A Novel Concept-level Approach for Ultra-concise Opinion Summarization

Elena Lloret, Ester Boldrini, Tatiana Vodolazova, Patricio Martínez-Barco, Rafael Muñoz, Manuel Palomar

Department of Software and Computing Systems
University of Alicante, E-03080, Alicante, Spain
{elloret,eboldrini,patricio,rafael,mpalomar}@dlsi.ua.es; naiadhe@gmail.com

Abstract

The Web 2.0 has resulted in a shift as to how users consume and interact with the information, and has introduced a wide range of new textual genres, such as reviews or microblogs, through which users comunicate, exchange, and share opinions. The explotation of all this user-generated content is of great value both for users and companies, in order to assist them in their decision-making processes. Given this context, the analysis and development of automatic methods that can help manage online information in a quicker manner are needed. Therefore, this article proposes and evaluates a novel concept-level approach for ultra-concise opinion abstractive summarization. Our approach is characterized by the integration of syntactic sentence simplification, sentence regeneration and internal concept representation into the summarization process, thus being able to generate abstractive summaries, which is one the most challenging issues for this task. In order to be able to analyze different settings for our approach, the use of the sentence regeneration module was made optional, leading to two different versions of the system (one with sentence regeneration and one without). For testing them, a corpus of 400 English texts, gathered from reviews and tweets belonging to two different domains, was used. Although both versions were shown to be reliable methods for generating this type of summaries, the results obtained indicate that the version without sentence regeneration yielded to better results, improving the results of a number of state-of-the-art systems by 9%, whereas the version with sentence regeneration proved to be more robust to noisy data.

1. Introduction and motivation

Each day more, within the framework of the current digital society of the Web 2.0, most of the information users produce can be found on-line and a great part of it is freely accessible. Thus we all have huge amounts of terabytes of user-generated content (e.g. in blogs, fora, review sites, social networks, etc.) that anybody can freely use (consciously or not) in our daily decision-making process. An example of this could be the typical situation when we are to buy a product; we first check the Web to find opinions, comments, information, on that item expressed by users that already have such product before deciding to acquire it or not. This phenomenon can be recognized as the modern (using data on the Web) version of the ancient Word of Mouth (WOM) originally developed in the streets many years ago when people relied on the opinions of others about a specific issue. However, the electronic Word of Mouth (eWOM) main difference with respect to the traditional one is that nowadays information is spread all over the world, in many languages, about a wide range of topics, services, places, etc., and there are vast amounts of information referring to the same topic. On the one hand, it is very useful to have all this information available; however, on the other hand, there is the risk in which important and key information may be lost or passed unnoticed, since it is impossible to manually manage all the information in an effective and efficient manner. In the current society there is a shortage of time for spending it and have an in-depth look at each and every single entry.

Given this context and these user needs, Natural Language Processing (NLP) provides the appropriate mechanisms that allow people to automatically deal with Web textual information, speeding up and simplifying the tasks of retrieving, extracting, classifying and summarizing, so that they can have just the necessary information users may need at a specific moment and in real time.

In this process and to obtain a high quality and reliable results, text summarization is a crucial process towards this goal; the objective of text summarization is to automatically produce a brief and concise fragment of text, containing only the essential and most relevant information about the sources that have been used (Spärck Jones, 1999). When it comes to text summarization, a wide range of summary types can be found (e.g., multi-document, indicative, informative, opinion-based, etc.). Thanks to this, this task becomes customizable to the users needs. In recent years, new summary types are emerging; one of this new types is ultra-concise opinion summarization, which poses greater challenges to the task, since it aims to just capture the main essence of an opinionated document with a few words. Taking into consideration the summary size, this task could be viewed as the task of headline extraction, but its main difference lies in the type of text where it is applied (Web 2.0 platforms, such as microblogs), thus having to address the particularities of the Web 2.0 textual genres, and meet the strict length restriction these services may have (e.g., 140 characters long).

The motivation of our research relies on the fact that at present microblogs are the trendy textual genre of the Web 2.0 that is increasing more in comparison with others and consequently, this communication channel is being widely used by different user types and considered as a point of reference for many users (Kim et al., 2014; Hennig-Thurau et al., 2014; Abdullah et al., 2014). This implies that the generation of such brief summaries has the potential of reaching a huge number of users.

Therefore, taking into account the importance of the eWOM and the potentiality of NLP technologies in the Web 2.0 scenario, this research work proposes the development and analysing of an innovative text summarization approach able to generate ultra-concise opinion abstractive summaries in the form of a tweet starting from different sources (reviews and microblogs). It is worth stressing that the main contribution of this article lies in the manner that this type of summaries is generated, since our approach performs abstractive summarization[1], which is one of the most challenging issues for current summarization approaches.

---

[1]Abstractive summarization differs from extractive in that the summary does not limit to the selection

In this manner, our proposed approach goes beyond the mere identification and extraction of the most relevant sentence by including a sentence simplification and language regeneration stage that allow us to effectively condense the information. We thus identify the key parts of the sentences and then, generating the information, we are able to minimize the effects on its cohesion and coherence. In order to verify the effectiveness of the approach proposed, we tested it with two macro topics that are usually discussed considerably in the framework of the Web 2.0: technology and motor. After that we perform a qualitative evaluation to determine the appropriateness of our approach with respect to several quality criteria. The results show that the proposed simplification and regeneration stages are an added value to the current summarization strategies, having been proven to generate appropriate ultra-concise opinion summaries.

In this research article, our approach is presented as a single expert intelligent system, and evaluated intrinsically on its own, focusing on the quality of the opinion summaries that is able to generate. Undoubtedly, opinion summarization is one of the most valued and powerful NLP technologies. This type of technologies are acting as expert systems, given their power to advice or assist humans and/or other automatic processed when it comes to make any decision (Wang et al., 2013), (Darling, 2014), so more and more they are becoming essential in our current society (Nassirtoussi et al., 2014). Therefore, having shown the appropriateness of our method, this could be integrated and exploited as part of more complex applications, such as ERPs and other business intelligence process to help companies to deal with the eWOM, for instance, by disseminating ultra-concise messages through different channels to advertise and attract more clients or to make its potential customers aware of the high reputation of their products Pai et al. (2013). On the other hand, from a user point of view, this type of application could be very useful for providing the most outstanding feature for a product, service, etc. based on the on-line available information related to it.

---

and output of the most relevant sentences. In abstractive summarization a sentence transformation process is involved (e.g. sentence compression)

As it can be seen, the technology proposed in this article, could be easily integrated in a real-life application that will allow users to save their time and effort since the system will do the job automatically analysing the texts selected and summarizing their content in a reliable way.

The remaining of this article is organized as follows. Section 2 covers the related work and puts our work in perspective. Section 3 explains our proposed approach for generating ultra-concise summaries, together with the tools and resources employed. Section 4 reports the corpus developed and the experimental framework. Section 5 contains the evaluation carried out and the results obtained providing a comparison with state-of-the-art approaches, as well as discussing the potentials and limitations of the approach. Finally, Section 6 concludes the article and outlines future work.

2. Related Work

Text Summarization was initiated by Luhn (Luhn, 1958) and Edmundson (Edmundson, 1969) when analysed the first approaches to generate summaries automatically. Since then, most of the research in text summarization has focused on summarizing news documents, exploiting a wide range of techniques (McKeown et al., 2002; Barzilay and McKeown, 2005; Sarkar, 2012; Kabadjov et al., 2013; Huang et al., 2013).

However, during the last decade, the expansion of the Web, and in particular, the birth of the Web 2.0 has raised the need for new types of summaries tightly related to the opinions users express about a specific topic, product, service, etc. In this context, new textual genres, such as blogs, reviews, or social networks have emerged, providing new forms of communication that highly differ from the traditional ones (i.e., news documents). Thus, the consequence is that there is the need to face new challenges due to different linguistic phenomena. For instance, the Web 2.0 is characterized by its informal nature (Mosquera and Moreda, 2012) that means new challenges for existing NLP tools and methodologies for language treatment.

Given the above context, there has been much interest in proposing summarization approaches for these new scenarios in recent years. Opinion or sentiment-based summarization are at a general level, a new type of summary that aim to provide the most relevant pros and cons of a specific product, service, etc. In (Pang and Lee, 2008), the importance of this type of summaries is acknowledged due to the vast amounts of opinions stated on the Internet. Having made an extensive analysis of the most relevant research on the field, we realised that almost nothing has been done on automatic summarization from different textual genres simultaneously and also with a content generation aspect. And here is the novelty of our research. We work with different textual genres (blogs and microblogs), retrieve data from these sources on a specific topic and we are able to generate a tweet that contains the most relevant content. Thus, the novelty is twofold: on the one hand we are able to treat texts from different sources, that present different linguistic challenges and on the other hand, we generate the tweet that is a summarization of the most relevant content.

Concerning approaches producing sentiment-based summaries initially focused mainly on blogs (Balahur-Dobrescu et al., 2009; Missen et al., 2009; Mithun, 2010); There is also a high number of approaches addressing the summarization of reviews (Zhuang et al., 2006; Lerman and McDonald, 2009; AleEbrahim and Fathian, 2013; Ravi Kumar and Raghuveer, 2013; Raut and Londhe, 2014), forums (Carbonaro, 2010; Ren et al., 2011) as well as microblogs (Sharifi et al., 2010; Harabagiu and Hickl, 2011; Chakrabarti and Punera, 2011; Bahrainian and Dengel, 2013). We also find Farzindar (2014), focusing on how summarization tasks can improve social media retrieval and event detection, and Li et al. (2014), in which the multi-document summarization by sentence compression is explored.

Regarding microblogging services, the most popular is Twitter. It is a huge repository of data that is gaining popularity especially in text summarization focusing on generating brief summaries starting from a collection of texts like microblog entries, like tweets Connor et al. (2010); Sharifi et al. (2010) or enriched with other sources of information (Liu et al., 2011). In (Sharifi et al., 2010), a trending topic was taken as a starting point, and then all

related posts were collected and summarized, using machine-learning algorithms to detect the sentences mostly related to the topic phrase. Similarly, the approach proposed in (Weng et al., 2011) attempted to summarize Twitter posts, following a two-stage approach. Firstly, the posts and the responses were classified according to their intention (interrogation, sharing, discussion and chat), and then, different strategies for producing the summary were analysed, including sentiment analysis techniques. An interesting aspect of this research work was the presentation of the final summary, which depended on its intention with respect to the classification made. For instance, if the summarized posts were classified as sharing, the summary was a pie chart showing the percentage of positive and negative opinions.

Different from the aforementioned approaches, automatic generation of tweets is gaining attention as well, as a new type of summaries (i.e., ultra-concise summaries). Although this task may be similar to that of generating headlines, it is more difficult, since a tweet can only contain 140 characters, and the information available in this source may be ill-formed and noisy. This implies that the generated summary has to be very short at the same time concise and informative enough. Although there is extensive research addressing headline generation (Zhou and Hovy, 2004; Sarkar and Bandyopadhyay, 2005; Xu et al., 2010; Lopez et al., 2012), the task of tweet generation is rather novel. In (Lloret and Palomar, 2013), the extent to which current automatic summarization systems were able to generate tweets by employing extractive techniques was analyzed, obtaining promising results for continuing investigating in this area. In the literature, the term of *"ultra-concise summaries"* appeared in the last two years for referring the new type of summaries that aim to express in the form of a tweet, very concise information. In (Ganesan et al., 2012), this type of summaries was generated in the context of opinionated information. Specifically, an automatic tweet was generated from a set of reviews, and each tweet was a summary of a key opinion in the text. The techniques employed for producing such summaries are based on Web N-grams, and the results obtained were very promising. In (Lofi and Krestel, 2012), a potential application of automatic tweet generation is described, where the tweets are used to increase the transparency of government

7

actions. The proposed approach combines several NLP techniques, such as topic classification, or text summarization; however, it is just a task proposal, and none of the techniques and the system as a whole are analyzed or evaluated. More recently, it is also worth mentioning the work of Liu et al. (2015), in which short text summarization from social network streams is performed. In their approach, the authors aim to model a novel incremental clustering problem based on the requirements of comment stream summarization, focusing on social networks. For this, they design at-a-glance presentation, which is concise, informative, and impressive, to help users easily and rapidly get an overview understanding of a comment stream, and thus, it could be considered a summary. For determining which comments are representative, the authors used a term vector representation, and applied a basic linguistic process removing punctuation signs, stopword, and performing stemming. Then, an heuristic to perform normalization, by identifying redundant characters and removing them (e.g. "soooo" will be transformed to "so"), and finally a stemming process is applied. To account for the relevance of comments, n-gram term extraction is employed, together with clustering techniques, in order to group the comments and extract the one which will constitute the final summary.

Based on the users' needs identified and the research carried out so far, it can be said that the generation of ultra-concise summaries (and more specifically ultra-concise opinion summaries) is of a great usefulness and at the same time represents a great potential for users (individuals and collectives). This is because it provides relevant, concise and reliable information related to a topic in real time. And to the best of our knowledge, most of the existing approaches have limited to follow an extractive strategy using surface summarization techniques (i.e., n-grams or word frequency).

After having reviewed the most relevant state of the art in the field of NLP dealing with social media summarization, our research is meant to go a step further. The approach we present in this article is highly innovative. This is due to the fact that we work with different textual genres of the Web 2.0 simultaneously, and this is done to create a comprehensive

summary of their content (i.e., an ultra concise summary in the form of a tweet). So far, previous research addressed the task of automatic summarization separately for each textual genre; our approach deals with multiple textual genres at the same time for the provision of a summary that includes their content. This implies a considerable progress with respect to the existing approaches.

3. Abstractive Concept-level Summarization

This section focuses on providing details about our ultra-concise opinion summarization approach. In this manner, the different stages that were designed and developed to build the approach are first explained (Section 3.1), followed by the specific tools and resources for implementing it (Section 3.2).

3.1. Our approach

For building our ultra-concise concept-level summarization approach, we relied on the classical generic flow for summarization process, proposed in Spärck Jones (1999). In this process, three broad phases are distinguished: i) *interpretation*; ii) *transformation*; and iii) *summary generation*. In the first one, the input text needs to be analyzed and understood; the second phase transform the input text into an internal representation, so that the main ideas can be further extracted, and finally the last phase determines which content is the most important among the input sentences (in the case of extractive summarization), or the rephrased sentences when natural language generation is also addressed (i.e., abstractive summarization).

As part of this summarization process, in our approach, each of these phases were broken into smaller modules, each one having a very concrete purpose. The main novelty of our approach with respect to other existing summarization approaches is the inclusion of a simplification and a regeneration stages. Specifically, the complete summarization pipeline goes through the following stages, which are graphically depicted in Figure 1):

9

- **Syntactic text simplification**: At this stage the long sentences are being split into two or more simple ones, ideally consisting of a single clause only. This will allow us to have less complex dependency trees, which in its turn simplifies the process of text analysis and information representation in the following stages.

- **Text analysis**: Here we analyse all the sentences generated in the previous stage to extract:

  - subject-verb-object (SVO) triples

  - noun phrases with the head noun explicitly specified

  - named entities

  This information is necessary for the concept representation generation in the next stage.

- **Concept representation**: The SVO triples are being stored internally using a representation scheme crafted specifically for this purpose. This scheme allows to store subjects, objects and complete noun phrases with the indication of head noun. This representation reflects the characteristics of the given triple that are necessary to further generate a valid sentence using this information. Once the representations have been created, we can manipulate the information that they store in a more efficient way than raw text data.

- **Concept analysis and summarization**: At this stage the concepts that represent the SVO triples are analysed taking into consideration the information that they share, namely the identical head nouns and other nouns of the subject and object noun phrases as well as verbs.

  - **additional summarization techniques**: This stage is optional. It involves extending the proposed concept summarization to include term frequency and named entity scoring.
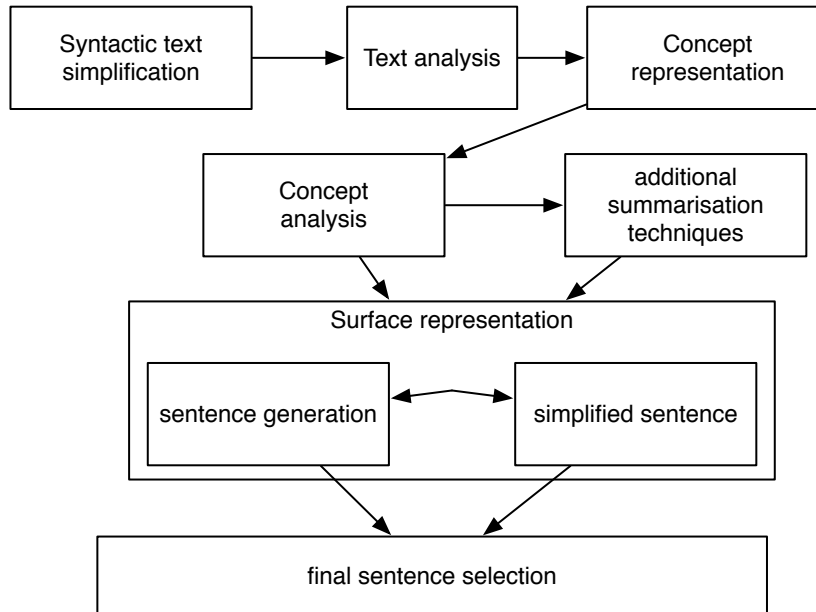
Figure 1: Stages of summarization process.

This is the stage where the actual process of concept rating and summarization takes place. In the case of integrating summarization techniques, statistical computations, based on term or named entity frequencies are performed.

- **Surface representation**: At this stage the selected concept representations are converted to text form. This can be done in two ways:

    - regenerating the simplified sentence

    - selecting the simplified sentences that produced the given SVO triples

- **Sentence selection**: Once the concepts were rated and translated into a corresponding surface representation, the system selects the highest rated sentence with respect to the maximum allowed size of the tweet.

Moreover, in this particular research work, we focus on information obtained from the Web 2.0, which is mainly subjective. Therefore, taking into account the general concept-level summarization approach, we can adapt it to produce the type of ultra-concise opinion summaries pursued. For this, we feed the system with polarity information that was obtained

after a manual annotation process (please see Section 4.1 for more information), although it would have been also possible to integrate an opinion mining system within the process.

3.2. Tools and Resources

To implement and complete all the stages of the proposed approach for abstractive concept-level summarization, the following tools were used, listed here in the stage order illustrated in Figure 1:

- Simplified Factual Statement Extractor (Heilman and Smith, 2010) was employed to solve the task of syntactic sentence simplification

- Stanford CoreNLP Tools[2] was used to carry out basic NLP analysis: tokenization, lemmatization, POS-tagging, named entity recognition, syntactic and dependency parsing

- Java provided DOM parser was used to extract noun phrases

- The Word Vector Tool[3] assisted to extract term frequencies

- The text regeneration stage was completed by means of SimpleNLG API (Gatt and Reiter, 2009)

Additionally, as it was previously mentioned, the generation of opinion summaries was possible since we used Emotiblog annotation scheme (Boldrini et al., 2010). This scheme served as the model for carrying out the annotation of opinionated information in terms of polarity and intensity.

4. Experimental Framework

This section introduces the corpus we gathered for conducting the experiments, as well as the experiments performed.

---

[2]http://nlp.stanford.edu/software/corenlp.shtml
[3]http://www-ai.cs.uni-dortmund.de/SOFTWARE/WVTOOL/index.html.en

4.1. Corpus Development

We automatically gathered a corpus using the crawler developed in (Fernández et al., 2010) that looked for reviews gathered from Amazon[4] and WhatCar[5], and tweets extracted from Twitter[6]. We focused on the English language, and we selected the top ranked mobile phones and cars retrieved by the crawler, obtaining 10 documents from each site and domain. In total, our corpus had around 400 English texts, which were preprocessed and only the main text was kept for being processed by our approach. This included the removal of elements, such as unnecessary information tags (e.g., timestamp, location).

The reason for selecting these sorts of documents and for these domains is that they are frequently discussed by people with different profiles and constitute the starting point for demonstrating the efficiency of our system in multi-domain and multi-genre contexts. Furthermore, mobile phones and car topics can be seen as in between the high-level terminology complexity of a specific and technical domain, such as medicine for example, and an easier one, like music or food. This means that we focus on a medium level complexity topic in order to determine whether our techniques could be appropriate for a real-life application. Table 1 summarizes some statistics of the corpus developed.

Table 1: Corpus statistics.

	Technology	Motor
Topic	Mobile phones	Cars
Number of topics	10	10
Number of microblogs per topic	13	10
Number of online reviews per topic	9	10
Microblogs (Avg. Number of words per tweet)	28.05	14.76
Microblogs (Number of words in total)	3647	1476
Reviews (Avg. Number of words per review)	146.30	156.13
Reviews (Number of words in total)	13167	15613

After the corpus collection, we started the manual annotation process, where one expert

---

[4]http://www.amazon.com
[5]http://www.whatcar.com
[6]https://twitter.com/

annotator used the coarse-grained version of the EmotiBlog annotation schema (Boldrini et al., 2010) for labeling the corpus. This annotation model is used to support automatic systems to detect the subjective expressions in the new textual genres of the Web 2.0. It has been employed to improve the performance of different NLP systems dealing with complex tasks, including opinion summarization, where it obtained satisfactory results. The expert annotator labeled the corpus at sentence level during 4 weeks in part time dedication using the following EmotiBlog elements: POLARITY (positive/negative/neutral) + INTENSITY (high/medium/low).

The information resulted from this annotation process will be very important for producing opinion summaries. As previously mentioned, taking into account our proposed approach for summarization, we integrate opinion information, in order to be able to generate ultra-concise opinion summaries, thus allowing to produce different types.

4.2. Experiments

As indicated in the description of the summarization approach in Section 3, our system offers two different ways of surface representation of the concepts, namely using natural language generation or the original simplified sentence that yielded the selected concept. Another choice that we have is whether to use term frequency and named entity-based scoring to aid the summarization process or not. This leaves us with 4 different system settings as listed below[7]:

- **S+SVO+TFNE**. This setting yields a simplified sentence in the surface representation phase and term and named entity frequencies scoring during the summarization phase. Here is an example tweed generated using this setting:

  The car is well planted on the road nothing like my brothers 318d but still very good for such a massive car.

---

[7]where the initial S and R correspond to "simplified" or "regenerated" sentences respectively, and TFNF indicates the additional scoring using term and named entity frequency

- **S+SVO**. This setting also yields a simplified sentence, however only the shared information between the concepts is used to rate them. Consider the following example tweet:

  I was impressed with the cabin space, quietness and big car feel.

- **R+SVO+TFNE**. This setting yields a regenerated sentence in the surface representation phase and term and named entity frequencies scoring during the summarization phase. Find an example below:

  My only disappointment with this car is the high noise level at highway speeds.

- **R+SVO**. This setting also yields a regenerated sentence, whose corresponding concept was rated using only the shared information between the concepts:

  The blackberry 9520 storm is a great little phone ; very stylish, no fiddly keys.

Those four system settings were used to generate ultra-concise summaries of the original documents for both of the domains represented in our corpus: mobile phones and cars. Besides the ultra-concise summaries (i.e., tweets) for the original documents, we also generated summaries targeting a different sentiment: a) positive; b) negative; c) objective; and d) mixed, that involves both objective and subjective information. This manner, we could see how our approach performs with different types of summaries, and whether it exhibits a different behaviour when generating each of them. It is worth mentioning that we took advantage of the information annotated in the corpus, and we extract the opinionnated information directly from it to produce the different types of ultra-concise summaries.

5. Evaluation and Discussion

After having underlined the added value of our research basing on the state of the art and having described the resources we use and our approach, this section focuses on the evaluation of our approach. We discuss the methodology used and the results obtained.

5.1. Evaluation metholodogy

A qualitative evaluation was performed, where three expert users assessed in an independent manner all the types of ultra-concise summaries automatically generated. For this, three quality criteria were established, based on the DUC/TAC readability criteria[8]: a) the content of the summary, b) its readability and c) its overall responsiveness. Specifically, the first criterion determines whether the tweet reflects important information of the source inputs; the second assesses whether the tweet is properly written and easy to understand; and the third evaluates if it is reliable and suitable for a real-life application. It is worth mentioning that the ultra-concise opinion summaries generated will be evaluated from a conceptual point of view and the general idea expressed, rather than from the individual words building up the summaries. Moreover, evaluating our results with these criteria allows us to determine if the automatic output obtained through our approach is appropriate without needing additional treatment.

Therefore, each of the tweets (original, positive, negative, objective and mixed) for each of the experimental settings (S+SVO+TFNE; S+SVO; R+SVO+TFNE; R+SVO) was rated according to a 3-level Likert scale. The values of the scale ranged from 1 to 3 (1=poor or very poor; 2=barely acceptable; and 3=good or very good). It is important to stress that the assessors had access to the original reviews and tweets, from which the automatic tweets were generated, and they read them in advance for being able to determine whether the automatic tweets were reliable and a good representative of the source documents.

5.2. Results and Discussion

Tables 2 and 3 show the results obtained that will be analysed from four perspectives.

Simplified vs Regenerated and SVO vs TFNE. Regarding the results, they show that for the system setting that involved natural language generation in the final phase, the summarization approach that includes term and named entity frequencies performed slightly better.

---

[8]http://www.nist.gov/tac/2008/summarization/op.summ.08.guidelines.html#evaluation

Table 2: Regenerated SVO+TF+NE

	Content		Readability		Overall Responsiveness		Global Average	
	Phones	**Cars**	**Phones**	**Cars**	**Phones**	**Cars**	**Phones**	**Cars**
Original	1.633	1.933	1.933	2.033	1.6	1.666	1.722	1.877
Positive	**2.037**	2.5	2.148	2.4	**1.888**	2.3	**2.024**	2.4
Negative	1.925	2.3	**2.259**	2.266	1.703	2.066	1.962	2.211
Objective	1.185	1.4	1.666	1.8	1.185	1.4	1.345	1.533
Mixed	1.777	**2.75**	2.037	**2.708**	1.629	**2.541**	1.814	**2.666**

Table 3: Simplified SVO

	Content		Readability		Overall Responsiveness		Global Average	
	Phones	**Cars**	**Phones**	**Cars**	**Phones**	**Cars**	**Phones**	**Cars**
Original	1.533	2.3	1.966	2.533	1.333	2.266	1.611	2.366
Positive	2.333	2.566	2.555	2.733	2.185	2.433	2.358	2.577
Negative	**2.444**	2.7	**2.814**	**2.833**	2.148	2.466	**2.469**	2.666
Objective	1.333	1.9	2	2.3	1.296	1.766	1.543	1.988
Mixed	2.407	**2.8**	2.666	2.766	**2.259**	**2.566**	2.444	**2.711**

However, the results were the opposite when no generation was employed: the mere SVO-scoring based on the mutual information of object and subject noun phrases as well as verbs performed better. Both for simplification and regeneration methods we provide only the best results (see Tables 2 and 3 respectively). To explain this phenomenon we computed the average number of tokens for all of the four system settings. The results show that the average number of tokens for the R+SVO setting is 15.1, R+SVO+TFNE is 15.4 and both S+SVO and R+SVO+TFNE have an average of 15.7 tokens. Thus we can conclude that the longer tweets were rated higher by the annotators. Moreover, since both the R and S settings select one and the same SVO triple to constitute the summary, we can use the results obtained for the S settings to determine which summarization method selects better sentences. The SVO evaluation numbers are higher than the SVO+TFNE ones, thus we conclude that term and named entity frequencies introduce unnecessary information and reduce the quality of generated summaries. On the other hand these sentences are easier converted into a concept representation and then regenerated from this representation. This explains the difference in numbers between S and R settings with regard to SVO and SVO+TFNE summarization methods.

Sentiment and original tweet summarization. Table 2 and 3 show that objective tweets are the most difficult to summarize. To better understand the challenges of objective tweet summarization, we analysed our corpus again and discovered that this type of tweets has the highest rate of advertisement that is often published in this kind of web pages. Here is an example of such a tweet:

Daily deals Nokia 2700 classic Black Grey (Unlocked Quadband) GSM Cell Phone

The next type of tweet at the bottom of the evaluation scale is the original tweet that due to the same problems of advertisement contains a lot of noise. The best results for the cars domain were obtained for the mixed tweets. The best results for the mobile phones differ between the S and the R settings: the S+SVO handled the negative tweets the best, while the R+SVO+TFNE yielded the best numbers for the positive tweets.

Cars vs mobile phones domains. Both the regenerated and the simplified system settings expose the same tendency: for the cars domain our system performs better. However, the domain specific information doesn't influence the system performance. Those are rather the review and website characteristics that are responsible for this behaviour. The used mobile phone reviews contain a high number of advertisements and tend to be shorter and less descriptive than the car reviews.

Simplified vs regenerated tweet. Taking into account that both the simplified and the regenerated system settings select the same concept, we have the possibility to evaluate the how the surface representation affects the overall quality of an ultra-concise summary. To be more precise, this system shows whether the state of the art natural language generation tools can compete with an extractive summarization approach that is represented by a simplified sentence version in this case. The global average numbers show that simplification outperforms regeneration in both domains and all the sentiment types. However, the original tweets for the mobile phone domain that due to the high rate of advertisement produced worse results, are better summarized using regeneration. The explanation of this phenomenon we see in the sentence analysis and sentence selection stages: the text analyser fails to generate a com-

plete parse and dependency tress, since advertisements often contain incomplete sentences, with URLs. If those sentences can't be analysed, then they can't be stored properly and used further to generate a valid sentence. Those cropped regenerated sentences are too short to carry any significant information and thus discarded during the sentence selection phase. The simplified setting on the other hand, selects those sentences as they are. Often they happen to be long enough to not to be discarded in the sentence selection stage. This is an encouraging tendency, since the regeneration approach proves to be more robust to the noise, than the simplification approach. The common error of the regeneration approach that caused the low readability index is connected to the generation of negations. However, this is tool specific (SimpleNLG). Please consider the following example:

The car does not does not give any mechanical problems, apart from an over smelly AC which made my car smell quite a bit.

The repetition of the auxiliary verb together with the particle cause the decrease in the readability scores.

5.3. State-of-the-art Comparison

After testing our approach with the different settings, we decided to compare the best results obtained by either regeneration or simplification for each topic with respect to the state of the art. In the case of the technology domain (mobile phones), the best results for R+SVO+TFNE were obtained for the positive tweets, whereas when only simplification was applied with SVO representation (S+SVO) the best results were negative ones. In the case of the cars, the mixed tweets were the best performing ones for both R+SVO+TFNE and S+SVO.

The systems we used for comparison were the following ones:

- COMPENDIUM$_{SEM}$ (Vodolazova et al., 2012). This summarization approach incorporates semantic knowledge, specifically, anaphora resolution, word sense disambiguation, and concept frequency using Wordnet(Fellbaum, 1998).

- The approach proposed by Ganesan, Zhai and Viegas(Ganesan et al., 2012), which is able to produce ultra-concise opinion summaries as well.

- Swotti[9], which is a commercial system that provides summarized opinion information for a wide range of products.

The reason for selecting those systems is that they could generate the same type of summary that ours. The system, which may differ the most, is COMPENDIUM$_{SEM}$, which is a generic summarizer; but it was adapted to produce ultra-concise opinion summaries also. The evaluation performed was identical as the previously explained, but after that we computed the global average for each domain, thus reporting only those results. Table 4 shows such comparison results.

Table 4: Comparison results (global average)

System/approach	Phones	Cars
R+SVO+TFNE	2.02	2.67
S+SVO	**2.47**	**2.71**
Swotti	2.32	2.42
Ganesan et al., 2012	1.30	1.48
COMPENDIUM$_{SEM}$	1.70	2.65

According to the results in Table 4, the implemented approach outperforms other systems in the cars domain, and only the regeneration setting performed slightly worse than the Swotti system on the mobiles phones domain, thus putting our system on the first and third positions. The overall picture is very encouraging. In particular, the regeneration approach proves that the state of the art natural language generation tools are competitive enough and developed enough to be used in a real life application.

5.4. Potentials and Limitations

As we can deduce from the results obtained our approach presents some advantages. We clearly advance the state of the art since we include language generation in our approach

---

[9]http://swotti.starmedia.com/

and this is done taking into account input given from different textual genres and the results obtained show that the approach is well performing. An advantage of this approach is that we produce abstractive ultra-concise summaries, which is one the main challenges of the present research in the field.

Having said that, this approach has also room for improvement. Although we have internally represented the textual genres treated by means of their syntactic structures (SVO), we perceived there is a strong need for deeper linguistic analysis (e.g., semantics) that is able to tackle aspects such as irony, non-literal language or even negation phenomena. Another relevant issue (since we are working with social media, in particular, microblogs and reviews) that we will address as future work would be the treatment of informal language, widely employed by most of the Internet users.

6. Conclusion and Future Directions

This paper proposed a novel concept-level approach for summarization that was applied and evaluated in the context of ultra-concise opinion summarization. The described approach is characterized by integration of syntactic sentence simplification and sentence regeneration stages into the summarization process, as well as the internal concept representation that allows to some extent to uncouple the actual information summarization stage from the surface realization of concepts.

Our motivation for this research derives from the fact that microblogs, such as Twitter that can be considered as ultra-concise summaries, have gained high popularity and have become a common platform to communicate and spread information, reaching a high number of users. They constitute one of the means that allow eWOM, assisting users in the process of decision making based on the opinions and comments stated by other users through the communication channels of the Web 2.0. More specifically, the main research contributions are: i) we are able to treat texts from different sources, that present different linguistic challenges and ii) we generate the tweet that is a summarization of the most relevant content.

Both issues are crucial in the development of expert and intelligent systems (Zhan et al., 2009), (He, 2013).

Therefore, in order to achieve them, different types of ultra-concise opinion summaries were generated and evaluated, analysing in depth the stages involved in the process. The obtained evaluation results show that the implemented approach is a reliable method for ultra-concise opinion summary generation. It is competitive enough to outperform a number of other available systems. Both the regeneration and simplification versions yielded encouraging results. However, the simplification approach performed better. Regeneration approach in its turn proved to be robust on the noisy data, outperforming in this case the simplification approach. As for the differences of the summarization approaches, no unanimous decision can be taken. It was observed that the regeneration approach performed better when term and named entity frequencies scoring was included, and the opposite phenomenon was registered for the simplification approach. This question should be investigated further and applied to a number of different corpora to have a wider range of results, also validating the appropriateness of the integration of an opinion mining system in the summarization process.

After having proposed and evaluated our approach, and despite obtaining promising results, which demonstrate the robustness of the NLP methods and tools employed, we also analyzed possible issues learned in this research to be improved for developing further versions of our ultra-concise opinion summarization approach. Specifically, the future work we are planning is to evaluate both surface representation approaches on a number of different corpora. The goal is to investigate, whether introducing even more noise in the data will decrease the performance of the simplification approach and maintain the performance of the regeneration approach on the same level. On the other hand, it is worth trying to test both of the approaches on a corpus of easy reading texts, i.e. text that are developed for language learners. We hypothesize that in such an environment the natural language generation tools should perform better and improve the performance of the regeneration approach.

Moreover, another research direction could be the improvement of the summarization approach. So far our method does not involve word sense disambiguation. The information sharing of the concepts is judged by their lexical representation and does not involve deeper semantics. More semantics can be included into the summarization method by introducing word sense disambiguation combined with anaphora resolution, or normalizing the informal language for its better treatment and comprehension within our approach.

Finally, another improvement can be obtained from the evaluation side. There are a number of commonly used readability metrics, such as FRE, that evaluate text complexity. The readability index is an important measure that could further help us to decide which of the surface representation and summarization methods that our system implements generate better ultra-concise summaries.

Acknowledgment

References

Abdullah, N., Nishioka, D., Tanaka, Y., and Murayama, Y. (2014). A preliminary study on users decision making towards retweet messages. In Cuppens-Boulahia, N., Cuppens, F., Jajodia, S., Abou El Kalam, A., and Sans, T., editors, *ICT Systems Security and Privacy Protection*, volume 428 of *IFIP Advances in Information and Communication Technology*, pages 359–365. Springer Berlin Heidelberg.

AleEbrahim, N. and Fathian, M. (2013). Summarising customer online reviews using a new text mining approach. *International Journal of Business Information Systems.*

Bahrainian, S.-A. and Dengel, A. (2013). Sentiment analysis and summarization of twitter data. In *Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on*, pages 227–234.

Balahur-Dobrescu, A., Kabadjov, M., Steinberger, J., Steinberger, R., and Montoyo, A. (2009). Summarizing opinions in blog threads. In *Proceedings of the Pacific Asia Conference on Language, INformation and Computation Conference*, pages 606–613.

Barzilay, R. and McKeown, K. R. (2005). Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31:297–328.

Boldrini, E., Balahur, A., Martínez-Barco, P., and Montoyo, A. (2010). Emotiblog: a finer-grained and more precise learning of subjectivity expression models. In *Proceedings of the Fourth Linguistic Annotation Workshop*, LAW IV '10, pages 1–10.

Carbonaro, A. (2010). Towards an automatic forum summarization to support tutoring. In Lytras, M., Ordonez De Pablos, P., Avison, D., Sipior, J., Jin, Q., Leal, W., Uden, L., Thomas, M., Cervai, S., and Horner, D., editors, *Technology Enhanced Learning. Quality of Teaching and Educational Reform*, volume 73 of *Communications in Computer and Information Science*, pages 141–147. Springer Berlin Heidelberg.

Chakrabarti, D. and Punera, K. (2011). Event summarization using tweets. In *Proceedings of the Fifth International Conference on Weblogs and Social Media.*

Connor, B. O., Krieger, M., and Ahn, D. (2010). TweetMotif: Exploratory Search and Topic Summarization for Twitter. *Artificial Intelligence*, (May):384–385.

Darling, W. (2014). *Innovative Document Summarization Techniques: Revolutionizing Knowledge Understanding: Revolutionizing Knowledge Understanding*, chapter Approaches to Large-Scale User Opinion Summarization for the Web, page . IGI Global.

Edmundson, H. P. (1969). New methods in automatic extracting. In *Inderjeet Mani and Mark Maybury, editors,* Advances in Automatic Text Summarization, pages 23–42. MIT Press.

Farzindar, A. (2014). *Social Network Integration in Document Summarization*, chapter 6, pages 139–162. IGI-Global.

Fellbaum, C. (1998). *WordNet: An Electronical Lexical Database.* The MIT Press, Cambridge, MA.

Fernández, J., Gómez, J. M., and Barco, P. M. (2010). Evaluation of web information retrieval systems on restricted domains. *Procesamiento del Lenguaje Natural*, 45:273–276.

Ganesan, K., Zhai, C., and Viegas, E. (2012). Micropinion generation: An unsupervised approach to generating ultra-concise summaries of opinions. In *Proceedings of the 21st International Conference on World Wide Web*, pages 869–878, New York, NY, USA. ACM.

Gatt, A. and Reiter, E. (2009). Simplenlg: a realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, ENLG '09, pages 90–93, Stroudsburg, PA, USA. Association for Computational Linguistics.

Harabagiu, S. M. and Hickl, A. (2011). Relevance modeling for microblog summarization. In Adamic, L. A., Baeza-Yates, R. A., and Counts, S., editors, *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia.* The AAAI Press.

He, W. (2013). Improving user experience with case-based reasoning systems using text mining and web 2.0. *Expert Systems with Applications*, 40(2):500 – 507.

Heilman, M. and Smith, N. A. (2010). Extracting simplified statements for factual question generation. In *Proceedings of the 3rd Workshop on Question Generation.*

Hennig-Thurau, T., Wiertz, C., and Feldhaus, F. (2014). Does twitter matter? the impact of microblogging word of mouth on consumers adoption of new movies. *Journal of the Academy of Marketing Science*, pages 1–20.

Huang, X., Wan, X., and Xiao, J. (2013). Comparative news summarization using concept-based optimization. *Knowledge and Information Systems*, pages 1–26.

Kabadjov, M., Steinberger, J., and Steinberger, R. (2013). Multilingual statistical news summarization. In Poibeau, T., Saggion, H., Piskorski, J., and Yangarber, R., editors, *Multi-source, Multilingual Information Extraction and Summarization*, Theory and Applications of Natural Language Processing, pages 229–252. Springer Berlin Heidelberg.

Kim, E., Sung, Y., and Kang, H. (2014). Brand followers retweeting behavior on twitter: How brand relationships influence brand electronic word-of-mouth. *Computers in Human Behavior*, 37(0):18 – 25.

Lerman, K. and McDonald, R. (2009). Contrastive summarization: An experiment with consumer reviews. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 113–116.

Li, C., Liu, Y., Liu, F., Zhao, L., and Weng, F. (2014). Improving multi-documents summarization by sentence compression based on expanded constituent parse trees. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 691–701, Doha, Qatar. Association for Computational Linguistics.

Liu, C., Tseng, C., and Chen, M. (2015). Increst: Towards real-time incremental short text summarization on comment streams from social network services. *IEEE Transactions on Knowledge and Data Engineering*, PP(99):1–14.

Liu, F., Liu, Y., and Weng, F. (2011). Why is "SXSW" trending? Exploring Multiple Text Sources for Twitter Topic Summarization. In *Proceedings of the Workshop on Language in Social Media* , pages 66–75.

Lloret, E. and Palomar, M. (2013). Towards automatic tweet generation: A comparative study from the text summarization perspective in the journalism genre. *Expert Systems with Applications*, 40(16):6624 – 6630.

Lofi, C. and Krestel, R. (2012). iParticipate: Automatic tweet generation from local government data. In *Database Systems for Advanced Applications*, volume 7239 of *Lecture Notes in Computer Science*, pages 295–298. Springer Berlin Heidelberg.

Lopez, C., Prince, V., and Roche, M. (2012). Just title it! (by an online application). In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 31–34, Avignon, France. Association for Computational Linguistics.

Luhn, H. P. (1958). The automatic creation of literature abstracts. In *Inderjeet Mani and Mark Maybury, editors,* Advances in Automatic Text Summarization, pages 15–22. MIT Press.

McKeown, K., Barzilay, R., Evans, D., Hatzivassiloglou, V., Klavans, J. L., Nenkova, A., Sable, C., Schiffman, B., and Sigelman, S. (2002). Tracking and summarizing news on a daily basis with the columbia's newsblaster. In *Proceedings of the Human Language Technology (HLT) Conference. San Diego, CA*.

Missen, M. M. S., Boughanem, M., and Cabanac, G. (2009). Challenges for sentence level opinion detection in blogs. In *ICIS '09: Proceedings of the 2009 Eigth IEEE/ACIS International Conference on Computer and Information Science*, pages 347–351, Washington, DC, USA. IEEE Computer Society.

Mithun, S. (2010). Exploiting rhetorical relations in blog summarization. In Farzindar, A. and Keelj, V., editors, *Advances in Artificial Intelligence*, volume 6085 of *Lecture Notes in Computer Science*, pages 388–392. Springer Berlin Heidelberg.

Mosquera, A. and Moreda, P. (2012). A qualitative analysis of informality levels in web 2.0 texts: The facebook case study. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, pages 28–45.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., and Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41(16):7653 – 7670.

Pai, M.-Y., Chu, H.-C., Wang, S.-C., and Chen, Y.-M. (2013). Electronic word of mouth analysis for service experience. *Expert Systems with Applications*, 40(6):1993 – 2006.

Pang, B. and Lee, L. (2008). Opinion mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Raut, V. and Londhe, D. (2014). Opinion mining and summarization of hotel reviews. In *Computational Intelligence and Communication Networks (CICN), 2014 International Conference on*, pages 556–559.

Ravi Kumar, V. and Raghuveer, K. (2013). Dependency driven semantic approach to product features extraction and summarization using customer reviews. In Meghanathan, N., Nagamalai, D., and Chaki, N., editors, *Advances in Computing and Information Technology*, volume 178 of *Advances in Intelligent Systems and Computing*, pages 225–238. Springer Berlin Heidelberg.

Ren, Z., Ma, J., Wang, S., and Liu, Y. (2011). Summarizing web forum threads based on a latent topic propagation process. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 879–884, New York, NY, USA. ACM.

Sarkar, K. (2012). An approach to summarizing bengali news documents. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, pages 857–862, New York, NY, USA. ACM.

Sarkar, K. and Bandyopadhyay, S. (2005). Generating headline summary from a document set. In *Proceedings of Computational Linguistics and Intelligent Text Processing, 6th International Conference, CICLing 2005, Mexico City, Mexico*, pages 649–652.

Sharifi, B., Hutton, M.-A., and Kalita, J. (2010). Summarizing Microblogs Automatically. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 685–688.

Spärck Jones, K. (1999). Automatic Summarizing: Factors and Directions. In *Inderjeet Mani and Mark Maybury, editors,* Advances in Automatic Text Summarization, pages 1–12. MIT Press.

Vodolazova, T., Lloret, E., Muñoz, R., and Palomar, M. (2012). A comparative study of the impact of statistical and semantic features in the framework of extractive text summarization. In *Proceedings of the 15th International Conference on Text, Speech, and Dialogue Conference*, TSD '12, pages 306–313.

Wang, D., Zhu, S., and Li, T. (2013). Sumview: A web-based engine for summarizing product reviews and customer opinions. *Expert Systems with Applications*, 40(1):27 – 33.

Weng, J.-Y., Yang, C.-L., Chen, B.-N., Wang, Y.-K., and Lin, S.-D. (2011). IMASS: An Intelligent Microblog Analysis and Summarization System. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 133–138.

Xu, S., Yang, S., and Lau, F. C.-M. (2010). Keyword extraction and headline generation using novel word features. In Fox, M. and Poole, D., editors, *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15,*. AAAI Press.

Zhan, J., Loh, H. T., and Liu, Y. (2009). Gather customer concerns from online product reviews a text summarization approach. *Expert Systems with Applications*, 36(2, Part 1):2107 – 2115.

Zhou, L. and Hovy, E. (2004). Template-filtered headline summarization. In Marie-Francine Moens, S. S., editor, *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 56–60, Barcelona, Spain. Association for Computational Linguistics.

Zhuang, L., Jing, F., and Zhu, X.-Y. (2006). Movie Review Mining and Summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, pages 43–50.