

Europe in the shadow of financial crisis: Policy Making via Stance Classification

Lefkothea
Spiliotopoulou
University of the
Aegean
lefkisp@aegean.gr

Dimitrios
Damopoulos
Stevens Institute of
Technology
ddamopou@stevens.edu

Yannis
Charalabidis
University of the
Aegean
yannisx@aegean.gr

Manolis
Maragoudakis
University of the
Aegean
mmarag@aegean.gr

Stefanos
Gritzalis
University of the
Aegean
sgritz@aegean.gr

Abstract

Since 2009, the European Union (EU) is phasing a multi-year financial crisis affecting the stability of its involved countries. Our goal is to gain useful insights on the societal impact of such a strong political issue through the exploitation of topic modeling and stance classification techniques. To perform this, we unravel public's stance towards this event and empower citizens' participation in the decision making process, taking policy's life cycle as a baseline. The paper introduces and evaluates a bilingual stance classification architecture, enabling a deeper understanding of how citizens' sentiment polarity changes based on the critical political decisions taken among European countries. Through three novel empirical studies, we aim to explore and answer whether stance classification can be used to: i) determine citizens' sentiment polarity for a series of political events by observing the diversity of opinion among European citizens, ii) predict political decisions outcome made by citizens such as a referendum call, iii) examine whether citizens' sentiments agree with governmental decisions during each stage of a policy life cycle.

With the proliferation of social media, governments focus on changing exorbitantly the way of communicating with citizens empowering their participation in the decision making process [6]. Governments focus on accelerating public sector regulations to reach society cost-effectively, avoiding any bureaucratic obstacles with stakeholders (i.e. citizens, public/private bodies), creating channels of offline participation and facilitating collaborative governance [7]. Their aim is to make a shift from e-government to we-government, through the exploitation of social media platforms [6] creating a new era of democratic involvement, transparency and accountability through political openness [1]. Extracting online information, analyzed and processed for sentiment classification can help us not only to initially understand how a certain community reacts to specific events but also to even try to predict their reactions to future events based on their behavior history. All these inferences are of key importance to policy makers that need to base their decisions on citizens' opinions and needs.

An event of great importance is the European debt crisis, that occurred due to the inability of Eurozone members, including Greece, to pay their governmental debt or bail-out over-indebted banks without the assistance of other Eurozone countries, or the European Central Bank (ECB), or the International Monetary Fund (IMF). European debt crisis has become a widely identified problem rising the speculation that other European countries will have the same outcome leading to a possible break-up of the Eurozone.

On 27 June 2015, under a “Grexit” threat, Greece’s Prime Minister announced the Greek Bailout Referendum, as a direct act of democracy, that took place on 5 July 2015, referring to whether Greece should accept bailout conditions offered by European Commission, IMF and ECB. Greek citizens would vote, either stating “Not approved/No” or “Approved/Yes” on two previous documents, entitled “Reforms for the Completion of the Current Program and Beyond” and “Preliminary Debt Sustainability Analysis”. The Referendum result proved that bailout conditions were rejected by a majority of over 61% to 39% approving. Although the result was negative, Greek government, requested for a three-year bailout from Eurozone’s rescue fund, reassuring to implement the needed measures and reforms. European finance leaders scheduled a “crisis summit” evaluating the Greek request

1. Introduction

Social media penetration has increased dramatically reshaping both structure and public discourse in society, reforming communities on a whole new level and resetting agendas in various topics ranging from social, religious to political issues [1].

These online interaction platforms, function as e-participation channels, crucially increasing the possibilities for users to express their opinions concerning any topic of discussion offering their personal position towards the topic or else their stance (i.e. positive/negative) [2].

A large part of these discussions refers to ideological dual-sided topics, considering political issues in which users' stance can take only two polarizing sides, specifically for or against [3]. Such topics, used for expressing and forming opinions, often stem heated discussions and attract large audience of people [4][5].

and a few days later, the package with the completed proposal was forwarded to the Eurogroup.

This extreme and unpredictable change in Greek government's political decision not to follow Referendum's result indicates that a series of multiple events affected its sentiment, altering the final political decision. Thus, the main topic of our research paper focuses on understanding the stance of online posts over a series of critical financial events that occurred in EU due to the Greek government–debt crisis. The period of time that we examine (26th June to 16th July), is considered unique in terms of the amount of critical political decisions taken among European countries and affecting the sentiment of two groups of online audience in Europe and Greece, respectively. From the day referendum was announced to the day the third memorandum was signed can be regarded as a policy life cycle that led to an enormous series of a daily sequential events. Each decision taken each day by one of the two groups, Europe and Greece, resulted to the creation of a new event in the next day producing a timeline of occurrences from which critical political events were identified.

1.1. Our contribution

Our goal is to gain a deeper understanding on whether citizens opinions are reconcilable with those of political figures, identifying public's opinion against governmental decisions, concerning policy formulation, promoting in that way the decision making process. The paper introduces a bilingual stance classification architecture that enables a deeper understanding of citizen's opinions and feelings for critical political decisions that affect their everyday lives. Additionally, examines if a new era of democratic involvement through political openness, social media and intelligent services can empower citizens' participation in the decision making process.

Summarizing the above, the contributions of this work consist of the following:

- We design and implement a new methodology for a bilingual stance classification architecture, concerning English and Greek language.
- Novel linguistic features were introduced, and a different approach was used to train the stance classification model.
- Our prototype is compared to models from similar research methods and is evaluated, in terms of performance, along two main axes: Accuracy and F-measure using real data.
- We performed 3 empirical studies based on real case scenarios, aiming to fully understand:
 - 1) Topic Stance Classification: determine citizens' sentiment polarity in a timeline of critical political

events, among European citizens, by indicating the diversity in the opinions.

- 2) Predicting the Greek Referendum Result: predict the outcome derived from citizens' political decisions, such as a referendum call and evaluate these predictions compared to the real results.
- 3) Policy Making: examine whether citizens' sentiments converge or diverge with governmental decisions made in each stage of a policy life cycle.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 analyzes the methodological approach of our research. Section 4 describes the overall architecture. Section 5 highlights the experiments and upcoming results on our stance classification system. In Section 6, we evaluate our stance classification system presenting the results of three empirical studies. Section 7 follows with the conclusions of our research and future work.

2. Related Work

In this section, we present previous works that are similar to ours, indicating the points in which our approach is more efficient than the others. Our work, specifically, consists of three main research pillars: topic modeling, sentiment analysis and stance classification. Thus, we offer an in-depth analysis of the related works in all three domains.

2.1. Topic Modeling

One part of our research is based on modeling online discussions in order to find and extract the topics of discussion. Topic modeling lies on analyzing large volumes of unlabeled text to create cluster of words that frequently occur together characterized by their distributional probability. LDA [8] [9], pLSA [10] and unsupervised pLSA [11], used for probabilistic latent semantic indexing, are considered as the most efficient choices when mining topics from large amounts of online context. However, there have been proposed numerous extensions of LDA, with Blei and Lafferty [12] proposing a correlated topic model (CTM), where topic proportions exhibited correlation via the logistic normal distribution, Titov and McDonald [13] developing a Multi-Grain topic model for extracting ratable aspects of objects in online reviews and Ramage et al. [14] proposing the Labeled LDA (LLDA), that constrains Latent Dirichlet Allocation by defining a one-to-one correspondence between LDAs latent topics and user tags. Moreover, Mukherjee and Liu [4] developed JTEP model, a semi-supervised generative model motivated by jointly occurred expression types, topics in online discussion posts and user pairwise interactions. Yuan et al. [15] developed a LightLDA,

enabling very large data sizes and models to be processed on a small compute cluster. Luo et al. [16] employed a Dirichlet Process Mixture Model (DPMM) to dynamically cluster detection responses into sets of objects and defined a semantic topic as co-occurring words encoding appearance and spatial information.

2.2. Sentiment analysis

Sentiment analysis determines positive or negative opinions expressed on topics of discussion [17] [18]. Main tasks include aspect extraction [19] [20], opinion polarity identification [21] and subjectivity analysis [22]. Sentiment Analysis is often used in researches that aim on the use of ICTs and social media platforms exploitation by government agencies for participative policy formulation utilizing sentiment analysis of online users' comments. Stylios et al. [7] examined 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. Sobkowicz et al. [23] proposed an opinion formation framework based on content analysis of social media and sociophysical system modeling. Charalabidis et al. [24] proposed an approach to sentiment analysis use aiming to leverage the extensive policy community of the European Union.

2.3. Classifying Stances

A relatively new challenging opinion mining task that needs to be further explored is stance classification: given a post written for a topic discussed online, we need to determine whether author's personal position towards the topic is either for or against.

Somasundaran and Wiebe [2] presented an unsupervised opinion analysis method for debate-side classification based on association rules and implementing Integer Linear Programming. Somasundaran and Wiebe [25] showed the benefit of modeling opinions and their targets at a fine-grained level using relational sentiment analysis techniques. Anand et al. [3] utilized meta-post features, contextual features, dependency features and word-based features from posts and parents posts to identify agreement and disagreement between the posts on an online debate website and that rebuttal posts are significantly harder to classify for stance. Walker et al. [5] classified posts in online debates focusing on the debate structure and the relations between speakers using MaxCut over rebuttal links between posts to separate them into opposite clusters. Ranade et al. [26] classified users' stance as pro or con examining users' intentions and debates structure. Hasan and Ng [27] use Conditional Random Fields (CRFs) to find opposite stances between sequences of posts. Boltuzic and Snajder

[28] and Ghosh et al. [29] studied various linguistic features to model stance and agreement interactions respectively. Sobhani et al. [30] proposed a minimally supervised model and merely a one-to-one mapping between the pre-defined argument set and the extracted topics.

3. Methodological Approach

Our aim is to define the top EU–Greek financial crisis topics discussed on social media, and determine through stance classification the citizens' sentiment (for–stance or against–stance) towards these topics. Due to the fact that the Greek's financial crisis is indissolubly linked with the economic policies of the EU partners, and indirectly affects all EU citizens, it is important to study both Greek's and EU citizens' feelings about the political and financial actions taken by their governments. Greek financial crisis, resulting in the Referendum incident, caused a series of upcoming parallel events affecting their economies in a large scale and creating the need to hear the sentiment pulse towards this event from audience inside and outside the Greek perimeter. That is the reason why we collected online posts with both English and Greek content.

3.1. System Evaluation

This section delivers the evaluation results stemmed from the proposed stance classification system. Initially, multiple machine learning classifiers were cross–evaluated through 10–fold validation in order to select the one with the greater percentage of Accuracy. From this preliminary study, we selected our classification engine.

3.2. Empirical Studies

Taking into consideration, the performance results derived from the evaluation phase, our proposed system is utilized in real case scenarios. To achieve this, we conduct 3 empirical studies focusing on how to employ a stance classification system in reality.

4. System Overview

This section describes the overall architecture of our proposed stance classification system, providing implementation details about the most important modules. Figure 1 depicts the overall system's overview of the 9 modules implemented to automatically collect, process and determine sentiment polarity.

4.1 Social Media Platforms Selection

Despite the rapid growth and popularity of social media platforms like Facebook and Twitter, in our research, we rely on online newspapers and weblogs to collect our data for analysis. Greek financial crisis resulting in the Referendum incident caused a sequence of daily events and political decisions that only online newspapers and weblogs report the news in a structured (*Title, article, discussion comments*), clear (built with structure CMS systems), and protected, by the freedom of communication and expression, way.

All newspapers, in a daily basis, publish articles related to the political and financial events occurred the same or previous day, allowing online users to express their personal opinion via discussion boards at the end of the articles. Utilizing *Social Media Platforms module*, we selected totally 20 online platforms based on their high popularity and broad user base measured by Alexa top news sites.

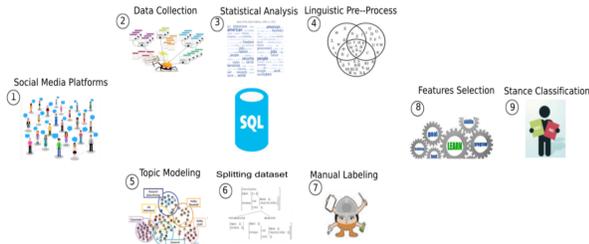


Figure 1. System Architecture

4.2. Data Collection

The data collection process started the day of the referendum call, 26th June of 2015, and lasted till the day the Greek government signed for a third memorandum, 16th July of 2015. We choose this period of time because it represents a full cycle of policy making by the Greek government and the EU, allowing us to track the feelings of the EU citizens, and also monitor all actions taken for the Greek crisis by the EU.

A dataset of 1734250 posts from 1129 topics was collected in a period of two summer months in 2015 utilizing the *Data Collection* module. During a preliminary study, we created two smaller datasets; the first one to train the stance classification model, and the second one to test our proposed system through three empirical studies.

To collect the data, it was important to monitor new topics and collect comments for both new and old topics for the whole period of time. Online platforms base their structure on Content Management Systems (CMFs) that facilitate the use of reusable components or customized software for managing web content. Data collection is performed by a customizable set of parsers developed in the Python programming language, and modified for each

online news website and weblog accordingly. The module is able to collect the *event title*, the *article*, and the *participants' posts, username*, and the *timestamp* (date and time) of each post written either in the Greek or the English language. To store the data, we utilize the Structured Query Language (SQL) Lite [31] database engine.

4.3. Statistical Analysis

Aim of the *Statistical Analysis* module is to distinguish the days when the number of posts and comments cumulatively in a high rate. We assume that a high number of post/comments indicate the occurrence of a dominant event that results in users to discuss it online and express their opinion. To automatically process the data a python script was build.

4.4. Linguistic pre-process

Next, a series of linguistic processes, follows with both grammatical and semantic analysis, creating n-grams, in our case both *uni-grams* and *bi-grams*. N-gram [17] is a contiguous sequence of n items (letters, words or base pairs), from a given sequence of text or speech. Here the goal is twofold; i) use the n-grams and via Topic modeling to identify the topics in each post, and ii) use them as features for stance classification. We are interested in building uni-grams and bi-grams by utilizing words that are nouns, adjectives and verbs. These words are considered opinionated words and can be used later as additional features.

Developing the *Linguistic Pre-process module*, we start with tokenization, splitting each posts' sentence into words. We continue with stemming, finding the root of each word and with part-of-speech (POS) tagging [17], marking up each word in the corpus as corresponding to a particular part of speech, based on both its definition and context. Last but not least, we end with n-grams generation, specifically uni-grams and bi-grams. For tokenization, sentence splitting, (POS) tagging and n-gram generation, we installed components of Natural Language Toolkit [32]. For stemming, we utilized and imported in our program Porter Stemming Algorithm [33]. The algorithms, implemented for textual analysis, are language dependent. This means that we had to create two different tools, that follow the same methodological approach, one for the Greek and one for the English language.

4.5. Topic Modeling

As described in the Social Media Platforms Selection subsection, we collect the comments from articles published in online newspapers and weblogs. Although

each article focuses on a specific event on that day, users tend to discuss in their comments related topics too (e.g., in an article related to the financial crisis, users may discuss also about topics like the education or health). Thus, the utilization of *Topic Modeling module*, is a crucial part in our research aiming to identify all the topics of each sentence in each post, and keep only those related to the topic we are analyzing each time.

Due to the bilingual comments, we decided to use two different approaches for topic modeling, one for English and a different one for the Greek language. Both methodologies are well evaluated in the recently literature providing the best results.

To analyze the English content and reveal the hidden thematic structures inside the post, we utilize Mallet [34], a tool for modeling our datasets and extracting the topics of discussion. Mallet used a generative statistical model called Latent Dirichlet Allocation (LDA) [12]. LDA allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. We trained the model under a set of commands preserving the document as a sequence of word features, converting all words to lowercase and removing stop-words. As an output of this modeling process, we obtain the most probable topics and the most probable words, called top-words, that appear with the highest frequency across posts.

In order to extract the topics being discussed using the Greek language, we developed a python script that relies on the Gliozzo et al. [35] study and the uni-grams of each post's sentence, specifically to those that contain adjectives and nouns. The study of Gliozzo et al. [35] point that posts containing adjectives have high probability of indicating implicit user opinions as opposed to posts that contain no adjectives at all, and topics are most likely to appear in a post in the form of a noun. Having the uni-grams and bi-grams from the previous stage, we employed a syntactic dependency parser to identify which adjectives refer to which nouns across the posts, making adjective-noun pairs that serve as bi-grams and then we counted their frequency. Finally, we selected all those n-grams with the highest appearance and we considered these as our topics and top-words.

4.6. Features selection

Aiming to build an automated tool that utilizes machine learning classifiers to determine the stance of a sentence, as for or against, it is important to evaluate and select correctly a set of linguistic features. As our baselines, we use unigrams and a set of three lexico-syntactic features proposed by Anand et al. [3].

Following Anand et al. [3] methodology, *Features Selection module*, retrieving data from online posts, is based on the composition of lexico-syntactic features:

basic, sentiment and argument. As basic features, we utilize the number of words and sentences in a post; posts' length; cue words representing posts' initial uni-gram and bigram sequence and repeated punctuation (e.g. !! or ??) normalized by the number of uni-grams in a post. As sentiment features, we employ pronominal forms, positive and negative emotion words extracted in the English comments via the Linguistics Inquiry Word Count (LIWC) tool [36] and in the Greek comments utilizing Greek Sentiment Lexicon by Tsakalidis et al. [37]. As argument features, for the English comments, we exploit repeated punctuation (e.g. !! or ??) normalized by the number of uni-grams in the post, POS generalized dependencies and opinion dependencies using MPQA Dictionary [38] of emotion words, and syntactic dependencies using the Stanford Parser [39]. For the Greek comments, we extract the exact same features utilizing our own syntactic dependency parser.

4.7. Splitting dataset

Although selecting the proper features is considered the key element in designing a modern machine learning system, the utilization of the right data to build the classification model is the proper way to achieve high Accuracy and predict correctly the stance. During a preliminary study, through the *Splitting Dataset module*, we test various combinations and percentages in splitting the dataset, before concluding that the best way to create the training dataset is by learning from the 20% of the daily topics that contain the top words. In this way, the classification model contains instances that appear in most topics.

4.8. Manual Labeling

We rely on manually annotation to label the training dataset. Hence, with *Manual Labeling module*, we label each post's stance towards a topic as a for-stance or against- stance, removing sentences that are objective, which contain no sentiment towards any topic. Furthermore, having in our possession the top-words that appear across all posts, we label them determining their sentiment polarity as positive or negative, and we create an additional feature.

Creating training instances by employing the two feature sets as well as its manually annotated stance as its class label, we train the stance classifiers determining the post's stance.

4.9. Stance Classification

The classification engine is considered the most important part of stance classification system. To choose the right classifier as our stance classification engine, we

conducted multiple experiments and cross-evaluated various algorithms. At the end, we selected the classifier with the highest performance. To build our *Stance Classification module*, we utilize the Weka library [40] that includes a collection of machine learning algorithms for data and opinion mining tasks such as classification, and the Random Forest classifier as our engine. We choose Random Forest, please refer to System Evaluation Section 5, due to its high Accuracy in automatically classifying the stance in online comments.

5. System evaluation

This section provides results derived from the evaluation of the proposed stance classification system. Figure 1 illustrates a high-level overview of the proposed architecture composed of 8 main components that follow our proposed methodology.

For the needs of this paper, various machine learning classifiers have been utilized during a small scale preliminary study. Specifically, we cross-evaluate four supervised machine learning algorithms, i.e., Bayesian Networks, Radial Basis function (RBF), K-Nearest Neighbor (KNN) and Random Forest [41]. The data analysis was carried out using Weka. Additionally, a 10-fold cross validation technique was used to assess how the results will generalize to an independent dataset. As a dataset for the preliminary study, we picked 5 random but sequential days. The analysis of the data has been performed on a laptop machine with a 2.53 GHz Intel Core 2 Duo T7200 CPU and 8 GB of RAM. The OS of this machine is OS X El Capitan.

We evaluate the performance of our stance classification system in terms of two metrics: i) the Accuracy, and ii) the F-measure. Accuracy denotes the correct instances that are classified [42]. F-measure [59] is the harmonic mean of two other metrics, precision and recall, whereas precision [42] indicates how well a classifier categorizes instances correctly and recall [42] measures the fraction of relevant instances correctly retrieved from all possible instances.

Table 1, presents the performance of our proposed stance classification system (see the numbers in red annotation) in comparison to the performance of the other research works mentioned in the related work Section 2, referring to stance classification. This comparison is based on the same evaluation metrics, specifically Accuracy and F-measure and these performance results, derived from each system's cross-evaluation, both ours and from the recent literature review, are also presented in the table. We note that Random Forest is the most promising method showing optimal results of 82.7% Accuracy and 79.3% F-measure.

We show that our system's performance exceeds significantly both the baseline and the similar researches'

approach with promising results. As a general remark, all the experiments present highly accurate results, thus providing strong evidence that designing a bilingual stance classification system can be a very precise way of analyzing big data.

Table 1. Comparison of Proposed Approach with previous work in terms of Accuracy and F-measure.

Model	Baseline%		Approach%	
	Accuracy	F1 Score	Accuracy	F1 Score
Somasudaran & Wiebe [2]	53.2	56.4	66.1	66.1
Somasudaran & Wiebe [25]	66.6	-	70.5	-
Anand et al. [3]	56.4	-	67.9	-
Walker et al. [5]	71	46	88	80
Hasan & NG [27]	66.9	56.6	75.4	61.1
Ranade et al. [26]	64.2	-	74.3	-
Boltuzic & Snajder [28]	-	77.9	-	81.6
Ghosh et al. [29]	-	59.9	-	66.9
Sobhani et al. [30]	-	41	-	77
Proposed Model	62.4	51.6	82.7	79.3

6. Empirical studies

Having achieved high performance results, our proposed system can be used in real case scenarios. We conducted 3 empirical studies on how to utilize a stance classification system in real life.

In the first scenario, our goal is to determine whether citizens' sentiment polarity in Europe and Greece may converge or diverge in a series of events occurred under the umbrella of a single political topic.

In the second scenario, we aim to explore whether a stance classification system can replace traditional mechanisms of extracting citizen's opinion towards a political event, such as gallups and online polls.

Taking policy's life cycle as a baseline, in our third scenario, we examine how public's sentiment polarity changes in all stages of a policy according to the political decision taken in each stage.

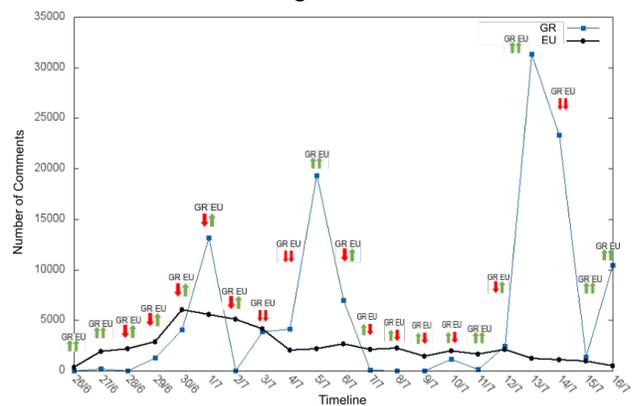


Figure 2. Stance classification

The following list presents the events utilized in the 3 real empirical studies. The starting day of the timeline corresponds to when the Greek bailout referendum was announced, while the last event to when Greek

Government signed the Third Memorandum. We selected these days because they are considered as the starting and ending point of a full policy life cycle.

26/6	Greek Referendum Call
27/6	Eurogroup declares that the crisis has commenced
28/6	Pause of emergency support to Greek banks by European Central Bank
29/6	Capital controls begin
30/6	Greek Prime Minister asks from Greeks to vote "NO" in Bailout Referendum
1/7	Europe prepares for a Grexit
2/7	Cash decrease in Greek banks
3/7	Capital controls leave Greece with shortages in multiple sectors
4/7	Europe claims Greek Government is worsening the crisis
5/7	Greece voted "NO" in Bailout Referendum
6/7	European Central Bank keeps Greek Banks Emergency Liquidity Assistance frozen
7/7	European Commission considers bridge program for Greece
8/7	Greek Bailout solution with a Third Memorandum proposed in EuroSummit or Grexit
9/7	Greek Government suggests Bailout proposals
10/7	Greek Prime Minister implores Syriza party to accept proposed reforms
11/7	Issue of trust between European creditors and Greek Government
12/7	German plan demands e50bn of state assets to be transferred to external fund
13/7	Greece gets Bailout deal in EU Summit
14/7	German financial minister discusses with European ministers for parallel currency in Greece
15/7	Greek Parliament votes for the Third Memorandum
16/7	Greek Parliament voted "YES" and Greek Government signs the Third Memorandum

6.1. Stance classification

In this empirical study, we aim to determine whether sentiment polarity of European citizens remains the same or differs on each event in the timeline of these political occurrences. Due to the fact that each group, Greek and Europe (EU governments, European Commission, IMF and ECB), corresponds to different decisions made by the politicians, sentiment polarity of each group, may be affected differently.

We perform topic stance classification, labeling each post as for or against towards the topic being discussed in the specific post. Having all stance classification results, we determine the final stance over each topic, in the timeline of events, formulating the sentiment polarity for both Europe and Greek online audience.

Hence, in Figure 2, we depict stance polarity with for-stances colored in green and against-stances colored in red, GR for Greece and EU for Europe. It is evident that in some events sentiment polarity for both groups remains the same when in others it changes orientation. Specifically, starting with the day when referendum was announced, both groups have positive feeling towards this political call. This is likely to have happened due to the fact that both groups believe in the existence of

democracy and consider that in such important political decisions affecting a country's future, it is the citizens that need to make the final decision. As days come closer to the voting day, we observe, that on 1st July, maybe Europe considers Grexit as an option on the table of negotiations and sentiment orientation changes. Europe continues being positive, unlike Greece's sentiment that turns into negative, most likely because capital controls and money shortage in Greek banks start affecting their decision towards the referendum vote.

We again notice a fluctuation of opinions but in Europe's side. This polarity alternation is maybe based on the belief that if Greece could not yet pay the loans already taken from IMF, then she would not be able, to pay additional loans, remaining in a constant financial debt. Thus, a possible Grexit may not have been a wrong decision. Nevertheless, a change in opinion orientation is formulated, transforming the sentiment in both groups to positive on 5th July, with Greeks voting "NO" to an agreement plan on the one side and Europeans being positive to what Greeks would decide on the other side. This specific day, as illustrated in our figure, is considered as a critical event due to the huge number of comments posted online, hitting approximately 20000 in our sample.

Another opinion fluctuation occurred from 7th to 10th July where both groups are assigned with opposite polarity. In this timeframe, although Greeks voted negatively, Greek government, under Grexit threat, starts a series of negotiations with EU Summit suggesting bailout proposals. This political action had most likely a negative effect on Europeans who considered that Greek government had not taken into consideration citizens' vote, creating a damaging attitude towards Greek government. Till the last day of our timeline, we observe that sentiment orientation in Greeks' side remains positive, with Greek government finally signing the third memorandum even with more austerity measures than the one proposed in the beginning of the timeline, indicating her urge to remain in the European Union and a possible Grexit to be avoided. Europeans on the other side, although their sentiment was negative due to Greek's government credibility issues, it finally altered into positive at the end, due to reassurance of employing all measures signed in the third memorandum offered by the Greek government.

At the end of this study, we were able to classify the stance for two groups of people towards each topic of discussion. It is very important to realize that their sentiment polarity was affected from each political decision taken in both European and National level and that change was very clear the days when critical events had taken place. We should not forget, though, that all events in the timeline, refer to a single central political topic of interest, the EU financial crisis, which is the topic of our research.

6.2. Predicting the Greek Referendum results

Greek bailout referendum, considered as the most direct act of democracy, is a crucial issue that not only affects Greece but Europe as a whole, making it inevitable to predict the outcome. Gallups and online polls are two traditional mechanisms introduced, in the past decades, in the dawn of a political critical event such as elections, referendums, etc [43]. These methods use analytics in order to extract citizens' opinion.

In this empirical study, we aim to figure if such traditional mechanisms still provide accurate predictions or new ways like sentiment classification can be used as potential tools.

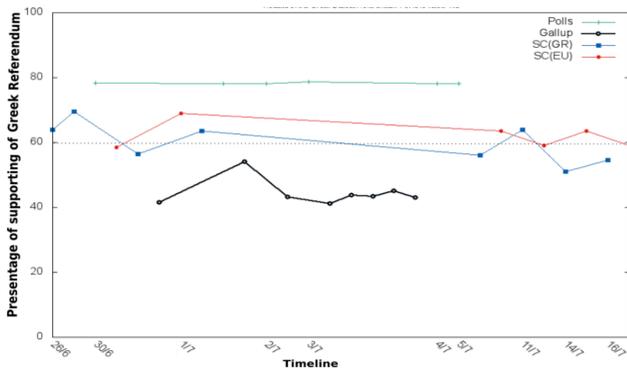


Figure 3. Predicting the Greek Referendum Result

The referendum results, indicating that 61.31% of Greek people voted “No”, are illustrated in Figure 3 with the black dotted line. The predictions derived from the online polls are depicted with the green line and the ones from gallups are highlighted with the black line. The online polls were initiated on 30th June, when Greek Prime Minister asked from citizens to vote “NO” in the bailout referendum and ended on 5th July with Greeks voting “NO” in the referendum. From our figure, we can observe that, from the first to the last day online polls were conducted, the percentage of “NO” vote was moving in the same range, hitting a spectacular rate of approximately 80%. Comparing online polls' predictions to the real referendum result, it is surprisingly unexpected that there is such a large deviation. The same happens also with the gallup's predictions but in reverse. More specifically, although online polls predicted correctly that “NO” vote would be the real outcome of the referendum, the percentage of “NO” was actually 20% higher than the real one. With the gallups, we have the reverse outcome. This means that, although gallups predicted also correctly that “NO” vote would win, the percentage of “NO” was approximately 20% lower than the referendum's real result, but still winning the “YES” vote.

It is evident that the predictions made from both traditional methods offered the correct result, but the aberration in the percentage rate was too large and this is an astonishingly uncommon phenomenon.

Focusing on our system's results, which are illustrated for Greece with blue line and Europe with red line, we notice that the rate of the predicted stance result is very close to reality, giving a percentage of 57% for Greece and 65% for Europe in the referendum's day, respectively. At some points, as shown in the figure, the blue and red line touch the black dotted line having a total match of our prediction with the real result.

This occurrence offers us the ability to believe that a stance classification system can perform greater than traditional mechanisms in predicting political events and potentially replace them. Hence, maybe stance classification can be viewed as an e-government tool promoting decision making and empowering citizens in the policy making formulation.

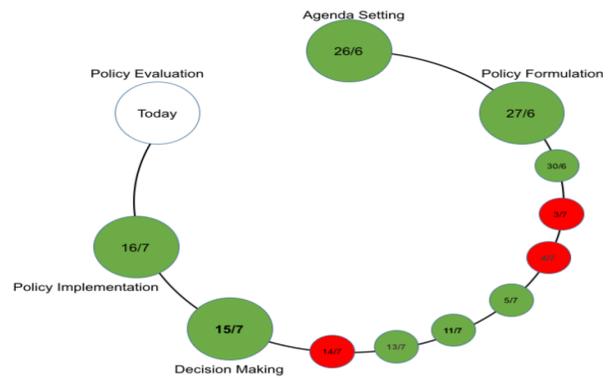


Figure 4. Policy Making Life Cycle via Stance Classification

6.3. Policy Making

Policy, as a product of a political process, can be viewed as a sealed black box. In politics, policy refers to the basic principles by which a government is guided.

A policy model can be treated as a cycle of various discrete stages, each one comprehended as a coherent chain of events with given context according to their chronological occurrence [44]. These events can be rationally related to each other and predictions can be made based on their sequential appearance. Taking policy's life cycle as a baseline, we make the following consideration. We regard the sequence of events, occurred in our timeline, from the day Greek referendum was announced to the day the third memorandum was signed by Greek government, as a policy. Our goal is to present how public's sentiment polarity changes during the different stages of policy life cycle, when a different political decision is taken at each policy level.

In Figure 4, we present how sentiment alters at each policy stage in the timeline of critical political events,

concerning the two groups of people in Europe and Greece. Having already performed stance classification, we indicate in which critical events the sentiment polarity of both groups remains the same or differs. If sentiment in both groups has the same polarity then the critical event is illustrated in green color, otherwise in red.

Thus, mirroring the fluctuation of opinions, the first phase of agenda-setting starts on 26th June with the announcement of the Greek bailout referendum. It is colored in green acquiring a for-stance for both Europeans and Greeks. The phase of policy formulation includes a series of sequential critical events that take place from 27th June till 15th July when the next phase begins. At this period of time, Greek government initiated a series of negotiations with the European Commission, IMF and ECB suggesting bailout proposals even though Greek citizens have voted negatively in the referendum call. As we can observe from our figure, each critical event, in this timeline, is determined by a specific sentiment polarity colored either in green with both groups of people sharing the same sentiment or in red in which the sentiment differs. Then, arriving at the stage of decision-making, with the Greek government asking from the parliament to vote positively in signing the third memorandum, sentiment polarity in both groups is positive. Finally, policy is implemented with the Greek government signing the third memorandum and assuring European creditors to follow measures and reforms proposed.

From this empirical study, we were able to indicate that the sentiment may has a certain polarity at the initiation of a policy but till the last phase of a policy, the political decision is constantly evolving, affecting dynamically the original sentiment causing a possible alternation in its orientation. Thus, we consider that a dynamic system should be developed based on the synergy of citizens and politicians bridging the gap between the two groups and empowering citizens in the decision-making process.

This online platform utilizing stance classification and policy making methods would offer to politicians the ability to acquire feedback from citizens' opinions and make a decision based on these opinions. If politicians were willing to use this mechanism effectively, then a policy would need less time to be formulated and the final decisions would be closer to citizens' will.

7. Conclusions and future work

In this paper, we focus on classifying the stance of online posts by determining the sentiment polarity of a popular political event, specifically, the EU – Greek financial crisis concerning the Greek Bailout Referendum call. The paper introduces a unified bilingual stance classification architecture with quite promising results, able to analyze the citizen's opinions and determine their

feeling towards various critical political topics. This was achieved by examining whether the sentiment for two groups of online audience, in Europe and Greece, had the same orientation or changed in a timeline of critical events. Finally, we relied on stance classification to validate the hypothesis that citizens' participation, through social media platforms, can efficiently contribute in the government's decision making process, utilizing policy's life cycle as baseline.

As future work, it would be interesting to explore additional new features in order to boost the performance of the classifier even further, considering features that contain the feeling of tolerance or irony. Moreover, we could explore more sophisticated machine learning algorithms trained in our augmented set of linguistic features, measuring their performance with new statistical metrics. Last but not least, we aim to analyze similar political events like the very recent Brexit.

8. References

- [1] J. Freeman and S. Quirke, "Understanding e-democracy government-led initiatives for democratic reform," *JeDEM- eJournal of eDemocracy and Open Government*, vol. 5, no. 2, pp. 141–154, 2013.
- [2] S. Somasundaran and J. Wiebe, "Recognizing stances in online debates," in *ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 2-7, Singapore, 2009, pp. 226–234.
- [3] P. Anand, M. Walker, R. Abbott, J. E. F. Tree, R. Bowmani, and M. Minor, "Cats rule and dogs drool!: Classifying stance in online debate," in *Proc. of the 2nd WASSA '11. ACL*, 2011.
- [4] A. Mukherjee and B. Liu, "Discovering user interactions in ideological discussions," in *Proc. of the 51st Annual Meeting of the ACL 2013*, pp. 671–681.
- [5] M. A. Walker, P. Anand, R. Abbott, and R. Grant, "Stance classification using dialogic properties of persuasion," in *Proc. of the 2012 Conference of the NAACL HLT '12. ACL 2012*, pp. 592– 596.
- [6] D. Linders, "From e-government to we-government: Defining a typology for citizen coproduction in the age of social media," *Government Information Quarterly*, vol. 29, no. 4, pp. 446 – 454, 2012.
- [7] G. Stylios, D. Tsohis, and D. Christodoulakis, *Mining and Estimating Users' Opinion Strength in Forum Texts Regarding Governmental Decisions*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 451– 459.
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
- [9] E. Graells-Garrido, M. Lalmas, and R. A. Baeza-Yates, ser. CEUR Workshop Proceedings, C. Trattner, D. Parra, P.

- Brusilovsky, and L. B. Marinho, Eds., vol. 1421. CEUR-WS.org, 2015, pp. 10–16.
- [10] T. Hofmann, “Probabilistic latent semantic indexing,” in *Proceedings of the 22Nd Annual International ACM SIGIR*. ACM, 1999, pp. 50–57.
- [11] T Hofmann, “Unsupervised learning by probabilistic latent semantic analysis,” *Mach. Learn.*, vol. 42, no. 1-2, pp. 177–196, Jan. 2001.
- [12] D. M. Blei and J. D. Lafferty, “Dynamic topic models,” in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML ’06. ACM, 2006, pp. 113–120.
- [13] I.Titov and R.T.McDonald, “Modeling online reviews with multi-grain topic models,” in *Proceedings of the 17th International Conference on WWW*, 2008, Beijing, China, April 21-25, 2008, 2008, pp. 111–120.
- [14] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, “Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora,” in *EMNLP*, 2009.
- [15] J. Yuan, F. Gao, Q. Ho, W. Dai, J. Wei, X. Zheng, E. P. Xing, T.-Y. Liu, and W.-Y. Ma, “Lightlda: Big topic models on modest computer clusters,” in *Proc. of the 24th International Conference on WWW ’15. USA*: ACM, 2015, pp. 1351–1361.
- [16] W. Luo, B. Stenger, X. Zhao, and T.-K. Kim, “Automatic topic discovery for multi-object tracking,” in *Proc. of the Association for the Advancement of Artificial Intelligence (AAAI)*, 2015.
- [17] B. Liu, *Sentiment Analysis and Opinion Mining*, ser. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
- [18] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 12, pp. 1–135, 2008.
- [19] A.-M. Popescu and O. Etzioni, “Extracting product features and opinions from reviews,” in *Proceedings of the Conference on HLT ’05. ACL 2005*, pp. 339–346.
- [20] A. Mukherjee and B. Liu, “Aspect extraction through semi-supervised modeling,” in *Proc. of the ACL 2012*, pp. 339–348.
- [21] A. Hassan, V. Qazvinian, and D. Radev, “What’s with the attitude?: Identifying sentences with attitude in online discussions,” in *Proc. of the 2010 Conference on Empirical Methods in Natural Language. ACL, 2010*, pp. 1245–1255.
- [22] V. Hatzivassiloglou and J. M. Wiebe, “Effects of adjective orientation and gradability on sentence subjectivity,” in *Proc. of the 18th Conference on ACL 2000*.
- [23] P. Sobkowicz, M. Kaschesky, and G. Bouchard, “Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web,” *Government Information Quarterly*, vol. 29, no. 4, pp. 470–479, 2012.
- [24] Y. Charalabidis, M. Maragoudakis, and E. Loukis, “Opinion mining and sentiment analysis in policy formulation initiatives: The eu-community approach,” in *Electronic Participation - 7th IFIP, ePart 2015, Proc. 2015*, pp. 147–160.
- [25] S. Somasundaran and J. Wiebe, “Recognizing stances in ideological on-line debates,” in *Proc. of the NAACL HLT 2010 Workshop. ACL, 2010*, pp. 116–124.
- [26] S. Ranade, R. Sangal, and R. Mamidi, “Stance classification in online debates by recognizing users’ intentions,” in *Proc. of the SIG-DIAL 2013 Conference. ACL 2013*.
- [27] K. S. Hasan and V. Ng, “Extra-linguistic constraints on stance recognition in ideological debates,” in *Proc. of the 51st Annual Meeting of the Association for ACL 2013*.
- [28] F. Boltuzic and J. Snajder, “Backup your stance: Recognizing arguments in online discussions,” in *Proc. of the 1st Workshop on Argumentation Mining. Maryland: ACL 2014*.
- [29] D. Ghosh, S. Muresan, N. Wacholder, M. Aakhus, and M. Mitsui, “Analyzing argumentative discourse units in online interactions,” in *Proc. of the 1st Workshop on Argumentation Mining. Maryland: ACL 2014*, pp. 39–48.
- [30] P. Sobhani, D. Inkpen, and S. Matwin, “From argumentation mining to stance classification,” *NAACL HLT 2015*, p. 67, 2015.
- [31] Sql lite,” 2015. Online: <https://www.sqlite.org/lang.html>
- [32] “Natural language toolkit,” 2015. Online: <http://www.nltk>.
- [33] “The porter stemming algorithm,” 2015. Online: <http://tartarus.org/martin/PorterStemmer/>
- [34] “Mallet,” 2015. Online: <http://mallet.cs.umass.edu/>
- [35] A. Gliozzo, C. Strapparava, and I. Dagan, “Unsupervised and supervised exploitation of semantic domains in lexical disambiguation,” *Computer Speech and Language*, vol. 18, no. 3, pp. 275 – 299, 2004.
- [36] “Linguistic inquiry and word count,” 2015. Online: <http://liwc.wpengine.com/>
- [37] A. Tsakalidis, S. Papadopoulos, and I. Kompatsiaris, *An Ensemble Model for Cross-Domain Polarity Classification on Twitter*. Springer International Publishing, 2014, pp. 168–177.
- [38] “Mpqa dictionary,” 2015. Online: <http://mpqa.cs.pitt.edu/>
- [39] “The stanford language processing group,” 2015. Online: <http://nlp.stanford.edu/software/lex-parser.shtml>
- [40] “Weka the university of waikato,” 2015. Online: <http://www.cs.waikato.ac.nz/ml/weka/>
- [41] D. Damopoulos, G. Kambourakis, S. Gritzalis, and S. O. Park, “Exposing mobile malware from the inside (or what is your mobile app really doing?),” *Peer-to-Peer Networking and Applications*, vol. 7, no. 4, pp. 687–697, 2014.
- [42] G. Forman, “An extensive empirical study of feature selection metrics for text classification,” *J. Mach. Learn. Res.*, vol. 3, pp. 1289–1305, Mar. 2003.
- [43] C. Holtz-Bacha, *Polls, Media and the Political System*. UK, 2012, pp. 267–280.
- [44] A. Macintosh, “Characterizing e-participation in policy-making,” in *Proc. of the 37th HICSS’04 - Track 5 - Volume 5, USA: IEEE Computer Society, 2004*.