Combining Text Summarization and Aspect-based Sentiment Analysis of Users' Reviews to Justify Recommendations

Cataldo Musto
University of Bari 'Aldo Moro'
cataldo.musto@uniba.it

Gaetano Rossiello
University of Bari 'Aldo Moro'
gaetano.rossiello@uniba.it

Marco de Gemmis
University of Bari 'Aldo Moro'
marco.degemmis@uniba.it

Pasquale Lops
University of Bari 'Aldo Moro'
pasquale.lops@uniba.it

Giovanni Semeraro
University of Bari 'Aldo Moro'
giovanni.semeraro@uniba.it

ABSTRACT

In this paper we present a methodology to *justify* recommendations that relies on the information extracted from *users' reviews* discussing the available items. The intuition behind the approach is to conceive the justification as a *summary* of the most relevant and distinguishing aspects of the item, automatically obtained by analyzing its reviews. To this end, we designed a pipeline of natural language processing techniques including *aspect extraction, sentiment analysis and text summarization* to gather the reviews, process the relevant excerpts, and generate a unique *synthesis* presenting the main characteristics of the item. Such a summary is finally presented to the target user as a *justification* of the received recommendation. In the experimental evaluation we carried out a user study in the *movie* domain (N=141) and the results showed that our approach is able to make the recommendation process more transparent, engaging and trustful for the users.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Recommender Systems; Explanation; Text Summarization; Sentiment Analysis

ACM Reference Format:
Cataldo Musto, Gaetano Rossiello, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2019. Combining Text Summarization and Aspect-based Sentiment Analysis of Users' Reviews to Justify Recommendations. In *Thirteenth ACM Conference on Recommender Systems (RecSys '19), September 16–20, 2019, Copenhagen, Denmark.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3298689.3347024

1 BACKGROUND AND MOTIVATIONS

A research line which is recently emerging in the context of the so-called *Data Explosion* [18] regards the analysis and the exploitation of the information extracted from user-generated content and, in particular, from *users' reviews*. The positive impact of the information extracted from users' reviews has been already acknowledged in many scenarios [1, 9]: as an example, several work showed that *review-aware recommendation models* tend to beat classical recommendation algorithms [4, 6, 7, 12].

Differently from most of the current literature, in this work we investigated the impact of the information conveyed by users' reviews in a different scenario, that is to say, the task of *justifying* a suggestion returned by a recommendation algorithm. In a nutshell, our methodology relies on the idea that an effective justification can be conceived as a *summary* of relevant and distinguishing aspects of the item, automatically obtained by analyzing its reviews. To this end, we designed a pipeline of natural language processing techniques to obtain and process reviews excerpts and to generate a unique *synthesis* presenting the main characteristics of the item, which is finally presented to the target user as *justification* of the received recommendation.

It is worth to note that we preferred to use the term *justification* over *explanation* since, as stated by Biran et al. [2], a *justification explains why a decision is a good one, without explaining exactly how it was made.* Accordingly, the strategy we implemented is closer to the idea of justification rather than explanation, since our methodology is more devoted to describe *why* a user would be interested in the item, in order to make a more *informed decision* about consuming the item or not.

The idea of exploiting users' reviews to explain or justify the recommendations was already investigated by Muhammad et al. [11]. However, differently from our work, the authors here focus on the identification of relevant aspects of the items (*e.g.*, food, service, etc.), without building a comprehensive natural language justification. A more sophisticated attempt is due to Chang et al., who propose in [3] a crowd-based natural language explanation based on review data. In this case, our approach distinguishes from this work since we exploited an automated natural language processing pipeline, while in [3] most of the work is carried out by a set of workers on a crowd-sourcing platform, who manually annotated the sentences to be included in the explanations.

2 METHODOLOGY

Our methodology to build *natural language justifications* by exploiting users' reviews relies on four steps: *aspect extraction*, *aspect ranking*, *sentence filtering* and *text summarization.*

Aspect Extraction. First, we assume that an effective justification should include relevant and distinguishing traits of the items,

RecSys '19, September 16–20, 2019, Copenhagen, Denmark

ACM ISBN 978-1-4503-6243-6/19/09…$15.00
https://doi.org/10.1145/3298689.3347024

which are often discussed in the reviews (with a *positive* sentiment, of course). Accordingly, the goal of the first phase is identifying the aspects that are worth to be included in the justification. Formally, our strategy takes as input a set of reviews $R = \{r_1, r_2 \dots r_n\}$ and produces a set of 4-tuples $r_i, a_{ij}, rel(a_{ij}, r_i), sent(a_{ij}, r_i)$ representing the review r_i, the *j-th* aspect a_{ij} extracted from the review r_i, the relevance $rel(a_{ij}, r_i)$ of the aspect a_{ij} in the review r_i, and the sentiment associated to that aspect.

To extract *aspects* from reviews, we used the *Kullback-Leibler divergence* [8] (KL-divergence, referred to as δ), a non-symmetric measure of the difference between two distributions. Formally, given two corpora c_a and c_b and a term t, pointwise KL-divergence is calculated as:

$$\delta_t(c_a || c_b) = p(t, c_a) log \frac{p(t, c_a)}{p(t, c_b)} \quad (1)$$

Where $p(t, c_a)$ is the number of occurrences of the term t in corpus c_a. In a nutshell, this measure rely on the idea that the use of language differs when talking about a specific domain with respect to a general topic, thus this method identifies the aspects whose distribution in a specific domain (e.g., *movie reviews*) diverges from that in a general corpus (e.g., the British National Corpus BNC[1]). In other terms, the measure identifies those aspects which are mentioned in the reviews *more often* than usual.

In our case, given a review r_i, we first extract all the *nouns* by running a POS-tagging algorithm. Next, we calculate KL-divergence for each noun by using as corpora our set of movie reviews and the BNC. The nouns having a KL-divergence greater then a threshold ϵ are labeled as *aspects* and the KL-divergence score is used as relevance score. Finally, we also run a sentiment analysis algorithm to identify the opinion is associated to the aspect a_{ij} in r_i and we stored that value. More details on the algorithms used in this phase are provided in the next section.

Aspect Ranking. Aspect extraction is run over all the reviews contained in R, and a set of aspects is extracted from each review. However, the goal of our strategy is to identify the most relevant aspects that overall describe the item, thus we also designed an *aspect ranking* phase to merge the information we extracted from each single review. Specifically, for each aspect a_j we calculate its *global score* as follows:

$$score(a_j) = \frac{\sum_{i=1}^{N} n_{a_j, r_i} * rel(a_j, r_i) * sent(a_j, r_i)}{N} \quad (2)$$

Our formula gives a higher score to the aspects that are often mentioned in the reviews (n_{a_j, r_i} is the number of the occurrences of a_j in r_i) with a *positive* sentiment. At the end of this step all the aspects are ranked and the *top-K* are labeled as *main aspects*. The *Top-5 Aspects* returned by our algorithm for three different movies are reported in Table 1.

Sentence Filtering. Once the main aspects are identified, we run a *sentence filtering* phase. The goal of this phase is to filter out *non-compliant* sentences that are supposed to be not useful in the final justification we want to build. In this case we first split each review $r_i \in R$ in sentences $s_{i1} \dots s_{im}$. Next, we verify the compliancy of each sentence s_i and we maintain only the sentences matching the following criteria: *(i)* s_i contains a *main aspect* $a_1 \dots a_k$; *(ii)* s_i is longer than 5 tokens; *(iii)* s_i expresses a *positive* sentiment; *(iv)* s_i does not contain first-person personal or possessive pronouns.

Chinatown	The Ring	Titanic
cast	actor	story
ending	thriller	love
nicholson	effects	effects
performance	horror	picture
story	character	music

Table 1: Top-5 Aspects returned by the Aspect Ranking algorithm for three different movies.

The rationale behind these heuristics is straightforward: we want to include in the justification only the sentences including a *relevant aspect* and expressing a *positive sentiment* about the item. Moreover, we decided to filter out very short and non-informative sentences as well as those using the *first person*. In this case, the intuition is to prefer excerpts having a more *impersonal* style (e.g., *'The movie has a great cast'*) than those expressing personal opinions (e.g., *I liked the cast of the movie, this is my favourite director*).

Text Summarization. At the end of the Sentence Filtering a set of potential *candidate sentences* is obtained. Such a set is used to feed a *text summarization* algorithm whose goal is to generate a unique summary to be used as justification of the recommendation. Such a summary is supposed to highlight the main contents discussed in the reviews of the item and to maximize both the *coverage* and *diversity* of the justification. This is done by selecting sentences which cover enough amount of topics discussed in the original reviews and by avoiding the redundancy as well.

To run text summarization we adapted the method described in [17], which proved to be effective in a multi-document summarization task. This makes the approach very suitable for our scenario, since each *review* can be easily considered as a *document*, thus the method can be very effective in summarizing all the reviews excerpts in a single *summary* that highlights the most salient features of the item.

Our approach combines centroid-based text summarization [16], which has the advantage of being unsupervised, with a pre-trained neural language model, such as Word2Vec [10], which is good in transferring information from web-scale textual corpora, and is based on two steps: first, all the information coming from the reviews of an item are condensed in a *centroid vector* which represents a pseudo-review. Next, the main idea is to project both the centroid and each sentence of the reviews in a vector space and to include in the summary only the sentences closer to the centroid.

Formally, given a set of reviews R and its vocabulary V with size $N = |V|$, we first define a matrix $M \in \mathbb{R}^{N,k}$, so-called *lookup table*, where the i-th row is a word embedding of size k, $k << N$, of the i-th word in V. The values of the word embeddings matrix M are learned by using Word2Vec. When the lookup table is learned, the summarization method consists of three phases:

Centroid Vector Building. The centroid vector that represents the *meaning* of the reviews is built in two steps. First, the most meaningful words occurring in the reviews (that is to say, those having their $tf * idf$ weight greater than a topic threshold) are

[1] http://www.natcorp.ox.ac.uk/

Movie	Justification
Chinatown	*This movie has a decent plot and great acting by Jack Nicholson. Unlike most noir films Chinatown has a great script with fantastic characters, a whirling, looping (but always believable) plot, and one of the best endings of any Hollywood film of any period.*
The Ring	*If you like or love the blood and gore kinds of films, this movie will certainly disappoint you as the focus is on character, story, mood and unique special effects. The Ring is a story about supernatural evil therefore, it is a horror film, done very much in the style of the psychological thriller.*
Titanic	*The highest grossing movie of all time, James Cameron's Titanic follows the love story of Jack and Rose set on the doomed 1912 maiden voyage of the Titanic. This is the greatest love story of our time and a very good drama with great special effects.*

Table 2: Example of justifications generated by our automatic text summarization technique.

selected. Next, the embedding of the centroid is computed as the sum of the embeddings of the top ranked words in the reviews using the lookup table M.

$$C = \sum_{w \in R, tfidf(w) > t} M[idx(w)] \tag{3}$$

In the eq. (3) we denote with C the centroid embedding related to the set of reviews R and with $idx(w)$ a function that returns the index of the word w in the vocabulary.

Sentence Scoring. Given the centroid vector, we need to identify the sentences to be included in the summary. For each sentence in the set of the reviews, we create an embedding representation by summing the vectors for each word in the sentence stored in the lookup table M.

$$S_j = \sum_{w \in S_j} M[idx(w)] \tag{4}$$

In the eq. (4) we denote with S_j the j-th sentence in the set of reviews R. Then, the sentence score is computed as the cosine similarity between the embedding of the sentence S_j and that of the centroid C of the set R.

$$sim(C, S_j) = \frac{C^T \bullet S_j}{||C|| \cdot ||S_j||} \tag{5}$$

Sentence Selection. The sentences are sorted in descending order of their similarity scores. The top ranked sentences are iteratively selected and added to the summary until the limit, in terms of the number of words in the summary, is reached. In order to minimize the redundancy of the information included in the summary, during the iteration we compute the cosine similarity between the next sentence and each one already in the summary. We discard the incoming sentence if the similarity value is greater than a threshold.

At the end of this process a *summary* is produced. An example of the summaries generated for three movies is reported in Table 2. These summaries were used in our experiments as justifications.

3 EXPERIMENTAL EVALUATION

The goal of the experimental session was twofold: to evaluate the effectiveness of different configurations of our justifications based on text summarization *(Experiment 1)*, and to compare our methodology to a baseline that exploits users' reviews without text summarization *(Experiment 2)*.

To this end, we designed a user study in the *movie* domain involving 141 subjects. Our sample was obtained through the *availability sampling* strategy, and it includes students, researchers in the area and people not skilled with computer science and recommender systems.

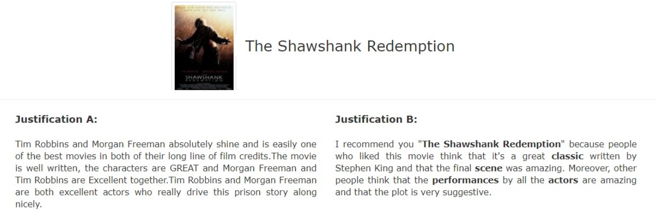
Experimental Protocol. To run the experiment, we deployed a web application[2]. As regards *Experiment 1*, we run a the experiment in a *between-subject* fashion, that is to say, each user was randomly assigned to a configuration of our pipeline, and he evaluated the justifications. Clearly, the user was not aware of the specific configuration he was interacting with. Conversely, for *Experiment 2*, we run a *within-subject* experiment, that is to say, all the users were provided with two different explanation styles (*i.e.*, our justifications exploiting text summarization and a review-based baseline presented in [13]). It is worth to note that we did not evaluate any other baseline since in our previous experiments [13] we already showed that our review-based baseline performs better than other explanation styles.

To run the experiment we built a dataset by mapping MovieLens data with Amazon reviews[3] about the movies. The resulting dataset contained 307 movies and 153,566 reviews. The average length of each review was 138.38 words. The mapping between MovieLens movies and Amazon products is available online[4].

Parameters of the Experiment. In order to evaluate different configurations of our pipeline, we compared four different different alternatives of the algorithm, obtained by varying *number of aspects* and *justification length*. As for the length, we compared *long* and *short* justifications, depending on the overall length of the justification (50 words for short justifications, 100 words for long justifications). These values were set in an empirical fashion. As for the number of aspects, we compared the justifications building by including the *top-10* aspects and the *top-30 aspects*.

Implementation details. Recommendations were generated by using Personalized PageRank (PR) [5] as recommendation algorithm, as in our previous work [14, 15]. To implement the recommendation algorithms we exploited Jung framework[5] and PR was run by using the default settings. As for the Aspect Extraction and the Aspect Ranking, we exploited the algorithms available in CoreNLP[6]. To identify the sentiment conveyed by the single sentences, we used the Stanford Sentiment Analysis algorithm[7], and we took into account sentences whose sentiment was labeled as *positive* or *very positive*. As regards the Text Summarizer, pretrained word embeddings learned through Word2Vec were used. The source code of the summarizer is available online as well[8].

[2] https://tinyurl.com/uniba-expl-summarizer
[3] http://jmcauley.ucsd.edu/data/amazon/
[4] https://tinyurl.com/uniba-reviews-mapping-dataset
[5] http://jung.sourceforge.net/
[6] https://stanfordnlp.github.io/CoreNLP/
[7] https://nlp.stanford.edu/sentiment/
[8] https://github.com/gaetangate/text-summarizer



The Shawshank Redemption

Justification A:

Tim Robbins and Morgan Freeman absolutely shine and is easily one of the best movies in both of their long line of film credits.The movie is well written, the characters are GREAT and Morgan Freeman and Tim Robbins are Excellent together.Tim Robbins and Morgan Freeman are both excellent actors who really drive this prison story along nicely.

Justification B:

I recommend you "**The Shawshank Redemption**" because people who liked this movie think that it's a great **classic** written by Stephen King and that the final **scene** was amazing. Moreover, other people think that the **performances** by all the **actors** are amazing and that the plot is very suggestive.

Figure 1: Screenshot of the platform during the within-subject experiment

Configuration		Metrics				
Aspects	*Length*	TRA	PER	ENG	TRU	EFF
Top-10	Short	2.83	3.06	3.06	2.83	0.89
Top-30	Short	3.16	3.06	2.69	3.19	0.94
Top-10	Long	**3.95**	**3.64**	**3.37**	**3.55**	0.55
Top-30	Long	3.24	3.18	3.12	3.22	**0.38**

Table 3: Results of Experiment 1. The best-performing configuration for each metric is reported in bold.

In a nutshell, each user involved in the experiment carried out the following steps: **(1) Preference Elicitation and Generation of the Justifications**. We asked users to explicitly rate at least three items, chosen among a randomly generated subset of 20 movies. Next, recommendations were generated. **(2) Between-subject Evaluation through Questionnaires**. We asked the users to fill in a questionnaire to evaluate the quality of the justification. Each user was asked to evaluate the previously presented *metrics* through a five-point scale (1=strongly disagree, 5=strongly agree). **(3) Within-subject Evaluation through Questionnaires**. We asked each user to evaluate the explanation style they preferred between our methodology and the baseline. As shown in Figure 1, we provided the user with both the explanations in the same screen, and we asked them to select the best one.

Evaluation Metrics. As in [19], we evaluated *transparency*, *persuasiveness*, *engagement* and *trust* of the recommendation as the average score collected through the user questionnaires, while the *effectiveness* was calculated as the normalized difference between the pre- and post-trailer ratings and pre- and post-excerpt ratings the user provided for the recommendation.

4 RESULTS AND CONCLUSIONS

The results we obtained for Experiment 1 are reported in Table 3. The values represent the average scores given by the users for that specific metric. The higher the better, with the exception of the effectiveness where the best configuration is the one closer to zero.

The first result emerging from the experiment is that *longer justifications* tend to beat their shorter counterparts. This means that very short justifications cannot convey the information which is needed by the final users to better understand the quality of the suggestions they received. Such a low quality is also confirmed by analyzing the scores we obtained for *short justifications*: they are often below 3.00 (equivalent to *'partially agree'* as answer), which is considered as the minimum score that makes a justification as *acceptable*. Conversely, when longer justifications are produced by our technique, the results we obtained are generally higher. This means that the addition of more sentences can make the justifications and the recommendation process more satisfying, engaging, transparent and trustful for the target users.

Movies	Aspects+Summar.	Aspects	Indifferent
Transparency	**54.55%**	40.91%	4.55%
Persuasion	**77.27%**	13.64%	9.09%
Engagement	**63.63%**	27.27%	9.09%
Trust	**68.18%**	4.55%	27.27%

Table 4: Results of Experiment 2. The configuration preferred by the higher percentage of users is reported in bold.

As regards the number of *aspects* to be included, the experiment showed that the best results are obtained by exploiting the *top-10 aspects* identified by our aspect ranking module. This is confirmed for all the metrics except of the effectiveness. This means that by selecting a subset of relevant aspects we can provide the user with a summary containing the most relevant and useful information to evaluate the quality of the suggestion. Conversely, when a higher number of aspects is exploited, some noise is probably introduced in the summarization process, thus non-relevant or non-interesting sentences are put in the final justification, and this leads to a decrease in the overall results.

Next, in Experiment 2 we compared our justification based on automatic text summarization to the to the justifications generated by exploiting a review-based baseline in a *within-subject* experiment. Due to space reasons, we only report the results of the comparison between the best-performing configurations emerging from *Experiment 1* and the baseline. Specifically, we used *long justifications based on top-10 aspects.* In a nutshell, our baseline identifies the most relevant aspects by using our Aspect Extraction and Aspect Ranking strategies and randomly selects a sentence expressing a *positive* sentiment that contains that aspects. For more details on the baseline we suggest to refer to our previous work [15].

As shown in Table 4, most of the users indicated that they preferred our methodology based on *automatic text summarization.* It is worth to note that we obtained the higher results for *persuasion* and *engagement.* This is a very encouraging outcome that confirmed the intuition behind this work, since the exploitation of text summarization can help to make interesting information about the recommended item emerge, and this can persuade the user in enjoying the item or can let the user discover new information about the item itself.

Overall, we can state the the results emerging from these experiments confirmed the effectiveness of the approach, and showed that automatic text summarization can be useful to identify the most relevant aspects of the items and support the suggestions generated by a recommendation algorithm. As future work we will introduce some strategy to automatically tune the parameters of the model, in order to dynamically select the optimal number of aspects and sentences to be included. Moreover, we will also evaluate more sophisticated techniques for natural language processing in order to include also entities and bigrams in our justifications.

Acknowledgement. This work is partially funded by the project "Electronic Shopping and Home delivery of Edible goods with Low environmental Footprint" (ESHELF), under the Apulian INNONETWORK programme, Italy.

REFERENCES

[1] Jorge A Balazs and Juan D Velásquez. 2016. Opinion mining and information fusion: a survey. *Information Fusion* 27 (2016), 95–110.

[2] Or Biran and Courtenay Cotton. 2017. Explanation and justification in machine learning: A survey. In *IJCAI-17 Workshop on Explainable AI (XAI)*. 8.

[3] Shuo Chang, F Maxwell Harper, and Loren Gilbert Terveen. 2016. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 175–182.

[4] Guanliang Chen and Li Chen. 2015. Augmenting service recommender systems by incorporating contextual opinions from user reviews. *User Modeling and User-Adapted Interaction* 25, 3 (2015), 295–329.

[5] Taher H. Haveliwala. 2003. Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search. *IEEE Trans. Knowl. Data Eng.* 15, 4 (2003), 784–796.

[6] Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. 2015. Trirank: Review-aware explainable recommendation by modeling aspects. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1661–1670.

[7] Zhong-kai HU, Xiao-lin ZHENG, Ya-feng WU, and De-ren CHEN. 2013. Product recommendation algorithm based on users reviews mining. *Journal of Zhejiang University (Engineering Science)* 8 (2013), 023.

[8] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.

[9] Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining text data*. Springer, 415–463.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *NIPS*. 3111–3119.

[11] Khalil Ibrahim Muhammad, Aonghus Lawlor, and Barry Smyth. 2016. A Live-User Study of Opinionated Explanations for Recommender Systems. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 256–260.

[12] Cataldo Musto, Marco de Gemmis, Giovanni Semeraro, and Pasquale Lops. 2017. A Multi-criteria Recommender System Exploiting Aspect-based Sentiment Analysis of Users' Reviews. In *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017*, Paolo Cremonesi, Francesco Ricci, Shlomo Berkovsky, and Alexander Tuzhilin (Eds.). ACM, 321–325. https://doi.org/10.1145/3109859.3109905

[13] Cataldo Musto, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2019. Justifying Recommendations through Aspect-based Sentiment Analysis of Users Reviews. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2019, Larnaca, Cyprus, June 9-12, 2019.*, George Angelos Papadopoulos, George Samaras, Stephan Weibelzahl, Dietmar Jannach, and Olga C. Santos (Eds.). ACM, 4–12. https://doi.org/10.1145/3320435.3320457

[14] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2016. ExpLOD: A Framework for Explaining Recommendations Based on the Linked Open Data Cloud. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 151–154. https://doi.org/10.1145/2959100.2959173

[15] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2019. Linked open data-based explanations for transparent recommender systems. *International Journal of Human-Computer Studies* 121 (2019), 93–107.

[16] Dragomir R. Radev, Hongyan Jing, Magorzata Sty, and Daniel Tam. 2004. Centroid-based summarization of multiple documents. *Inf. Process. Manage.* 40, 6 (2004), 919–938.

[17] Gaetano Rossiello, Pierpaolo Basile, and Giovanni Semeraro. 2017. Centroid-based Text Summarization through Compositionality of Word Embeddings. In *Proceedings of the Workshop on Summarization and Summary Evaluation Across Source Types and Genres, MultiLing@EACL 2017, Valencia, Spain, April 3, 2017*, George Giannakopoulos, Elena Lloret, John M. Conroy, Josef Steinberger, Marina Litvak, Peter A. Rankel, and Benoît Favre (Eds.). Association for Computational Linguistics, 12–21. https://aclanthology.info/papers/W17-1003/w17-1003

[18] Michael Schrage. 2016. How the big data explosion has changed decision making. *Harvard Business Review* (2016).

[19] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *UMUAI* 22, 4-5 (2012), 399–439.