-2012)

PADGETS ("Policy Gadgets Mashing Underlying Group Knowledge in Web 2.0 Media") focused on multilingual SA of citizens postings in government social media accounts, as a response to government policy campaigns and postings on specific topics/policies of interest. The main research objective of this project was to develop a methodology and a technological platform for the systematic and centrally managed exploitation of the emerging Web 2.0 social media by government organizations in their policy and decision making processes (Ferro et al., 2013). Citizens' postings (concerning opinions and comments) in government accounts in a variety of social media platforms, such as Twitter, Facebook, YouTube, Blogger, etc., were analyzed in order to identify citizens' sentiments. Since texts in social media tend to be very small, a machine learning approach was followed, in which limited linguistic resources were required. There was also a sentiment analysis module that incorporated sentiment lexicons, but this proved to perform significantly worse than machine learning models. Moreover, inclusion of emotional writing style attribute was taken into account in order to augment the performance of SA. From a technical point of view, feature selection was performed using a hybrid scheme of Support Vector Machines and Genetic algorithms. The system was implemented using RapidMiner®.

Nomad (2012-2015)

The NOMAD project ("Policy Formulation and Validation through non-moderated crowd sourcing") aimed at enabling government agencies to exploit the extensive

political content created in the social media beyond their own accounts, in multiple external sources (e.g. political blogs and forums, news websites, and various Twitter, Facebook, etc. accounts) (Loukis and Charalabidis, 2014). For this purpose there is extensive use of sophisticated OM/SA techniques, such as semantically driven textual data acquisition, sentiment analysis, thematic analysis, topics extraction and arguments extraction, and summarization. The main technology used was polarity lexica with some learning algorithms that exploit name entities.

3.2 Innovative Features of the OM/SA of the EU-Community Project

The EU-Community project, in order to meet its particular objectives and requirements, has developed a novel approach to OM/SA (described in the following section), which includes some interesting innovative features with respect to the above similar projects, shown below in Table 1.

Table 1. Innovative Features of the OM/SA Approach of the EU-Community Project

Other Projects	EU-Community	Explanation
OM/SA is mostly performed on short-sized texts from social media opinion channels.	OM/SA is functioning in a dual manner, both for short-sized opinions posted by users in various social media channels, and for larger documents, such as articles, positions and RSS feeds	While most existing projects focus on small texts, originating from social media platforms, our approach supports OM/SA in larger documents as well.
Are based on rule-based systems and polarity lexicons	It is using a hybrid model that combines polarity resources and machine learning	The proposed approach is utilizing both rule-based systems and classification rules extracted from data
Provide a single polarity estimation for a whole document	It can support a dual mode of both document and sentence-level OM/SA	Usually, the sentiment score is a single number that describes the whole document. Our approach can support either a single score, or a detailed sentiment score for each sentence.
Provide an OM/SA system that performs solely based on the sentiment of documents	It can support topic-based OM/SA applied on a collection of documents (= extraction of sentiment for each topic separately)	It allows extraction of sentiment for a document collection, based on the main topics are mentioned in them, for each topic separately. Users can focus on the documents posted by certain people, filter them according to their main topics, and see the corresponding sentiment of each.

4   An Opinion Mining and Sentiment Analysis Approach

The EU-Community project, aligned with the recent trends in OM/SA, is seeking to exploit key ICT ideas in order to leverage the huge amount of user-generated political content existing in multiple external sources, beyond government websites and social media accounts, focusing on the high quality content generated by experts, in order to support the formulation of public policies. In particular, the general idea of the project is that EU policy stakeholders need to be better informed on the most knowledgeable and credible people for each specific topic/policy they are interested in, and also the most relevant documents. The OM/SA module is mostly concerned with the analysis of the documents retrieved from various sources (either large ones such as positions, RSS Feeds, articles, etc., or small ones such as comments and posts in various social media sources).

This section describes the proposed OM/SA approach, which initially performs information extraction from crawled documents, aiming at conducting sentiment analysis on the document as well as on the sentence level, based on filtering according to a topic modelling procedure. The linguistic pipeline followed is shown in Figure 2.
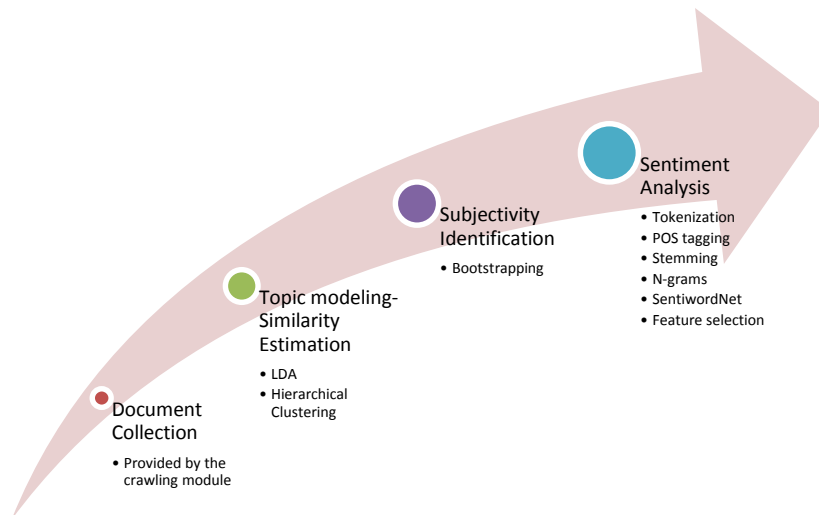
Sentiment Analysis
• Tokenization
• POS tagging
• Stemming
• N-grams
• SentiwordNet
• Feature selection

Subjectivity Identification
• Bootstrapping

Topic modeling-Similarity Estimation
• LDA
• Hierarchical Clustering

Document Collection
• Provided by the crawling module

Fig. 2. The linguistic pipeline of the proposed approach.

The first component activated is the crawler, which fetches data from various sources, such as RSS feeds, comments on the **Euractory** platform, Tweets, LinkedIn articles, etc. Since one of the main tasks of the EU-Community project is to identify relevant documents, we process the fetched documents using a topic-modeling procedure, which can identify the main "concepts" of a document. Topic modeling refers to a family of machine learning algorithms, most of them consisted of a probabilistic nature, which infer the hidden semantic structure within a set of input documents. The main idea behind topic modeling lies to the fact that documents are comprised of

some "concepts" (expressed through collections of terms). For the task at hand, we performed two different approaches. The former is the well-known Latent Dirichlet Allocation (LDA) method (Blei et al., 2003) and the latter is an agglomerative hierarchical clustering approach.

Upon grouping the documents according to their topics, the third module deployed was the subjectivity identifier, namely, a tool for discovering whether a document contains subjective or objective text. Work in opinion mining often adopts the hypothesis that the incoming documents are opinionated. However, in real-world situations there is a need to decide whether a given document contains subjective information or not, since the subjective sentences are of high importance for our purposes as they contain sentiment. The implementation of our subjectivity classifier is centered upon the methodology proposed by Riloff (2004). In fact, it is a bootstrapping approach that learns linguistically-rich extraction patterns for subjective expressions. The main principle is that high-precision classifiers, trained by a small dataset of labeled instances, perform labeling of a larger set of unannotated data in order to automatically create a large training set, which is then given to an extraction pattern learning algorithm. The learned patterns are then used to identify more subjective sentences. The bootstrapping process learns many subjective patterns and increases recall while maintaining high precision (recall measures the ability of the module to find all available subjective parts of a document, while precision measures the ability of identifying correctly parts as subjective).

The final step involves the core component, which performs sentiment analysis. In our approach, the subjective sentences are transformed into a vector of weights, according to the TF-IDF (Term Frequency-Inverse Document Frequency) model. Prior to this transformation, a series of pre-processing steps is undertaken, as shown in Figure 1. Tokenization includes the steps needed for removing any non-linguistic token and then taking care of contractions (e.g. didn't) and transforming all cases to lower for reducing the vocabulary size. Upon tokenization, the removal of non-informative, trivial tokens, called as "stopwords" was applied. Usually, these lists include tokens such as articles, personal pronouns, etc. In the present module, a custom list of stopwords was used as provided by RapidMiner®. Stemming is about keeping the core lexical form of a word from its various forms (e.g. *say*, *says*, *saying*, and *said* are all replaced by *say*). The next steps was about analyzing the part of speech of each token and keeping the most prone to contain sentiment such as adjectives, adverbs and verbs. In order to deal with text appearing in social media, which is characterized by idioms, abbreviations, emoticons, etc., a special gazetteer was tailored to identify them, and then replaced them by lexicons tokens (e.g. the emoticon ☺ was replaced by *happy*). The gazetteer also took punctuation marks into account and transformed them into tokens as well, since it is evident that there is a certain style when using such marks in social media texts (e.g. *!!!* was transformed as *frustration*).

Afterwards, a sliding window of *N* tokens was considered, in order to retain neighboring lexical content. *N*-grams are well known in computational linguistics and provide useful information about the context of a given term. For example, when we consider N-grams of N=2 (i.e. bigrams) we could capture sentiment information such

as "pretty stupid", while if we only consider single tokens then the word *pretty* would be assigned a positive sentiment and the word *stupid* a negative one. For our experiments, a value of *N*=2 and *N*=3 was respectively selected, with *N*=3 providing the best classification results. The next step utilized SentiWordNet, a sophisticated resource that encompassed semantic information (Miller, 1995). The motivation for this is based on the following idea: collection of affective words in a sentiment lexicon is tricky because such lexicons are limited to their domain and do not take into account the relation between words. Finally, due to the large amount of available features, generated by all the aforementioned pre-processing steps, we engaged a Support Vector Machine (SVM) classifier to weight each feature and eliminate those with lower weights.

5 Experimental Results

With respect to the data used in the experiments, initially we manually annotated some instances that were returned from the crawling component, upon removing non-English texts. As for evaluation, it was conducted for three different variations:

- Using Polarity Lexicon (PL) only.
- Using Machine Learning (ML) with all previous steps apart from the SentiWord-Net feature;
- The hybrid ML method, i.e. the combined Polarity Lexicons and ML approach (HML).

In order to calculate the performance, a method of 10-fold cross-validation using stratification (i.e. keeping the ratio of the class instances in the training and test sets constant) was followed. The performance was measured using precision, recall and accuracy, using three ML algorithms (Naïve Bayes, Support Vector Machines and Multi-Layer Perceptron Neural Networks), also comparing with a baseline classifier (i.e. a majority vote classifier, which always replies by selecting the most represented class in the dataset). The distribution of classes in the data was as follows: 103 negative, 614 neutral and 118 positive. The equations of each metric are provided below:

$$Recall = \frac{t_p}{t_p + f_n} \qquad (1) \qquad Precision = \frac{t_p}{t_p + f_p} \qquad (2)$$

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \qquad (3)$$

Here, **tp** are the correct positive predictions (true positives), **fp** are the incorrect ones (false positives), **tn** are the correct negative predictions (true negatives) and **fn** are the incorrect ones (false negatives). In other words, recall is the relative number of the correctly classified instances that were actually classified, precision is the relative number of correctly classified instances among all those classified and accuracy is the proportion of the instances of the testing set that were classified correctly against all instances.

In the Figures 3, 4 and 5 we can see the evaluation results for each metric (accuracy, precision and recall respectively). It is evident that the hybrid method HML of combining Polarity Lexicons with Machine Learning (Support Vector Machines) gives the best results.
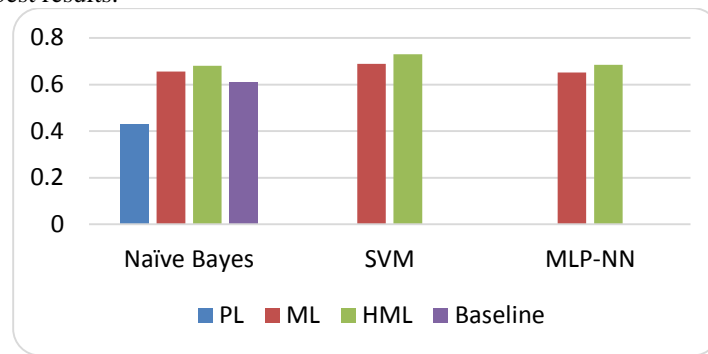

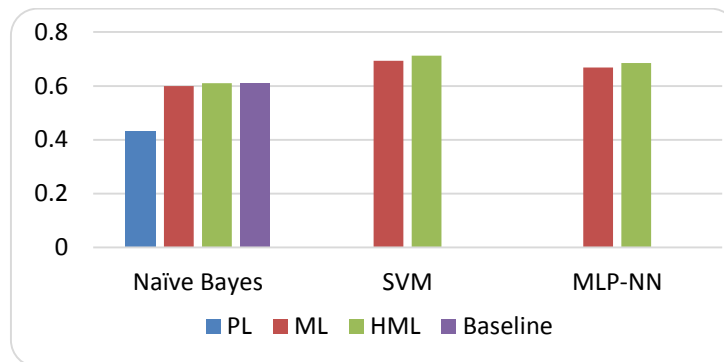**Fig. 3.** Evaluation results of the proposed approach - Accuracy

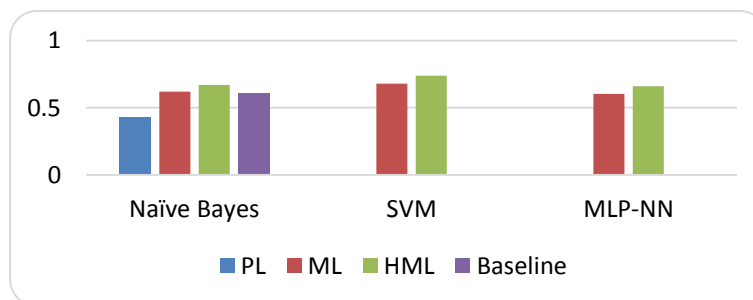**Fig. 4.** Evaluation results of the proposed approach - Precision

Fig. 5. Evaluation results of the proposed approach - Recall

5.1 Subjectivity Identification Results

In order to evaluate this task, a set of well-known datasets, also used by other relevant works, was adopted. Specifically, as a lexicon we used both the subjectivity lexicon considered in Rodrigues et al. (2008), as well as a slang lexicon created by our team. The subjectivity lexicon consisted of about 8.000 words, stored in a form that retains the following information: Term, Subjectivity Type (*strongsubj* or *weaksubj*), POS tag and Polarity as returned by SentiWordNet. Furthermore, for our subjectivity and polarity classification we used the datasets of IMDb movie reviews used also in Pang et al. (2002) and Wiegand et al. (2013). In the following Table 2, we can see the results (accuracy, precision and recall metrics). By observing the outcome, the proposed system can predict subjective sentences quite satisfactory (79.1%) however, it has some difficulties in identifying all of the existing subjective sentences in the dataset.

Table 2. Subjectivity Identification Results

Accuracy	Precision	Recall
66.3%	79.1%	58.6%

6   Conclusions

Governments are increasingly interested in exploiting the extensive political content created in various web sites and social media accounts, in order to obtain a better understanding of the needs and problems of society, and also the perceptions and feelings of the citizens, and to formulate effective public policies. This has led to the development of a variety of ICT based methods for this purpose, which are rapidly evolving and become increasingly sophisticated. The most critical technology for their effectiveness is the OM/SA. The evolution of these methods creates new requirements with respect to OM/SA use, and necessitates the development of new ways of using OM/SA for meeting the particular needs and serving the objectives of each new method.

In the previous sections of this paper has been described a novel approach to OM and SA use, as part of an advanced ICT-based method of exploiting political content created in the Internet, and especially in social media, by experts ('expertsourcing'). It includes interesting innovative features: it performs OM/SA both for short-sized opinions posted by users in various social media channels, and for larger documents, such as articles, positions and RSS feeds; it is using a hybrid model that combines polarity resources and machine learning; it can support a dual mode of both document and sentence-level OM/SA; and also it can support topic-based OM/SA applied on a collection of documents (= extraction of sentiment for each topic separately).

The first experimental results are encouraging. However, it requires further evaluation using more real life data (so that a better training of the classifiers can be made), which will identify strengths and weaknesses, and lead to improvements. This is already in progress, as part of the EU-Community project: pilot applications of the

whole ICT-based expertsourcing method developed in this project (Figure 1) have been planned for the near future, which will enable a better evaluation of the proposed OM/SA approach with more data. Also, further research is required on the combination of such fourth generation methods (oriented towards expertsourcing) with the ones of the previous generations (oriented towards the general public), so that knowledge, perceptions, opinions and ideas from the general public on one hand and from experts on the other hand can be combined for the formulation of better public policies.

7 References

1. Bekkers, V., Edwards, A.R., Moody, R., & Beunders, H. (2011). Caught by surprise? Micro-mobilization, new media and the management of strategic surprises. Public Management Review, 13(7), 1003–1021.
2. Bekkers, V., Edwards, A., & de Kool, D. (2013). Social media monitoring: Responsive governance in the shadow of surveillance? Government Information Quarterly, 30(4), 335-342.
3. Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In Proceedings of the 23rd international conference on Machine learning (ICML '06). ACM, New York, NY, USA, 113-120.
4. Chadwick, A. (2009). Web 2.0: New Challenges for the Study of E-Democracy in an Era of Informational Exuberance. I/S: A Journal of Law and Policy for the Information Society, 5(1), 9-41.
5. Charalabidis, Y. and Loukis, E. (2012). Participative Public Policy Making Through Multiple Social Media Platforms Utilization. International Journal of Electronic Government Research, 8 (3), 78-97.
6. Charalabidis, Y., Loukis, E., Androutsopoulou, A., Karkaletsis, V., Triantafillou, A. (2014b). Passive Crowdsourcing in Government Using Social Media. Transforming Government: People, Process and Policy, 8(2), 283-308.
7. Charalabidis, Y., Loukis, E., Koulizakis, Y., Mekkaoui, D., Ramfos, A. (2014b), 'Leveraging European Union Policy Community Through Advanced Exploitation of Social Media', IFIP Sixfth International Conference on e-Participation - ePart 2014, 1-3 September, 2014, Dublin, Ireland.
8. Charalabidis, Y., Loukis, E., Koulizakis, Y. (2014c), 'Social Media in Policy Making: the EU Community project approach', International Conference on e-Democracy and Open Government Asia 2014 (CeDEM Asia 2014), 4-5 December, 2014, Hong Kong.
9. Conklin, J. and Begeman, M. (1989). gIBIS: A tool for all reasons. Journal of the American Society for Information Science, 40 (3), 200–213.
10. Conklin, J. (2003). Dialog Mapping: Reflections on an Industrial Strength Case Study. In Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making (Kirschner, P., Buckingham Shum, P. and Carr C. Eds.), Springer Verlag, London, UK.
11. Ferro, E., Loukis, E., Charalabidis, Y., Osella, M. (2013). Policy Making 2.0: From Theory to Practice. Government Information Quarterly, 30(4), 359–368.
12. Kokkinakos, P., Koussouris, S., Panopoulos, D., Askounis, D., Ramfos, A., Georgousopoulos, C., et al. (2012). Citizens collaboration and co-creation in public service de-

livery: The COCKPIT project. International Journal of Electronic Government Research, 8(3), 44–62.

13. Kunz, W. and Rittel, H. (1979). Issues as Elements of Information Systems. Working Paper No. 131, University of California, Berkley.

14. Loukis, E., Charalabidis, Y. (2014), 'Active and Passive Crowdsourcing in Government', in Marijn Janssen, Maria Wimmer and Ameneh Deljoo (Eds) 'Policy practice and digital science: Integrating complex systems, social simulation and public administration in policy research', Spinger Verlag – Public Administration and Information Technology Series.

15. Maragoudakis, M., Loukis, E., Charalabidis, Y. (2011). A Review of Opinion Mining Methods for Analyzing Citizens' Contributions in Public Policy Debate. IFIP Third International Conference on e-Participation - ePart 2011, August 29 - September 1, 2011, Delft, The Netherlands.

16. Maynard, D. and Funk, A. (2011). Automatic detection of political opinions in tweets. In Dieter Fensel Raúl García-Castro and Grigoris Antoniou (eds), The Semantic Web: ESWC 2011 Selected Workshop Papers, Lecture Notes in Computer Science. Springer Verlag.

17. Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM 38, 11, 3941.

18. Pang, B. Lee, L. and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. Proceedings of EMNLP, pp. 79--86,.

19. Rhodes, R. A. W. (2006). Policy Network Analysis. In M. Moran, M. Rein and R. E. Goodin (Eds.) The Oxford Handbook of Public Policy. Oxford: Oxford University Press), pp. 423-445.

20. Rhodes, R. A. W. (2007). Understanding Governance: Ten Years On. Organization Studies, 28, 8, 1243–1264.

21. Riloff, E. (2004). Learning extraction patterns for subjective expressions, In Proceedings of the 2003 conference on Empirical methods in natural language processing (EMNLP '03). Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 105-112.

22. Rodrigues, P. P., J. Gama, and J. P. Pedroso (2008). Hierarchical clustering of time series data streams. IEEE Transactions on Knowledge and Data Engineering 20 (5), 615-627.

23. Shirky, C. (2011). The political power of social media. Foreign Affairs, 90, 28–41.

24. Skogstad, G. (2005). Policy Networks and Policy Communities: Conceptual Evolution and Governing Realities. Workshop on "Canada's Contribution to Comparative Theorizing" Annual Meeting of the Canadian Political Science Association, University of Western Ontario, London, Ontario.

25. Wandhöfer, T., Taylor, S., Alani, H., Joshi, S., Sizov, S., Walland, P., Mark Thamm (GESIS–Leibniz-Institute for the Social Sciences, Germany), Bleier, A. Mutschke, P. (2012). Engaging politicians with citizens on social networking sites: The WeGov toolbox. International Journal of Electronic Government Research, 8(3), 22–43.

26. Wandhöfer, T., Allen, B., Taylor, S., Walland, P., and Sizov, Z. (2013). Online Forums vs. Social Networks: Two Case Studies to Support eGovernment with Topic Opinion Analysis. IFIP 12th International Conference on e-Government - eGov 2013, September 16-19, 2013, Koblenz, Germany.

27. Wiegand, M., Klenner, M. and Klakow, D. (2013). Bootstrapping polarity classifiers with rule-based classification, Language Resources and Evaluation, pp. 1049-1088.