

Ultra-Concise Multi-genre Summarisation of Web2.0: towards Intelligent Content Generation

Elena Lloret, Ester Boldrini, Patricio Martínez-Barco, Manuel Palomar

Department of Software and Computing Systems

University of Alicante

Apdo. de correos, 99

E-03080 Alicante, Spain

{elloret, eboldrini, patricio, mpalomar}@dlsi.ua.es

Abstract

The electronic Word of Mouth has become the most powerful communication channel thanks to the wide usage of the Social Media. Our research proposes an approach towards the production of automatic ultra-concise summaries from multiple Web 2.0 sources. We exploit user-generated content from reviews and microblogs in different domains, and compile and analyse four types of ultra-concise summaries: a) positive information, b) negative information; c) both or d) objective information. The appropriateness and usefulness of our model is demonstrated by its successful results and great potential in real-life applications, thus meaning a relevant advancement of the state-of-the-art approaches.

1 Introduction and Motivation

The Web 2.0 has created a framework where users from all over the world express their opinion on a wide range of topics via different communication Social Media channels, such as blogs, fora, micro-blogs, reviews, etc. Undoubtedly, all this information is of great value in today's competitive business environment, increasing the need for businesses to collect, monitor, and analyse user-generated data on their own and on their competitors' Social Media, such as Twitter (He et al., 2015). Moreover, this context is also fostering the electronic Word of Mouth (eWOM) (Jansen et al., 2009), an unpaid form of promotion (Duan et al., 2008) in which customers share with other users their experience with the product they bought, for example. WOM is an ancient phenomenon originated in the streets orally, but now, with the flourishing of the Web 2.0 it has been evolved in eWOM, whose essence is the same; the only difference is that it is not implemented orally, but

using the Social Media instead (Boldrini et al., 2010): fora, blogs, online reviews and microblogs. However, the huge amount and the heterogeneity of online data poses great challenges to the development of applications able to effectively retrieve, extract and synthesise the main content spread within Social Media. Due to the richness of Social Media data, its exploitation is being crucial for business-oriented applications, such as market analysis, competence monitoring or simply understanding the reasons behind customers' opinion on a product. Having at their disposal effective applications for information analysis and exploitation would mean for them having competitive advantage.

Recently, three Natural Language Processing (NLP) applications are gaining predominance, especially in the field of Social Media content analysis: i) information retrieval (Croft et al., 2009), ii) opinion mining (Pang and Lee, 2008) and iii) automatic text summarisation (Nenkova and McKeown, 2011). Information retrieval aims to search and determine relevant documents on the Web according to a specific user need or topic. The goal of opinion mining is to identify subjective language and classify it according to its sentiment or polarity (i.e., positive, negative or neutral information). Finally, text summarisation detects the most relevant pieces of information from one or multiple texts and presents the main ideas in a coherent fragment of text.

The main objective of this article is to apply the aforementioned NLP techniques to exploit the Social Media data generated through online reviews and microblogs. In particular, our aim is to produce innovative automatic ultra-concise summaries in the form of tweets (140 characters) reliable in terms of content (they reflect the opinions expressed on a topic positive/negative -) and form. Even if there are some previous studies on this topic (Ganesan et al., 2012), the novelty of our

approach comes from the usage of multiple textual genres simultaneously and the production of an ultra-concise summary (multi-genre summarisation). This means that our final summary is presented to the user in the form of a Tweet. The summary is representative of what has been said on a pre-defined topic. It is reliable since we perform a robust treatment of our source data. We treat each of the sources separately (because each textual genre has specific needs) and then merge the distinctive and relevant information for automatically building up the tweet as final outcome.

Microblogs, and more especially tweets, have a direct impact on eWOM communication. They empower people to share these brand-affecting points of view anywhere to almost anyone. Moreover, Princeton Survey Research Associates International¹ found that the microblogging site Twitter experienced massive amounts of growth over the past years with millions of new users joining and engaging with the site on a daily basis (Smith and Brennen, 2012). While the conciseness of microblogs keeps people from writing extensive reflections, it is exactly the micro part that makes microblogs peculiar if compared with other eWOM channels, such as blogs, webs, etc. Moreover, the advantage of having a tweet as a final outcome is that: a) we provide immediate information, b) users can take it and exploit it in the way they prefer (i.e. post it) and c) have a complete overview, comprehensive of different genres content in a friendly format, and d) save a lot of their time since the system carries out the job for them: retrieving, analysing, selecting, and providing them with the information they are looking for. Due to the limited length of a tweet, it is necessary to analyse to what extent and how current approaches could be adapted.

The motivation of our article lies on the fact that microblog is one of the most used Social Media channels and thus considered as a point of reference for many users. This implies that the generation we do of brief summaries would be useful for users that can use it directly in their Social Networks. Microblogs have the potential of reaching a huge number of users. But not only normal users can take advantage of it; for instance a company can exploit such ultra-concise summaries disseminating them through different channels for advertising purposes, to attract more customers or to

make its potential customers aware of the high reputation of their products; thus, making this technology a real-life application that will allow users to save a lot of time and effort since the system will do the job automatically analysing the texts selected and summarising their content in a reliable way.

The paper is organised as follows: Section 2 presents the most relevant related work, while Section 3 the corpus creation and Section 4 its annotation. Section 5 describes the methodology for generating ultra-concise summaries and Section 6 the evaluation of the results obtained. Section 7 analyses our approach in-depth, and finally, Section 8 outlines the main conclusions and future work.

2 Related Work

In the last years, there has been much interest in summarisation from Social Media, within the wider context of opinion summarisation. Twitter is now the most popular microblogging service. It is a huge repository of data and it is gaining popularity in different NLP tasks, especially in automatic text summarisation focusing on generating brief summaries starting from a collection of texts like microblog entries, like Tweets (O'Connor et al., 2010; Sharifi et al., 2010; Kim et al., 2014) or enriched with other sources of information, like Webpage links or newswire (Liu et al., 2011). In (Sharifi et al., 2010) a trending topic is considered as a starting point from which all related posts are collected and summarised. They generally use machine learning algorithms to detect the sentences mostly related to the topic phrase. In (Inouye and Kalita, 2011) a comparative analysis of different summarisation techniques is carried out to determine which is the most adequate for this type of summaries, and it is concluded that simple word frequency and redundancy reduction are the best techniques for the Twitter topics summarisation. In (Weng et al., 2011), the approach to summarise Twitter posts consists of two stages: i) classification of the posts and responses in different groups, according to their intention (interrogation, sharing, discussion and chat) and ii) analysis of different strategies for building the summary through sentiment analysis techniques or simply analysing the responses for each post. Their final summaries are generated depending on the category they belong (e.g., if the summarised posts are within the sharing groups, the summary is a pie

¹<http://www.psraai.com/> (last access 30 January 2017)

Domain	Technology	Motor
Topic	Mobile phones	Cars
Number of topics	10	10
Number of tweets 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

Table 1: Corpus statistics.

chart showing the percentage of positive and negative opinions). Other relevant studies aim at generating other types of summaries, like event summaries (Chakrabarti and Punera, 2011) or ultra-concise opinion summaries (Ganesan et al., 2012). In the former, the goal is to produce a real-time summary of events, focusing on the American Football games, and analysing the performance of different summarisers. In the latter, the objective is to generate a tweet from a set of reviews, where each tweet is a summary of a key opinion in text, and it relies on techniques based on Web N-grams.

Our research idea is based on (Ganesan et al., 2012), but the main novelty and added value of our research with respect to it is precisely the multi-genre perspective addressed: starting simultaneously from multiple sources of information, tweets and online reviews we treat them and produce an ultra concise summary that is reliable and representative in terms of content. This output can be directly used by the user in his own Social Networks and not to waste their time in analysing what they found and prepare a summary of the huge amount of information available.

3 Corpus Creation

We automatically gathered a corpus of online reviews and tweets in English for 10 different mobile phones and 10 cars using the crawler developed in (Fernández et al., 2010). A crawler is an automatic process in charge of retrieving and extracting the HTML content of a Website. In this process, relevant information sources are identified (e.g. Websites of reviews), and then the content of useful web pages within the selected sources of information are retrieved, downloaded, and extracted in an automatic, quick and easy manner using the well-known vector space

model. This manner, a corpus with a large number of documents can be created automatically. Using this crawler, 10 mobile phone brands and 10 car brand models were selected to conduct the experiments. For each brand, we obtained on average 10 microblogs extracted from Twitter and 10 online reviews from Amazon² and WhatCar³, having in total 400 texts. We selected cars and mobile phones, since they are frequently discussed by people with different profiles and at the same time they are from two different domains (important for demonstrating the efficiency of our approach in multi-domain 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 food or music. In our case, for this step of our research we focused on a medium level complexity topic to check if our techniques were pertinent for a real-life application.

Table 1 summarises the main features of our corpus. Analysing the specific properties and features for each of the Social Media textual genres, the most outstanding difference is the length of the texts. While the tweets are very short, having no more than 28 words on average, the reviews are quite longer, with more than 145 words on average. Both textual genres will be therefore complementary, tweets providing conciseness, whereas reviews providing opinions in a specific context.

4 Annotation Process

Once the documents were retrieved, we pre-processed them by extracting only the main text; a very important step since we eliminate all pos-

²<http://www.amazon.com> (last access 30 January 2017)

³<http://www.whatcar.com> (last access 30 January 2017)

sible characters that are not part of the information, links, and emoticons. Then, we started the annotation process, where one expert annotator manually labelled the documents using the coarse-grained version of the EmotiBlog annotation schema (Boldrini et al., 2010). To have the documents annotated by just one expert was considered enough for this preliminary study, since the EmotiBlog model had been previously evaluated (Boldrini, 2012) and proved to be easy to use. The annotation schema developed in EmotiBlog was created for automatic systems to detect the subjective expressions in the new textual genres of the Web 2.0 and has been employed to improve the performance of different NLP applications dealing with complex tasks, including opinion summarisation, where it obtained satisfactory results (Balahur et al., 2009).

Although EmotiBlog was originally a very fine-grained annotation scheme, we decided to use the less grained part of the resource, since our research purposes at this level only required to detect subjectivity and classify the polarity of a statement. Therefore, the expert annotator labelled 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 fragment below shows an example of labelled sentence for the topic “Nokia 2700”, a subjective sentence whose polarity is positive, with a high intensity.

```
<phenomenon degree1="high"
category="phrase"
polarity="positive"
source="w" target="Nokia
2700" confidence="high">It's
inexpensive, it's primarily a
phone but it has some useful
features, it's neat and it works
well.</phenomenon>
```

We selected the abovementioned elements of the EmotiBlog model since our purpose is to discriminate sentences between objective/subjective and from the subjective ones discriminate them into positive/negative and finally the summarisation system will treat those two groups to produce a reliable ultra-concise summary that reflects the reality of users feelings. This process implies the added challenge for the summarisation system to be able to treat the two typologies (objective/subjective), quite challenging task if we take

into consideration the high language variability present in the Web 2.0 (Pacea, 2011). The reason why we performed manual annotation was because we wanted to ensure a precise labelling and minimise cascade errors derived from the use of NLP tools. This manner we can focus more on the evaluation of the automatic summarisation approach.

5 Ultra-concise Summary Generation

To the best of our knowledge, this is one of the first attempts aiming at generating this new type of summaries starting from text sources of different nature (i.e., multi-genre), and about different opinion polarities (positive/negative/objective). This requires that the summarisation system has high coverage to produce such a small text with full meaning, relevant to the topic, and with grammatical adequacy. These types of summaries have enormous benefits as presented in Section 1. As it was previously said, the ultra-concise summarisation generation approach employed in this research is novel considering the following main aspects. First of all, it is able to deal with different types of Web 2.0 textual genres (tweets and reviews) thus employing NLP techniques able to treat the linguistic phenomena encountered (thanks to the annotation performed beforehand). The proposed approach takes as a starting point the corpus previously created, and then it performs a series of steps to determine the relevant information, and analysing the most appropriate manner to present it in the form of a tweet. In addition its modular architecture allows the inclusion of tools for deeper language/linguistic/content analysis. Figure 1 depicts an overview of our proposed approach. Next, the stages involved in the process are explained in more detail:

1. **Polarity detection and classification:** our main goal in this stage is to group the sentences into three sets: one for the positive, one for the negatives, and one for the objectives. In this manner, taking as a starting point a set of topic-related reviews and tweets, the first stage is to detect and classify the opinions contained. In our research work, we take advantage of the manual annotation process previously explained, because this manner, precise information concerning the positive, negative and objective sentences is obtained.

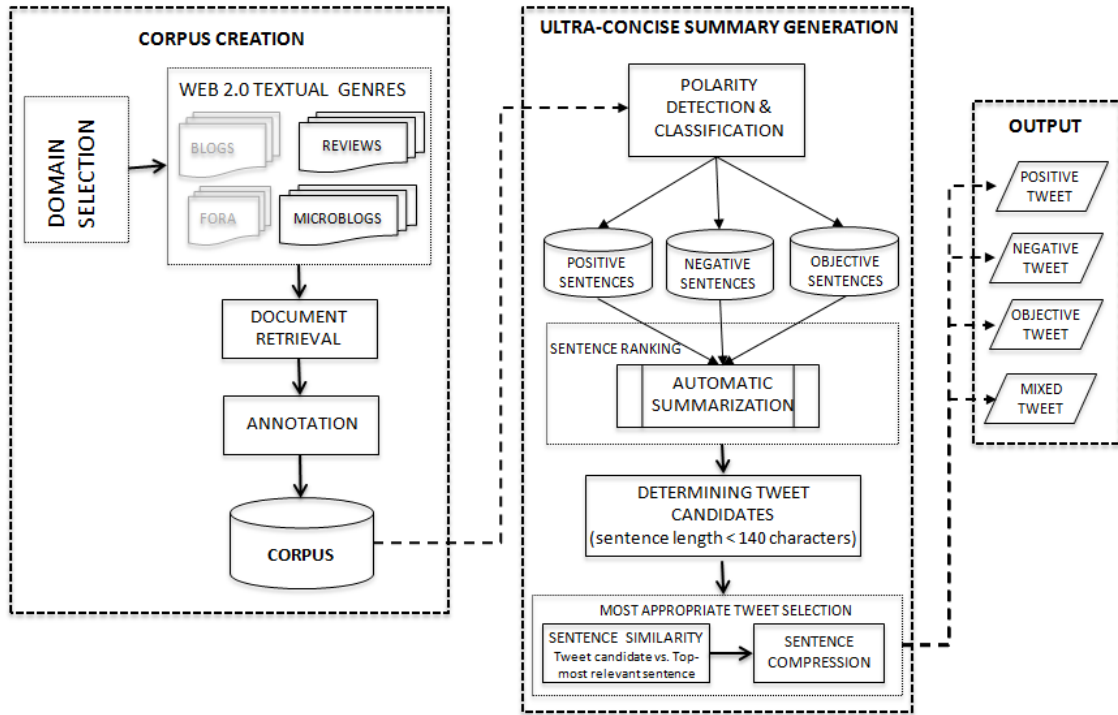


Figure 1: Overview of our proposed approach.

2. **Sentence ranking:** the aim of this stage is to assign a relevance score to each positive, negative or objective sentence. This relevance score is determined automatically, relying on automatic summarisation techniques. Specifically, two heuristics were used to compute the relevance for a sentence: term frequency and noun-phrase length. On the one hand, term frequency is a statistical technique that assumes that the more important sentences are determined by the most frequent words, without taking into account the words that do not carry any semantic information (i.e., stopwords such as “the”, “a”, etc.). On the other hand, the use of noun-phrase length has a linguistic motivation (Givón, 1990), where it is stated that longer noun-phrases carry more important information. Then, for calculating the score of each sentence, these two heuristics were combined, thus considering more important those sentences containing longer noun-phrases composed of high frequent words. The combination of both techniques has been proven to work fine for producing automatic summaries by means of COMPENDIUM summariser (Lloret and Palomar, 2013).
3. **Determining candidate text fragments for tweets:** although a relevance score was assigned to each sentence, one of the key issues of the task we are facing is how to produce tweet informative enough, ensuring at the same time, the 140 characters length restriction. At this stage, regardless of the relevance for each sentence, we identify from each group of sentences those ones not exceeding the 140 characters length to be combined with the relevance score in the next stage. Although this may seem a trivial approach, the current semantic analysis and natural language generation tools are not capable of dealing with the textual genres of the Web 2.0, making a lot of errors that could be detrimental for the final summaries.
4. **Selecting the most appropriate tweet:** having on the one hand, the score for each sentences, and on the other hand, the group of sentences that would fit the tweet length, an added value of our proposed approach is to take into account in a joint manner the relevance sentence score, distinguishing them with respect to their polarity, and the potential tweet candidates that fit with the right length. The strategy followed in this stage is to select as tweets the ones that are most

similar to the top-most relevant sentence in each group, but at the same time, not surpassing 140 characters. For achieving this we used the cosine similarity measure, where the tweet candidate whose cosine score is most similar to the sentences that have been identified are more relevant is extracted as final tweet. This works fine for the generation of positive, negative or objective tweets. However, when a mixed tweet needs to be produced, the result of combining subjective and objective information may be ending with a sentence longer than 140 characters. If this happens, we use the same method as in (Lloret et al., 2013) for compressing sentences. In the end, we obtain four tweets (positive, negative, objective, and one containing a mix of subjective/objective information). We therefore propose these four tweets as reliable ultra-concise summaries.

All these stages together with the corpus creation are then included in a semi-automatic sequential process, where basic and intermediate NLP components are used in order to analyse and pre-process the input documents to the summarisation engine.

6 Evaluation and Results

The evaluation of our ultra-concise summarisation approach was carried out by two expert users who evaluated in an independent manner the automatic ultra-concise summaries (i.e. the generated tweets). For performing this evaluation, we relied on the qualitative criteria employed in the TAC conferences⁴. These criteria were: a) the content of the summary, b) its readability and c) its overall responsiveness. The first criterion determines whether the tweet reflects important information of the source inputs; the second assesses whether the tweet is well-written and easy to understand; and the third evaluates if it is reliable and suitable for a real-life application. Evaluating our results with these criteria allows us determining if the product we reach is useful because it is of a high quality, so that no additional treatment is needed. They were evaluated from a conceptual point of view and the general idea expressed, rather than from the individual words building up the summaries.

⁴<http://www.nist.gov/tac/> (last access 30 January 2017)

For each topic (mobile phone and cars) we produce different alternatives as tweets: i) positive information; ii) negative information; iii) objective information; and iv) mixing subjective and objective information. Each of these tweets (40 in total) was rated according to a 3-level Likert scale, with values ranging from 1 to 3 (1=poor or very poor; 2=barely acceptable; and 3=good or very good). The reason for choosing such scale and not a 5-level one was to avoid assessment dispersion. In our case, the agreement between the two assessors was quite high: 60% for content, 75% for readability, and 65% for overall responsiveness criteria, meaning that they both agree in the score assigned to the summary. 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. Table 2 shows the average results obtained for each type of automatically generated tweet. Last two columns provide the global average obtained taking into account all the results.

The results are very encouraging, meaning that our proposed approach can be considered as appropriate for synthesising in one tweet or ultra-concise summary relevant information. Next, we show an example of a potential mixed Tweet with 126 characters that contains subjective and objective information that belong to the group that has obtained the best results:

“Most annoyingly, the alarm is MUCH too quiet. Aside from the alarm though, this is a fab little phone which I highly recommend.”

Generally, the results for each criterion are high, since they are over 2.50, and in most of the cases, very close to 3, the maximum value. Thus, tweets are readable, easy to understand, and reflect relevant information or the key aspects of the phones. If we have a look at the results obtained, we can also notice a better performance in the case of negative sentences. This could be due to the fact that generally negative concepts are expressed in a much more strong and direct way than positive ones, thus are easier to detect and treat from a linguistic point of view. Moreover, we can also appreciate that the objective sentences have lower

	Content		Readability		Over. Resp.		Global	
	Phones	Cars	Phones	Cars	Phones	Cars	Phones	Cars
Positive	2.55	2.70	2.65	2.65	2.55	2.70	2.58	2.68
Negative	2.72	2.50	2.83	2.60	2.76	2.65	2.76	2.58
Objective	2.01	2.35	2.61	2.40	2.22	2.35	2.22	2.37
Mixed	2.75	2.70	2.85	2.80	2.77	2.75	2.77	2.75

Table 2: Average results for the ultra-concise summaries generated with our approach

	Content		Readability		Over. Resp.		Global	
	Phones	Cars	Phones	Cars	Phones	Cars	Phones	Cars
Our approach	2.75	2.70	2.85	2.80	2.77	2.75	2.77	2.75
Swotti	2.22	2.35	2.45	2.50	2.30	2.40	2.32	2.42
Ganesan	1.20	1.30	1.45	1.85	1.25	1.45	1.30	1.48
COMPENDIUMsem	1.75	2.65	1.65	2.75	1.70	2.60	1.70	2.67

Table 3: Comparison results for mixed tweet generation across systems

performance and after having analysed our test set, we reach the conclusion that this is because of the huge amount of advertisement that companies launch in such communication channels, adding noise and influencing in a negative way the retrieval of the sentences and thus the performance of the system. Analysing more in detail the generated tweets, we find some beneficial aspects of our approach and some other aspects with room for improvement. Concerning the positive ones, the modularity of our approach makes it easy to adapt it to other textual genres and languages. Moreover, a strong advantage of our approach is that it is able to produce different types of tweets to be used for different purposes. Each type of tweet can be more or less suitable depending on the users' needs and interests, thus allowing taking into account the user/company's profile. For instance, if a company wants to emphasise a good feature of a product, a positive tweet entry would be the best. In contrast, if they want to improve the weaknesses of their products, a negative tweet could be more appropriate (always seeing this automatic generated tweet as an additional tool to be properly managed by market experts). On the other hand, we also observed some cases in which the automatic tweets generated did not meet our expectations. We found that some original tweets were in other languages than English, such as Spanish or French, and therefore they were counted as incorrect, since we are focusing only in the English language. This was probably due to the fact that the crawler used did not filter the language of the

tweets, because this did not occur for the reviews. However, this stresses the necessity and importance of multilingual approaches that could deal with these challenges and exploit a larger amount of data. Another issue that is worth discussing is the informality, frequently employed in Web 2.0 content. In our case, this was higher in tweets than in on-line reviews, and although, the automatic tweets were not much affected for this, this could be a problematic issue when dealing with highly informal texts.

For both domains tested, the mixed summaries were the ones with best results. To compare our summaries in the form of a tweet with respect to other existing approaches, we took into account the following systems, and we evaluated the results following the same criteria as for the previous evaluation of our approach:

- COMPENDIUMsem (Vodolazova et al., 2012) for determining relevant sentences. This summariser takes into account semantic features, such as concept identification and disambiguation, textual entailment and anaphora resolution.
- The approach proposed in (Ganesan et al., 2012), which is able to produce ultra-concise opinion summaries as well.
- Swotti⁵, which is a commercial system that provides summarised opinion information for

⁵<http://swotti.starmedia.com/> (last access 30 January 2017)

a wide range of products.

Table 3 shows the comparison results of our approach (mixed tweet configuration) with respect to the other systems, where the last two columns provide the global average per row. As it can be seen, our approach obtains the best performance, showing again its appropriateness to be used in real-life applications. From the comparison results, we would like to note that even if we also tested summarisation approaches relying on semantic knowledge the results of these summaries were lower than when using simple word frequency techniques, as in our approach. This confirms the findings in (Inouye and Kalita, 2011), and verifies the power of lexical-based techniques for summarisation. This could be also seen as a competitive method for the generation of ultra-concise summaries. In addition, the results achieved with Swotti, a real on-line platform that provides opinion summarisation, could not surpass the ones obtained with our approach either.

7 Potentials and Limitations of Our Approach

After having described our approach and the results obtained it is worth underlying that, since this the first step of an experimental approach, we decided to take into account only the text. As explained above, all non-conventional characters were deleted. In addition, emoticons have also been removed, however our idea is to take them into account for next steps, since they are usually charged with polarity that could also be an added value for the correct text interpretation.

The retrieval and pre processing stages confirmed the fact that the text available in the Social Media is extremely informal and it does not always follow the conventional rules. Many are the cases of informal languages that contain sayings or collocations hard to interpret automatically. They will be taken into account in future work, since the EmotiBlog annotation model is able to capture and interpret them in terms of polarity. In addition to semantic challenges, the issue of informality is a main concern. Last but not least, the fact of working with different textual genres poses the problem of having to deal with different linguistic phenomena, typical in one or other genre. The result of this is that, despite the satisfactory performance we also obtained cases with room for improvement. Example of this are:

- *Help please?!*
- *2.5mm Standard Headset Adapter for LG GW520 [Wireless Phone Accessory]: A top quality standard headset adapter (... <http://amzn.to/wnzLVP>)*
- *Easy AdSense Pro by Unreal Here is download link LG KS360/ KS365 PC suite/Sync/USB Driver for <http://goo.gl/fb/uxV5o>*

From the examples above, we can deduce that in the cases in which the system performance was not satisfactory, we obtain sentences with no relevant content in most of the cases. In all the cases in which the system performance was satisfactory, examples of sentences are the following ones:

- *The only annoying thing I do find is all the Blackberry apps are geared up for the USA, not UK, which limits their desirability.*
- *The battery life is good and will last me 2-3 days if I am careful.*
- *All the Blackberry apps are geared up for the USA. The internet drains is pretty quickly. The battery is good and will last me 2-3 days.*

As it can be seen, the content is relevant to the topic search. In addition to this, in the third case we obtain a mixed tweet, which includes different ideas taken from the positive, negative and objective sentences. This is very important to maintain the relevance of the content and thus the quality and reliability of the system output.

8 Conclusion and Future Work

In this article, we presented an innovative method for creating an effective system that produces ultra-concise summaries with no more than 140 characters (i.e., in the form of a tweet). The approach represents an advancement of the state-of-the-art approach since we work with multiple textual genres simultaneously and produce a reliable ultra concise summary. We start from the information contained in a selected corpus composed of on-line reviews and microblogs written in English about “mobile phones and cars” and more concretely we selected a set of 10 cell phones and 10 cars of different brands. We presented our corpus and the annotation process carried out. For the corpus annotation, we applied the coarse grained version of the EmotiBlog annotation schema that is

a fine-grained annotation schema for the detection of the subjectivity in the new textual genres of the Web 2.0. We took advantage of such annotation to propose a novel summarisation approach that employed statistical and linguistic techniques for detecting relevant sentences. We generated four types of ultra-concise summaries (or tweets) that were then evaluated following a standard qualitative framework, and compared to similar approaches. From the results obtained we conclude that our approach is reliable and appropriate for generating this types of summaries. The successful performance of the tweets (in terms of content reliability and syntactic adequacy) clearly indicates that they can be used within a real-life application. The selection of which tweet to show is out of the scope of this paper, and this would depend on the target users and purpose of the tweet. However, one strategy that could be adopted would be to rely on existing automatic rating models, such as EmotiReview (Boldrini et al., 2011). This manner depending on the rating given to a specific product, we could decide which type of tweet should be generated.

Despite the good results achieved, there are several issues that have to be tackled for improving the generation of ultra-concise summaries and that we plan to tackle as future work. On the one hand, in the short-term, we will mainly focus on two aspects: the multilinguality and the inclusion of more Web 2.0 textual genres and domains. This manner, we will be able to extend our approach to languages, such as Italian or Spanish, as well as to deal with other type of texts, such as fora, or blogs and increase the domains such as tourism, politics, etc. Moreover, we want to analyse in more detail the impact of each proposed stage in the summarisation process, as well as the influence of each textual genre. In this manner, we will substitute the manual annotation for sentiment analysis for an automatic one as well as we will analyse more features for sentiment analysis (e.g. target, intensity, etc.), and we will test other summarisation techniques. On the other hand, for the medium and long-term research, we will increase the size of the annotated corpus, and we will use it to train a machine learning system that automatically detects and classifies the objective/subjective information. In addition, a topic detection stage able to identify concepts and their relationships would be necessary in order to personalise the re-

sulting summaries. Other important issues to take into account will be the informality of the text and some sentiment-analysis-related phenomena, such as irony, that were out of the scope of the paper, due to its great difficulty. The informality of a text could be detected and used to normalize the texts using for instance the TENOR tool (Mosquera and Moreda, 2012), or vice versa, to produce an informal summary in the form of a tweet. For detecting ironic expressions, we could also rely on already existing approaches, such as the one described in (Reyes et al., 2012).

Acknowledgments

This research work has been partially funded by the Generalitat Valenciana and the Spanish Government through the projects DIIM2.0 (PROMETEOII/2014/001), TIN2015-65100-R and TIN2015-65136-C2-2-R.

References

- Alexandra Balahur, Elena Lloret, Ester Boldrini, Andrés Montoyo, Manuel Palomar, and Patricio Martínez-Barco. 2009. Summarizing threads in blogs using opinion polarity. In *Proceedings of the Workshop on Events in Emerging Text Types*, pages 23–31, Borovets, Bulgaria, September. Association for Computational Linguistics.
- Ester Boldrini, Alexandra Balahur, Patricio Martínez-Barco, and Andrés Montoyo. 2010. Emotiblog: A finer-grained and more precise learning of subjectivity expression models. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 1–10, Uppsala, Sweden, July. Association for Computational Linguistics.
- Ester Boldrini, Javier Fernández, José Manuel Gómez, and Patricio Martínez-Barco. 2011. Emotiblog: Towards a finer-grained sentiment analysis and its application to opinion mining. In *Proceedings of IV Jornadas TIMM*, pages 49–54.
- Ester Boldrini. 2012. *EmotiBlog: A Model to Learn Subjective Information Detection in the New Textual Genres of the Web 2.0 -a Multilingual and Multi-Genre Approach*. Ph.D. thesis, University of Alicante.
- Deepayan Chakrabarti and Kunal Punera. 2011. Event summarization using tweets. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*.
- Bruce Croft, Donald Metzler, and Trevor Strohman. 2009. *Search Engines: Information Retrieval in Practice*. Addison-Wesley Publishing Company, USA, 1st edition.

- Wenjing Duan, Bin Gu, and Andrew B. Whinston. 2008. Do Online Reviews Matter? - An Empirical Investigation of Panel Data. *Decision Support Systems*, 45(4):1007–1016, November.
- Javi Fernández, José Manuel Gómez, and Patricio Martínez Barco. 2010. Evaluation of web information retrieval systems on restricted domains. *Procesamiento del Lenguaje Natural*, 45:273–276.
- Kavita Ganesan, ChengXiang Zhai, and Evelyne Viegas. 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.
- Talmy Givón. 1990. *Syntax: A functional-typological introduction. Volume II. Amsterdam: John Benjamins. XXV, 553 pages. (Volume 1: 1984 (XX, 464 pages).*
- Wu He, Harris Wu, Gongjun Yan, Vasudeva Akula, and Jiancheng Shen. 2015. A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, 52(7):801–812.
- David Inouye and Jugal K. Kalita. 2011. Comparing twitter summarization algorithms for multiple post summaries. In *PASSAT/SocialCom 2011, Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom), Boston, MA, USA, 9-11 Oct., 2011*, pages 298–306.
- B.J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. 2009. Twitter power: Tweets as electronic word of mouth. *Journal of American Society of Information Science*, 60(11):2169–2188, November.
- Tae-Yeon Kim, Jaekwang Kim, Jaedong Lee, and Jee-Hyong Lee. 2014. A tweet summarization method based on a keyword graph. In *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, ICUIMC '14*, pages 96:1–96:8, New York, NY, USA. ACM.
- Fei Liu, Yang Liu, and Fuliang Weng. 2011. Why is "sxsw" trending?: Exploring multiple text sources for twitter topic summarization. In *Proceedings of the Workshop on Languages in Social Media, LSM '11*, pages 66–75, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Elena Lloret and Manuel Palomar. 2013. COMPENDIUM: a text summarisation tool for generating summaries of multiple purposes, domains, and genres. *Natural Language Engineering*, 19:147–186, 4.
- Elena Lloret, María Teresa Romá-Ferri, and Manuel Palomar. 2013. COMPENDIUM: A text summarization system for generating abstracts of research papers. *Data & Knowledge Engineering*, 88:164–175.
- Alejandro Mosquera and Paloma Moreda. 2012. Tenor: A lexical normalisation tool for spanish web 2.0 texts. In Petr Sojka, Ales Hork, Ivan Kopeček, and Karel Pala, editors, *Text, Speech and Dialogue - 15th International Conference, TSD 2012, Brno, Czech Republic, September 3-7, 2012. Proceedings*, volume 7499 of *Lecture Notes in Computer Science*, pages 535–542. Springer.
- Ani Nenkova and Kathleen McKeown. 2011. Automatic summarization. *Foundations and Trends in Information Retrieval*, 5(2-3):103–233.
- Brendan O'Connor, Michel Krieger, and David Ahn. 2010. Tweetmotif: Exploratory search and topic summarization for twitter. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*.
- O. Pacea. 2011. CorpTweet: Brands, Language and Identity in Web 2.0. *SERIA FILOLOGIE*, 22(2):157–168.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*., 2(1-2):1–135, January.
- Antonio Reyes, Paolo Rosso, and Davide Buscaldi. 2012. From Humor Recognition to Irony Detection: The Figurative Language of Social Media. *Data & Knowledge Engineering*, 74:1–12.
- Beaux Sharifi, Mark-Anthony Hutton, and Jugal Kalita. 2010. Summarizing microblogs automatically. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 685–688, Los Angeles, California, June. Association for Computational Linguistics.
- A. Smith and J. Brennen. 2012. Twitter use 2012. technical report. Technical report.
- Tatiana Vodolazova, Elena Lloret, Rafael Muñoz, and Manuel Palomar. 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.
- Jui-Yu Weng, Cheng-Lun Yang, Bo-Nian Chen, Yen-Kai Wang, and Shou-De Lin. 2011. IMASS: An intelligent microblog analysis and summarization system. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 133–138, Portland, Oregon, June. Association for Computational Linguistics.