Introducing the Key Stages for Addressing Multi-perspective Summarization

Elena Lloret
Department of Software and Computing Systems, University of Alicante, Apdo. de Correos 99, E-03080, Alicante, Spain

Keywords: Natural Language Processing, Text Summarization, Multi-perspective Summarizaiton, Opinion Mining, Topic Identification, Knowledge Discovery.

Abstract: Generating summaries from evaluative text (e.g., reviews) is a challenging task, in which available metadata is hardly exploited, thus leading to the creation of very generic and biased summaries. In this paper, the novel task of multi-perspective summarization is introduced. The key stages for generating this type of summaries are defined, and a preliminary analysis of their feasibility is conducted. The main novelties of this study include: i) the linguistic treatment of the text at the level of basic information units, instead of sentences; and ii) the analysis carried out over the distribution of opinions and topics.

1 INTRODUCTION

The Web 2.0 has allowed users to become the main actors on the Internet, by proactively expressing their comments/opinions, sharing their experiences, complaining about a service or product, interacting with other users, etc. In addition, it has been established as a means of supporting decision-making processes. However, being able to make a well-informed decision is not a trivial task, since it would imply reading and processing millions of documents, which may be not feasible for a human being in terms of time and processing cost. In this respect, text summarization can help users by automatically identifying relevant information, highlighting and outlining the advantages and disadvantages of a service/product, as well as providing the gist of a document or collection of documents.

Specifically, when producing summaries from reviews, existing summarization approaches only take into account the polarity of the sentences (whether they are positive, negative or neutral), or the topic (also known as aspect), generating multi-aspect based summaries (Gerani et al., 2014). However, this type of text, like any evaluative text, has a strong subjective component and its content depends on the specific experience a user had. For instance, the same hotel may be considered as good for a type of user (e.g., business person) and not good for another (e.g., families), as it can be seen from the following real extracts[1], respectively:

"Hotel located in quiet area near Vatican. It is a hotel that is not in the center but is very close…"

"Very far from the center. Regarding this hotel, the most important criticism is HOW FAR it is from the center of Rome…"

This issue makes the summarization task even more challenging, because it is very difficult to determine which opinion should be considered as important and appropriate to be included in the final summary. This could be partially solved by exploiting the metadata associated to a text which can provide useful information not explicitly included in the documents (e.g., traveller type, specific ratings, user nationality, etc.). Following the previous examples, if we had used the information about the traveller type related to the comment, and having known whether the opinion about the hotel was written by a family or by a business person, the summary could have either provided a general overview indicating also the profile behind the statement, or it could have been personalized depending on the target readership. This leads to a new type of summaries: multi-perspective summaries, that attempt to provide an overview of a topic from different points of views.

The goal of this paper is to introduce the task of multi-perspective summarization by defining and

[1]These are real opinions extracted from TripAdvisor: https://www.tripadvisor.com/

Lloret, E.
Introducing the Key Stages for Addressing Multi-perspective Summarization.
DOI: 10.5220/0006066603210326
In *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) - Volume 1: KDIR*, pages 321-326
ISBN: 978-989-758-203-5

proposing the key elements that should be involved in the design and development of such an approach. The preliminary analysis carried out provided insights about the development of a plausible approach, as well as allowing the identification of the challenges that need to be addressed.

2 RELATED WORK

When producing summaries from reviews, the most common strategy adopted in the literature is to perform multi-aspect review summarization, taking into account the different aspects discussed in a review (e.g., location). In (Ly et al., 2011), the different features expressed in product reviews are called facets. Once determined, the summarization approach is restricted only to those opinionated sentences, obtained through an opinion mining approach. These opinion sentences are clustered based on their content similarity. Finally, the most representative sentence for each group (i.e., the centroid of the cluster) is selected to form the summary. SumView (Wang et al., 2013) was developed as a graphical interface for product review summarization that was more focused on the summarization stage rather than in the opinion mining one. For determining the relevant sentences, a feature-based weighted non-negative matrix factorization model was used, outperforming well-known summarization methods (e.g., those based on PageRank). Different from the existing approaches, (Gerani et al., 2014) was pioneering in generating abstractive summaries from reviews. The authors proposed a full natural language generation pipeline for guiding the process of summary production, which exploits the discourse structure of the reviews by means of rhetorical relations, obtaining good summaries. In the existing approaches, the exploitation of metadata is not frequent, despite its usefulness for aiding in the selection of relevant sentences. Only few works that took into account metadata were found (Kokkoras et al., 2008), (Dubey and Kumar, 2008), where the relevance score associated to a sentence is modified (normally increased) according to the contribution of specific metadata (e.g., the number of users that found a review useful).

Finally, concerning the multi-perspective summaries, this is a novel task in the context of summarization. To the best of our knowledge, the only previous work related to our proposed research is the one presented in (Bouayad-Agha et al., 2012), where a base ontology was designed in order to represent the information about a football match and then generate a text from key information in the ontology, but taking into account two perspectives: the winner and the loser team. Although the use of an ontology may be useful for representing knowledge, it has two main drawbacks: i) the cost of designing and populating it; and ii) it is domain-specific, and therefore it cannot be adapted to other domains and scenarios.

3 DEFINITION OF KEY STAGES

In this section, we define the crucial stages that would be necessary to develop a multi-perspective summarization approach. Given a set of reviews with metadata information (e.g., the type of traveller), the approach shall produce a summary that could be adapted and personalized with respect to the user needs or settings.

3.1 Text Processing

This module carries out the linguistic analysis of the input texts. Specifically, it is divided in three stages:

- **Basic Information Units (BIUs) Identification.** A BIU is our linguistic working unit. It is represented by a fragment of text, that could range from a complete sentence to clauses, phrases, triples or even keywords. In this stage, the goal is to split the input text into the desired BIUs and extract them.
- **Topic Detection.** The aim of this stage is to identify the topic associated to a BIU, i.e., about what is being talked in the BIU. This will be useful for further stages to determine whether different BIUs are discussing the same issue or not.
- **Polarity Detection.** Finally, a process able to distinguish between subjective and objective information, as well as to classify it into positive, negative and neutral is crucial.

3.2 Semantic Content Analysis

This module plays a key role in the process, since it is responsible for identifying common and contradictory BIUs, focusing on those ones that express a subjective opinion.

- **Common BIUs Identification.** We consider that two BIUs are expressing the same information if the associated topics are the same, and they both have the same polarity.
- **Contradictory BIUs Identification.** Two BIUs will be considered contradictory, if their associated topics are the same, but they have opposite polarity (positive vs. negative).

3.3 Summary Generation

This module is in charge of producing the final summary. In this module, two intermediate stages are designed:

- **Settings Selection.** Given that we have the BIUs classified at different levels (topic, polarity and metadata), different types of multi-perspective summaries could be generated. Therefore, in this stage, the summary could be oriented to a single aspect, for instance, the topic, the polarity or the obtained metadata, or a combination of them. This would be very useful to create adaptive and personalized summaries.
- **Template-based Generation.** The technique of template-based generation is commonly used for abstractive summarization (Oya et al., 2014; Gerani et al., 2014). Therefore, we foresee a template-based generation proposes the dynamic combination of linking phrases (e.g., *"The most discussed topics about..."*; *"The aspects positively rated include..."*; *''Among the reasons for supporting their claims we found:"*) with the BIUs themselves and the information associated to them.

These two stages are combined at the same time, in the sense that the decision of the type of summary to be created will directly determine the kind of templates to be used.

4 EXPERIMENTAL ENVIRONMENT

An initial study of the key stages previously defined was carried out in the context of on-line hotel reviews, although the proposed method can be easily adapted to other domains and languages as long as the necessary natural language processing tools were available.

4.1 Corpus Retrieval and Analysis

Our experiments were focused on a corpus of hotel reviews extracted from TripAdvisor[2], which also provides metadata information, such as the global rating, the type of traveller, and the user location, among others. Specifically, we decided to use the type of traveller, since this manner we could analyze different points of view with respect to the same hotel. TripAdvisor classifies the types of travellers into four groups: families, business, couples and solo.

[2]https://www.tripadvisor.es/

For the experiments, 10 hotels in Rome were selected. Before gathering the reviews, a preliminary study was conducted, in order to ensure that: i) the hotels selected had enough number of reviews for all the traveller types (hotels with less than 500 reviews were discarded); and ii) there was a balance between the user ratings, so positive as well as negative opinions could be found to make it possible the existence of opinion variety and contrast. With these constraints, a crawler was developed to automatically collect the content for each review together with the metadata concerning the traveller type. More than 6,500 reviews (1,243 associated to families; 2,364 for couples; 324 for solo; and 871 for business) were initially obtained. Since this was a great number of reviews to be able to carry out a detailed analysis and evaluation of all the intermediate stages of the multi-perspective summarization approach, a representative sample for each group was extracted to work only with a subportion of this corpus. For this, the formula described in equation 1 was employed (Pita Fernández, 1996):

$$M = \frac{N * K^2 * P * Q}{E^2 * (N-1) + K^2 * P * Q} \quad (1)$$

where M is the number of samples to extract, N is the population, K is confidence interval, E is the error rate, P is the probability of success and Q is the probability of failure. The value for each parameter was set taking into account what was suggested in (Gutiérrez Vázquez et al., 2011):$K = 0.95; E = 0.05; P = 0.5; Q = 0.5$. Table 1 shows the final number of reviews per traveller group used for the experiments, having a total of 1,808 reviews.

Table 1: Number of reviews (Id refers to the hotel id).

Id	Families	Couples	Solo	Business
1	48	51	19	62
2	53	67	30	55
3	50	66	17	46
4	55	65	21	26
5	52	68	33	15
6	59	67	30	54
7	49	73	26	38
8	55	62	25	36
9	47	67	20	46
10	54	58	14	30

4.2 Tools and Experiments

All the reviews written for the same hotel were passed through the text processing stage, explained in Section 3. Therefore, the internal representation of the reviews consisted of a set of tuples with four elements:

$$< BIU \| topic \| polarity \| traveller\_group >$$

For extracting them, only a parser and a sentiment analyzer were used. On the one hand, Stanford Parser (Klein and Manning, 2003) was used for the identification and extraction of BIUs, since it has been shown a very good performance and it allows to do the parsing in different languages. Different granularity levels were established for detecting BIUs: i) complete sentences (level 0); ii) top-level clauses (level 1); and iii) second-level clauses (level 2). In this manner, ii) and iii) could be more appropriate for abstractive summarization, since, meaningful chunks of sentences will be extracted. Moreover, we took profit of the annotations provided by the Stanford Parser about the nucleus of a sentence/clause, considering at the moment the main nouns involved in them as topics. On the other hand, once the BIUs were identified and extracted, they were passed through a polarity detection and classification process, using the Sentiment Analysis tool developed in (Fernández et al., 2015). Finally, the traveller group was directly obtained at the crawling stage, as part of the metadata associated to the reviews.

Several examples of instances of different granularities are shown in Figure 1. It is worth stressing upon the fact the the process is fully automatic, so the errors made by the linguistic and semantic tools may influence the quality of the output. During the process, we made several decisions, such as discarding the BIUs for which the topic could not be identified.

Level 0:
The lobby is minimalist but quite large and modern. | |lobby| |positive| |Business
A bit shabby, and the window did not close. | |window| |negative| |Business
Level 1:
The only real problem with this hotel is its location | |hotel,location| |negative| |Couple
The location is not great, | |location| |negative| |Families
Level 2:
the service is impeccable , | |service| |positive| |Business
the service is terrible | |service| |negative| |Families

Figure 1: Examples of extracted BIUs for different granularity levels for hotel id 6.

For the semantic content analysis stage, both the common and contradictory BIUs were identified using a rule-based process, that compared the topic and the polarity obtained for each BIU.

5 ANALYSIS AND DISCUSSION

A qualitative analysis for the results obtained at the different intermediate key stages of the multi-perspective summarization approach are presented and discussed. This manner, we could determine their feasibility and appropriateness, outlining the potentials and limitations found in each step. The current assessment of the whole summarization method would be still too early since it would lead to ungrammatical summaries due to the direct insertion and combination of the with the linking phrases of the templates.

Table 2: Identified BIUs.

Id	Level 0	Level 1	Level 2
1	1201	1029	588
2	1655	1498	833
3	4473	4349	2385
4	3498	3199	1547
5	1540	1412	761
6	4714	4323	2526
7	4159	4139	2254
8	4415	4212	2302
9	4923	4758	2841
10	1893	1759	824

Table 2 shows the total number of identified and extracted BIUs from the original reviews for the three granularity linguistic levels experimented with. As it can be seen, the larger number of BIUs are detected when full sentences are taken into account. One could think that finer granularity levels that work at the clause-level should provide a larger number of BIUs, since a sentence may be composed of several clauses. However, the obtained results show exactly the oppositve: the finer level of granularity, the lower number of BIUs for all the selected hotels. This was due to the fact that not all sentences could be split into clauses according to the parser, and if this happened, the sentence was not taken into account. Possible solutions would be to include the full sentence if it does not contain any clause; test other parsers, or use a chunker. Nevertheless, we still got a sufficient number of BIUs, so we decided to continue evaluating the next stages of the approach.

Table 3: Polarity classification for the identified BIUs at each granularity level (Pos= positive; Neg=negative; Ntr= neutral).

	Level 0			Level 1			Level 2		
Id	Pos	Neg	Ntr	Pos	Neg	Ntr	Pos	Neg	Ntr
1	669	293	239	437	252	340	218	142	228
2	907	423	325	607	376	515	277	201	355
3	2507	1174	792	2005	1009	1335	833	522	1030
4	2116	722	660	1476	608	1115	550	284	713
5	863	371	306	615	320	477	268	172	321
6	2255	1391	1068	1646	1214	1463	815	617	1094
7	2426	918	815	1891	821	1427	836	474	944
8	2313	1187	915	1770	986	1456	814	562	926
9	2243	1574	1106	1809	1398	1551	787	769	1285
10	856	546	491	607	514	638	255	198	371

Concerning the analysis of polarity detection, Table 3 reports the statistics related to the number of positive, negative and neutral BIUs identified at each

Table 4: Top 20 most frequent topics shared among all traveller types and common to the three granularity linguistic BIUs levels.

Id	Topics
1	**area**, return, life, trip, time, Vatican, holidays, **location**, transport, traffic, tour, toast, towels, type, weather, taxi, supermarket, floor, services, noise
2	**area**, **wifi**, time, lobby, Vatican, **location**, TV, treatment, transport, job, shops, taxi, size, place, bus, suite, services, reception, property, price
3	juice, **area**, trainers, **wifi**, return, volume, peek, view, visitant, wine, villa, trip, traveller, lobby, windows, vegetarian, neighbourhood, time, Vatican, **location**
4	juice, **area**, yoghurt, **wifi**, return, flight, Vittorio, view, visit, villa, glass, trip, traveller, road, time, truth, windows, Venice, neighbourhood, Vatican
5	**area**, **wifi**, return, trip, road, Summer, windows, times, weather, Vatican, variety, **location**, train, transport, Termini, ceiling, sheet, Rome, restaurant, reception
6	**area**, shoes, yoghurt, **wifi**, vision, virus, time, lobby, garbage, windows, advantage, neighbours, Vatican, variety, holidays, **location**, TV, tourism, transport, towels
7	juice, **area**, ham, yoghurt, gypsum, **wifi**, Web, return, flights, views, wine, villa, road, trip, lobby, truth, Summer, window, time, peek
8	juice, **area**, yoghurt, **wifi**, Web, WC, flight, peek, view, visit, trip, time, lobby, ventilation, windows, vehicle, Vatican, variety, **location**, tourist
9	juice, **area**, **wifi**, walk, return, voice, view, trip, traveller, time, lobby, truth, Summer, windows, sale, variety, utility, **location**, cloths, transport
10	**area**, **wifi**, view, trip, traveller, time, windows, Vatican, variety, holidays, **location**, towels, terrace, TV, place, insurance, room, Rome, restaurant, price

granularity level. As it can be seen, for level 0 (i.e., when complete sentences are used as BIUs), the positive opinons prevail over the negative and neutral ones for all the analyzed hotels, being the neutral opinions the less frequent ones. For the remaining granularity levels, an increase of neutral opinions is observed, being higher than the negative ones for level 1 and even higher than both, the negatives and positive, for level 2. This is explained due to the sentence segmentation done at a clause level, where short fragments included in long positive sentence lead to neutral statements. For instance, the sentence*"Regarding the dinner, it was a wonderful experience ... The restaurant is on the roof, and one can admire a spectacular view of the entire city from it, with the Basilica of San Pedro in the background illuminated."* was detected as positive at level 0, but after its segmentation, some fragments were detected as neutral in levels 1 and 2, such as *"Regarding the dinner"*, or *"The restaurant is on the roof"*.

From the analysis performed, it seems that in general reviews tend to have more positive comments than negative. As it can be seen, there is also a considerable number of neutral opinions, where the users narrate their personal experience giving objective facts and data (e.g., *"the room is equipped with a safe box"*). Finally, it is also worth noting that the polarity for some of the sentence may be wrongly detected, due to performance errors of the tool, which has to face the challenging task of detecting the interpretation/intention behind the meaning (e.g., *"the hotel is out of the city centre"* is detected as neutral, but the user stating such comment may want to indicate a negative issue).

In addition to the analysis of the polarity, a study of the most frequent common topics that were present in the BIUs and shared between all traveller types was also conducted. A topic detection process was applied for each granularity level, obtaining the top 20 most frequent topics of the BIUs. However, in light that a great portion of the topics was common for all the granularity levels, Table 4 only shows the non-repeated ones for each hotel. This analysis allowed us to compare different topics across all the hotels. From the results obtained, it is interesting to note that there are a considerable number of common aspects, normally mentioned about a hotel, that include aspects like *area*, *wifi*, or *location*, regardless of the name and type of hotel.

The last analysis carried out was concerning the presence of common and contradictory opinions ac-

Table 5: Results for the semantic content analysis module (% of common -*Comm*- and contradictory -*Contr*- BIUs).

	Level 0		Level 1		Level 2	
Id	**Contr**	**Comm**	**Contr**	**Comm**	**Contr**	**Comm**
1	100%	0%	80%	20%	80%	20%
2	70%	30%	80%	20%	70%	30%
3	50%	50%	35%	65%	30%	70%
4	50%	50%	40%	60%	15%	85%
5	85%	15%	70%	30%	45%	55%
6	70%	30%	70%	30%	50%	50%
7	40%	60%	40%	60%	15%	85%
8	40%	60%	20%	80%	30%	70%
9	50%	50%	65%	35%	45%	55%
10	70%	30%	75%	25%	50%	50%

cording to the rules defined for the proposed approach (see Section 3.2). The BIUs associated to the top 20 most frequent topics were extracted, and were annotated as contradictory or common with respect to these topics. Table 5 reports the results obtained for each granularity level. As it can be observed, for all the hotels different perspectives can be found with respect to an aspect. For instance, for the same hotel we find BIUs reporting information about the hotel, like *"in general, the hotel seemed very good to me"*, *"the hotel was great"*, or *"we regretted booking the hotel"*. The first two share the same positive opinion, whereas the last one show the opposite, so this sentence contradicts the other two. This confirms our expectations and highlights the need to take this issue into account in order to be able to create better and precise summaries. In our results, when the text is treated at a sentence level, the number of contradictory opinions increases with respect to the common ones. However, this decreases when finer granularity levels are employed, where it gets more balanced.

6 CONCLUSIONS AND FUTURE WORK

In this paper, the task of multi-perspective summarization was introduced. The key stages of such an approach were defined and analyzed in the context of hotel reviews. One of its relevant novelties is the definition of BIUs that allow the text segmentation into smaller meaningful information units than sentences. The analysis carried out through all the stages in the process shows the potentials of the approach concerning its versatility and flexibility, but it also outlines its current limitations, leading to a lot of room for improvement.

In the future, several directions are planned: i) to develop a better method to perform topic detection and BIUs clustering; ii) to define and test a machine learning method to better identify contradictory and common BIUs ; and iii) to substitute the template-based method by natural language generation techniques to create fully abstractive summaries.

ACKNOWLEDGEMENTS

This research work has been funded by the University of Alicante through the project "Explotación y tratamiento de la información disponible en Internet para la anotación y generación de textos adaptados al usuario" (GRE13-15). Moreover, it has been partially funded by the Generalitat Valenciana, Spanish Government and the European Commission through the projects PROMETEOII/2014/001; TIN2015-65100-R, TIN2015-65136-C2-2-R; and FP7-611312.

REFERENCES

Bouayad-Agha, N., Casamayor, G., Mille, S., and Wanner, L. (2012). Perspective-oriented generation of football match summaries: Old tasks, new challenges. *ACM Trans. Speech Lang. Process.*, 9(2):3:1–3:31.

Dubey, A. and Kumar, A. (2008). A technique for summarizing web reviews. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '08, pages 711–717.

Fernández, J., Gutiérrez, Y., Tomás, D., Gómez, J. M., and Martínez-Barco, P. (2015). Evaluating a sentiment analysis approach from a business point of view. In *Proceedings of TASS 2015: Workshop on Sentiment Analysis co-located with 31st SEPLN*, pages 93–98.

Gerani, S., Mehdad, Y., Carenini, G., Ng, R. T., and Nejat, B. (2014). Abstractive summarization of product reviews using discourse structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1602–1613.

Gutiérrez Vázquez, Y., Fernández Orquín, A., Montoyo Guijarro, A., and Vázquez Pérez, S. (2011). Integración de recursos semánticos basados en WordNet. *Procesamiento del Lenguaje Natural*, 47:161–168.

Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pages 423–430.

Kokkoras, F., Lampridou, E., Ntonas, K., and Vlahavas, I. (2008). Mopis: A multiple opinion summarizer. In *Artificial Intelligence: Theories, Models and Applications*, volume 5138 of *Lecture Notes in Computer Science*, pages 110–122.

Ly, D. K., Sugiyama, K., Lin, Z., and Kan, M.-Y. (2011). Product review summarization from a deeper perspective. In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, pages 311–314.

Oya, T., Mehdad, Y., Carenini, G., and Ng, R. (2014). A template-based abstractive meeting summarization: Leveraging summary and source text relationships. In *Proceedings of the 8th International Natural Language Generation Conference*, pages 45–53.

Pita Fernández, S. (1996). Determinación del tamaño muestral. *CAD ATEN PRIMARIA 1996*, 3:114–138.

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.