information

MDPI

Article

Towards Context-Aware Opinion Summarization for Monitoring Social Impact of News

Alejandro Ramón-Hernández [1], Alfredo Simón-Cuevas [2,*], María Matilde García Lorenzo [1], Leticia Arco [3] and Jesús Serrano-Guerrero [4]

1 Centro de Investigaciones de la Informática, Universidad Central "Marta Abreu" de Las Villas, Villa Clara 54830, Cuba; aramon@uclv.cu (A.R.-H.); mmgarcia@uclv.edu.cu (M.M.G.L.)
2 Facultad de Ingeniería Informática, Universidad Tecnológica de La Habana "José Antonio Echeverría", La Habana 11500, Cuba
3 Computer Science Department, Vrije Universiteit Brussel, 1050 Brussels, Belgium; larcogar@vub.be
4 Department of Technologies and Information Systems, University College of Computer Science, University of Castilla-La Mancha, 13071 Ciudad Real, Spain; Jesus.Serrano@uclm.es
* Correspondence: asimon@ceis.cujae.edu.cu; Tel.: +53-7-2607242

Received: 10 October 2020; Accepted: 13 November 2020; Published: 18 November 2020

Abstract: Opinion mining and summarization of the increasing user-generated content on different digital platforms (e.g., news platforms) are playing significant roles in the success of government programs and initiatives in digital governance, from extracting and analyzing citizen's sentiments for decision-making. Opinion mining provides the sentiment from contents, whereas summarization aims to condense the most relevant information. However, most of the reported opinion summarization methods are conceived to obtain generic summaries, and the context that originates the opinions (e.g., the news) has not usually been considered. In this paper, we present a context-aware opinion summarization model for monitoring the generated opinions from news. In this approach, the topic modeling and the news content are combined to determine the "importance" of opinionated sentences. The effectiveness of different developed settings of our model was evaluated through several experiments carried out over Spanish news and opinions collected from a real news platform. The obtained results show that our model can generate opinion summaries focused on essential aspects of the news, as well as cover the main topics in the opinionated texts well. The integration of term clustering, word embeddings, and the similarity-based sentence-to-news scoring turned out the more promising and effective setting of our model.

Keywords: opinion mining; opinion summarization; topic modeling; semantic similarity measures; word embeddings

1. Introduction

The globalization of the use of the Internet and the development of technologies such as Cloud Computing, Internet of Things, social networks, Mobile Computing, and others has favored the increase of user-generated content on the web. Nowadays, a surprisingly high quantity of news, messages, and reviews of products or services are generated in online social media, news portals, e-commerce sites, etc. The data and information produced by users have proven useful in many domains (e.g., marketing studies, business intelligence, health, governance, and others) [1]. The processing of user-generated content on digital platforms (e.g., news platforms) is playing significant roles in the success of government programs and initiatives in digital governance, from extracting and analyzing citizens' sentiments for decision-making [2]. Several efforts have been dedicated to deal with extracting knowledge and efficient processing of this unstructured information produced by users [3], resulted

in increasing research interest in tasks within Natural Language Processing (NLP) such as sentiment analysis, also called opinion mining [4].

Opinion mining is the field of study that analyzes people's opinions, sentiments, appraisals, attitudes, and emotions towards entities and their attributes expressed in written texts [3]. Opinion mining (or sentiment analysis) is a broad area that includes many tasks, such as sentiment classification, aspect-based sentiment analysis, lexicon construction, opinion summarization, and others [5]. Opinion summarization is the task of automatically generating summaries for a set of opinions that are related to the same topic or specific target [6]. The aspect-based opinion summarization is one of the main approaches [7], but it would not be very appropriate in contexts where the opinions are not about products or services (e.g., opinions about news). Although summaries generated by several of the reported approaches are focused on specific topics [1,8,9], they are generally identified by looking only at the content in opinionated texts, whereas the context that originates the opinions (e.g., news) is not usually taken into account, being this a weakness. A comprehensive summary of the users' reactions concerning a news article can be crucial due to various reasons, such as (1) understanding the sensitivity/importance of the news, (2) obtaining insights about the diverse opinions of the readers regarding the news, and (3) understanding the key aspects that draw the interest of the readers [10]. On the other hand, to integrate both topic-opinion analysis and semantic information can yield satisfactory results in opinion summarization [1]. Nevertheless, the use of WordNet [11], as well as the deep-learning-based word embedding [12,13] (e.g., word2vec [14]) to represent and analyze the semantic of words when dealing with opinion summarization problems has been limited. Our work is addressed to the application of these models and resources to cope with opinion summarization challenges.

In this paper, a news-focused opinion summarization model is presented, which is conceived according to the conception of extractive and topic-based text summarization methods. Our model combines topic modeling, sentiment analysis, and the news-focused relevance scoring in seven phases: preprocessing, topic detection, sentiment scoring, topic-sentence mapping, topic contextualization, sentence ranking, and summary construction. The integration of these techniques allows us to deal with the problem in which the relevance focus not only comes from the texts of the opinions, but also comes from the news articles as the context that originates them. Semantic analysis is included in several phases, to improve text processing. The semantic characteristics of words are captured through the word2vec representation model [14] and from WordNet [11]. Besides, semantic similarity measures are used to assess the semantic relatedness between sentences-to-sentences and sentences-to-news.

The model was evaluated across two datasets containing Spanish news and opinions collected from a real digital news platform. The selected news and opinions are related to telecommunication services and the COVID-19 pandemic. The performance of our proposal was measured, using the Silhouette [15] and the Jensen–Shannon divergence (JSD) measures [16]. The first one is used to measure the quality of the clustering process, and then to estimate the prospective quality of the topic detection phase. The second one is used to measure the quality of the obtained summaries. Several experiments were carried out, to provide a deeper grounding for the contribution of our approach. Different settings of the proposed model were evaluated and compared, to analyze the behavior of the different techniques integrated into the model and to identify the best solution for the news-focused opinion summarization process. The analysis of the experimental results and obtained conclusions were substantiated through the well-known Wilcoxon's Statistics Test.

The rest of the paper is organized as follows: Section 2 summarizes the analysis of related works; Section 3 describes the proposed opinion summarization model; and Section 4 presents the datasets, metric description, and the experimental results and discussion. Conclusions and future work are pointed out in Section 5.

2. Related Works

Automatic text summarization is the task of producing a concise and fluent summary, condensing the most relevant and essential information contained in one or several textual documents, while preserving key information content and overall meaning of the information source [17]. Summarizing texts is still an active research field and needs further developments due to the huge data increase on the web [18] (e.g., user-generated content). These methods and techniques have been addressed for processing user-generated opinionated content on social networks and digital platforms, emerging as a new challenge [6]. Summaries can be automatically obtained through extractive (i.e., selecting the most important sentences from documents) or abstractive methods (i.e., generating new cohesive text that may not be present in the original information) [6,19]. Most of the opinion summarization models follow extractive methods [7,20]. Unlike traditional text summarization, the opinion-oriented summaries have to take into consideration the sentiment a person has towards a topic, product, place, or service [1]. Since a text summarization aims to generate a concise version of factual information, a sentiment summarization summarizes sentiments from a large number of reviewers or multiple reviews [21]. The opinion mining provides the sentiment associated with a document at different levels through the polarity detection task, whereas text summarization techniques identify the most relevant parts of one or more documents and build a coherent fragment of text (the summary) from them [1].

One of the main approaches to generate opinion summaries is the aspect-based opinion summarization [7,22], which summarizes opinions depending on different aspects or features (attributes or components) of an entity (objects, organizations, services, and products). In the context in which the aspects or features do not stand out, topic detection turning out critical for dismissing non-relevant sentences. However, achieving high effectiveness in this process constitutes a challenging task in contexts of the great diversity of opinions. Identifying topics is of great importance to determine regarding which issues users are giving their criteria [23], being one of the reasons that some opinion summarization approaches detect topics in their textual analysis [1,8,9,24,25]. Although the resulting summaries are generally focused on aspects or topics, they are mainly identified taking into account only the content of the opinionated texts and do not focus on specific information-context interests. Nevertheless, there are approaches where the relevance focus not only comes from the texts of the opinions, such as query-based opinion summarization, which aims to extract and summarize the opinionated sentences related to the user's query [6,26,27]. In these systems, classical summarization techniques are applied, and the context (query) is used as a relevant focus, to generate a coherent and useful summary for the user [28]. Other challenges are implicit in these opinion summarization methods, such as the following: how to retrieve query relevant sentences, how to cover the main topics in the opinionated text set, and how to balance these two requests [29]. Our proposal is addressed to a similar problem, where news articles are used as the relevant focus instead of users' queries, although few approaches dealing with this problem have been identified [10]. For instance, Chakraborty et al. reported a method of summarizing news article tweets that initially captures the diverse opinions from the tweets by creating a unique tweet similarity graph, followed by a community detection technique to identify the tweets representing these diverse opinions [10]. Representative keywords of the news articles are extracted to identify related tweets. The similarity scoring between news-tweets and a pair of tweets is based on the overlapping keywords (content similarity), and the word vectors' similarity (context similarity), respectively.

According to the results reported in Reference [1], integrating both topic-opinion analysis and semantic information can yield satisfactory results in opinion summarization. In this sense, for the analysis of opinions which are generally short texts, it is more useful to represent terms and to capture semantic information about them. Two fundamental approaches collect semantic characteristics of terms. One of them depends on the context, and the other one depends on the meaning. Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) are more commonly used methods for topic modeling in opinions and to capture the semantic information from the context, as reported

in [1,8,24,25]. However, some researchers consider LDA- and LSA-based approaches to not proprerly model the aspects of the reviews made on the web [3]; instead, clustering text segment approaches have the advantage of keeping the document structure through segments, to capture the semantics of texts [30]. On the other hand, word embedding models [12] (e.g., word2vec [14], Glove [31], and FastText) have been less applied; only a few approaches have been identified [8,10]. A word embedding is a learned representation for text where words that have the same meaning have a similar representation. This kind of representation has been successful in extractive summarization [32]. WordNet [11] is the most commonly used technique for capturing and processing the semantic meaning of terms; however, it has not been so much when summarizing opinions. In this context, the use of WordNet is mainly limited to capture synonyms, and few approaches have been identified [26,33,34]. Nevertheless, the use of WordNet in our proposal goes further on.

3. News-Focused Opinion Summarization Model

The conception of the proposed model is based on the extractive and topic-based text summarization approach, where the relevance scoring of sentences not only requires processing the information content to be summarized (e.g., the set of opinions), but also requires to carry out an alignment process with external or contextual information of interest—in our case, news content. An overview of the proposed model is shown in Figure 1. The proposed model combines the topic modeling (phase 2) and the news content, to determine the "importance" of opinionated sentences; it also includes the sentiment analysis process (phase 3) to determine the polarity strength of sentences and avoid the inclusion of non-opinionated sentences in the automatic summary. The topic-sentence mapping (phase 4) and topic contextualization (phase 5) allow us to align the sentences to the corresponding identified opinion topics and to determine the most relevant topics concerning the news. The least relevant topics are discarded, following the sentence ranking (phase 6) and summary construction (phase 7) processes.

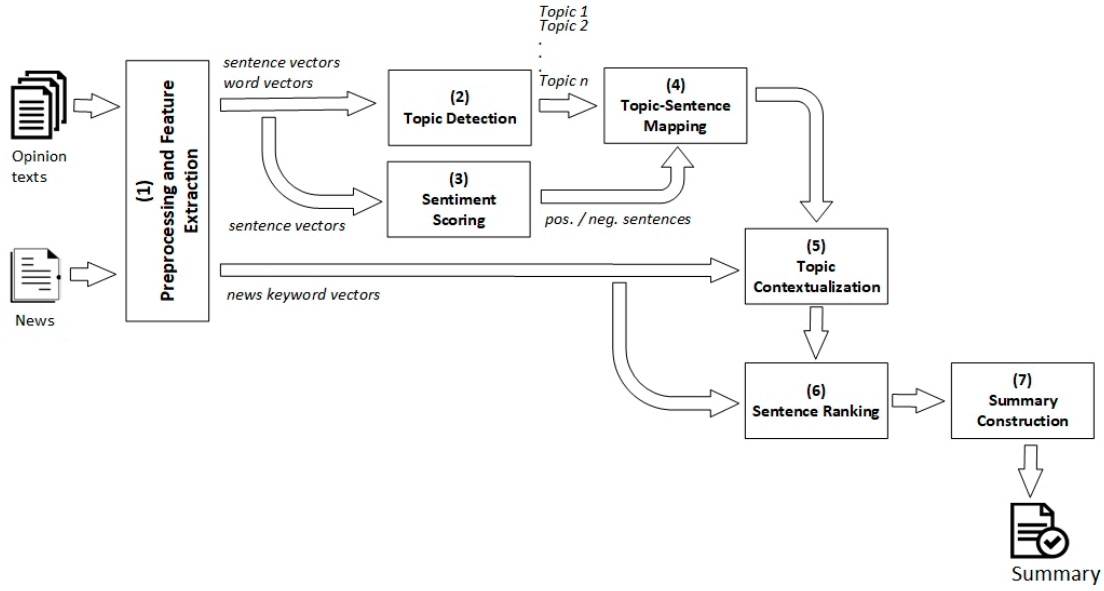


Figure 1. Workflow overview of the proposed model.

Several model settings and techniques were developed and evaluated, which are centered to address three important problems in the proposed model, such as (1) granularity in the topic modeling, (2) semantic processing of words and sentences, and (3) sentence relevance scoring. All of these developed alternatives are explained in the following subsections.

3.1. Preprocessing and Feature Extraction

In this phase, several Natural Language Processing tasks are performed for structuring the text (news and opinions) and extracting features, according to the preprocessing steps commonly reported in the opinion mining solutions [4]. Initially, the texts are split into sentences, and the tokenization task is applied to each sentence, for obtaining words or phrases. Some stop words, such as "la", "de", "y" and "o" (experiments were developed using Spanish text), are removed, considering that these words provide little useful information. Besides this, the lemmatization process of all words is carried out. Subsequently, the Part-of-Speech (POS) tagging is performed to determine the POS tag corresponding to each word belonging to sentences that make up opinions and news. The spaCy library of Python was used to support these tasks.

A crucial phase in opinion summarization is the feature-extraction phase, which simplifies the complexity of the involved tasks (e.g., topic modeling, sentiment classification, and semantic processing) by reducing the feature space. POS tags, such as adjective and noun, are quite helpful because the opinion words are usually adjectives and opinion targets (e.g., entities, aspects, or topics) are nouns or combinations of nouns [4]. Consequently, opinion features are constituted by noun phrases, adjectives, and adverbs. In the case of news texts, noun phrases play an important role as keywords in the content; therefore, they are used to construct the news keyword vector.

The vector space model was adopted for representing words and sentences (features). Two semantic representation approaches to reinforce the semantic processing were developed and evaluated, which are conceived through the use of (1) WordNet [11] and (2) word embeddings [12]. WordNet groups nouns, verbs, adjectives, and adverbs into sets of cognitive synonyms (*synsets*), each expressing a distinct concept meaning. Synsets are interlinked by means of conceptual–semantic and lexical relations. In the first case, the semantic characteristics of words are captured depending on their meaning. The feature vector is constructed with the *synset* of each word included in the sentence; in the case of ambiguous words (more than one *synset* in WordNet), the first *synset* that appears is selected. In the second case, the semantic characteristics of words are captured depending on their context. Word embedding vectors are obtained by applying the automatic learning model word2vec [14] on the sentences and news texts. Specifically, those vectors are generated by using the word2vec pre-trained model included in the es_core_news_md model of the spaCy library, which includes 300-dimensional vectors trained using FastText CBOW on Wikipedia and OSCAR (Common Crawl) containing 20 k unique words in Spanish.

3.2. Topic Detection

Topic detection is a way for monitoring and summarizing information generated from social sources, about which the participants discuss or argue or express their opinions. Therefore, identifying topics is of great importance to determine the relevant sentences of the opinion source to be included in the automatic summary. A topic can be analyzed and represented by considering different textual unit granularity, such as a group of terms, keywords, or sentences [30]. Term and sentence-based topic modeling approaches were applied and evaluated, adopting finally the first one in our proposal, as a consequence of the experimental results.

In our proposal, topic detection from all opinions is based on a clustering process, specifically of the terms extracted in the preprocessing task. In this sense, the cluster of terms represents the topics that have been boarded in the opinions. The objective of the clustering algorithms is to create groups that are coherent internally. In brief, cluster analysis groups data objects into clusters such that objects belonging to the same cluster are similar, while those belonging to different ones are dissimilar [35]. Both term and sentence clustering are carried out by applying a Hierarchical Agglomerative Clustering (HAC) algorithm [35]. HAC algorithm build hierarchies until obtaining a single cluster where all the objects are included. However, we need to obtain a certain quantity of groups of sentences that represent the topics boarded in the opinions. In this way, it is necessary to cut the hierarchy at some level for obtaining a partition. Although some variants to obtain a partition from a dendrogram are

reported in Reference [35], we adopted the definition of a threshold to achieve a standard cut-point for the hierarchies, which allows us to compare the results of the similarity measures of the clusters with this threshold in the cluster-construction process. Thus, terms are clustered until their higher similarities are less than the specified threshold; otherwise, the clustering process will be stopped. To obtain the threshold value, the mean of the maximum values of the similarities among any pair of objects was considered.

Two semantic processing approaches for measuring the similarity between text units in the clustering process were evaluated: (1) WordNet and (2) word embedding based, with the last one being the most promising. The Wu and Palmer measure included in WordNet::Similarity [36] is applied for computing the similarity of terms where the WordNet-based semantic processing is applied. The cosine similarity measure is applied over the word embeddings based term representation. The similarity between the sentences S_1 and S_2 is determined by using the following sentence-to-sentence similarity function [37] expressed in Equation (1):

$$sem\_sim(S_1,\ S_2) = \frac{1}{2}\left(\frac{\sum_{w\in\{S_1\}}(maxSim(w,\ S_2) * idf(w))}{\sum_{w\in\{S_1\}} idf(w)} + \frac{\sum_{w\in\{S_2\}}(maxSim(w,\ S_1) * idf(w))}{\sum_{w\in\{S_2\}} idf(w)}\right) \quad (1)$$

In this function, given two sentences, S_1 and S_2, for each word (w) in S_1, it is identified the word w' in the sentence S_2 that has the highest semantic similarity $maxSim(w_i, S_2)$, according to one of the word-to-word similarity measures (in our proposal, Wu and Palmer or cosine measures).

3.3. Sentiment Scoring

Different from traditional extractive text summarization, whose fundamental goal is extracting "important" sentences from single or multi-documents according to some features, the opinion-oriented summaries have to take into consideration the sentiment a person has towards a topic, product, place, service, etc. Opinion mining provides the sentiment associated with a document at different levels and through the polarity detection task, whereas text summarization techniques identify the most relevant parts of a document and build from them a coherent fragment of text (the summary) [1].

In this step, the sentiment analysis processing is performed based on a lexicon-based method, using the SpanishSentiWordNet (Spanish adjustment of SentiWordNet [38]) to extract sentiment-related words in texts. The SpanishSentiWordNet [39] lexicon is the result of the automatic annotation of all *synsets* of Spanish WordNet, according to the notions of "positivity" and "negativity". In this process, each WordNet *synset* is associated with two numerical scores, which indicate degrees of positivity and negativity of the contained terms (noun, verb, adjective, and adverb) in the *synset* [39]. The sentences that do not include sentiment content, or that have lower sentiment scores than a threshold value, are filtered. Words with a positive or negative SpanishSentiWordNet score greater than 0.4 are considered when computing the sentiment scores. The polarity scoring of a sentence is calculated as shown in Equations (2) and (3) [30]:

$$PosSentenceScore(j) = \sum_{t_i \in Opinion(j)} PosValue(t_i) \quad (2)$$

$$NegSentenceScore(j) = \sum_{t_i \in Opinion(j)} NegValue(t_i) \quad (3)$$

where $PosValue(t_i)$ and $NegValue(t_i)$ are the polarity values in SpanishSentiWordNet of the identified sentiment word t_i in the opinion j. The opinion polarity is determined according to the highest obtained polarity scores. According to Reference [30], the sum operator reached better accuracy achieved in the experimental results between four compared classical compensatory operators. The topic polarity

scores are measured by using the sum of the polarity scores $PosSentenceScore(S_j)$ and $NegSentenceScore(S_j)$ of each sentence S_j included in each cluster, according to Equations (4) and (5).

$$PosTopicScore(i) = \sum_{S_j \in Cluster(i)} PosSentenceScore\left(S_j\right) \tag{4}$$

$$NegTopicScore(i) = \sum_{S_j \in Cluster(i)} NegSentenceScore\left(S_j\right) \tag{5}$$

The highest obtained value of the cluster polarity score ($TopicScore(i)$) is used for determining which judgment (positives or negative) about the detected topics is the most representative in the processed opinion.

3.4. Topic-Sentence Mapping

Topic-based opinion-summarization systems, as our proposal, should be able not only to detect sentences that express a sentiment, but, more important, they should detect sentences that contain sentiment expressions towards the topic we are considering [1]. Once the opinion topics are identified and the sentences are classified as positive or negative, a mapping process between topics and sentences is performed. This process avoids the introduction of irrelevant sentences in the automatic summary. Mapping is carried out through computing the semantic similarity between the vocabulary that describes the topic and the sentences. For each sentence, Equation (1) is applied to compute sentences-to-topic similarity scores concerning all identified topics. Finally, the sentence is mapped onto the topic of the highest similarity score.

3.5. Topic Contextualization

Topic contextualization is one of the distinguishing tasks of our methodological proposal, concerning the generic opinion summarization systems that have been reported. In those systems, the generated summaries are generally focused on aspects or topics that are mainly identified while taking into account only the content of the opinionated texts. However, the purpose of our model is to provide automatic summaries focused on contexts of interest. In our model, these contexts are news articles, due the to fact they are the generators of the opinion comments.

In this phase, the news-based topic-ranking process is performed through computing the topic salience concerning the news content, obtaining a salience score for each topic. The topic salience is obtained by measuring the semantic similarity between the vocabulary associated with the topic and the news content. Topics with the lowest score (smaller or equal to a predefined threshold, which empirically was fixed in 0.5) are eliminated for the next steps of the summary construction process. This procedure means that the automatic summary will be built by extracting sentences from relevant topics of the news.

Similar to previous phases, Equation (1) and the conception for word-to-word semantic similarity are also applied. Topics are represented through term vectors, since the news is represented through the previously generated news feature vector. Formally, the salience score of a topic T_i for piece of news n_j is defined according to Equation (6). In the case of using sentence-based topic modeling (another developed and evaluated approach), topic salience is computed by averaging the semantic similarity between the sentence $S_k/S_k \in T_i$ and the news keyword vector, as shown in Equation (7).

$$salience\_score_1\left(T_i, n_j\right) = sem\_sim\left(T_i, n_j\right) \tag{6}$$

$$salience\_score_2\left(T_i, n_j\right) = \frac{\sum_{S_k \in T_i} sem\_sim\left(S_k, n_j\right)}{|T_i|} \tag{7}$$

3.6. Sentences Ranking

In this phase, the relevance assessment process applied to each opinionated sentence is carried out for generating the sentence ranking, according to a relevance score. Three approaches were developed and evaluated for measuring the relevance score:

1. Explanatoriness scoring [40]: In this approach, the ranking of sentences in opinions is based on their usefulness for helping users understand the reasons of sentiments (e.g., "explanatoriness"). It is one of the reported proposals in which the context is considered for determining the importance of the sentences. Kin et al. [40] proposed three heuristics for scoring explanatoriness of a sentence (i.e., length, popularity, and discriminativeness):

    - *Sentence length*: A longer sentence is very likely to be more explanatory than than a shorter one, since a longer sentence, in general, conveys more information.
    - *Popularity and representativeness*: A sentence is very likely to be more explanatory if it contains more terms that occur frequently in all sentences.
    - *Discriminativeness relative to background*: A sentence containing more discriminative terms that can distinguish opinionated sentences from background information is more likely explanatory.

In our proposal setting, for each sentence S_k, the clustered content by the contextualized topic to which the sentence S_k belongs is used as a reference for computing the *representativeness*. In addition, sentences from all opinions are used as background for computing the *discriminativeness*. It is important to point out that contextualized topics are the most important opinion topics for the news; therefore, this setting allows us to indirectly align the sentence relevance scoring process with the news context.

2 TextRank scoring [41]: TextRank is one of the most recognized standard and popular text summarization methods. This approach is conceived as a graph-based ranking model that is applied to an undirected graph extracted from natural language texts. In the graph, a sentence is represented as a vertex, and the "similarity" relation between two sentences determines the connexion (edge) between them. PageRank algorithm [42] is applied for computing the importance of a vertex (i.e., a sentence) within a graph.

3 Sentences-to-news scoring: This approach consists of computing the relevance score of each sentence S_k through measuring the semantic similarity between the sentence and the keyword vector of the news. For this purpose, Mihalcea et al. similarity function [37] (Equation (1)) is applied. Besides, two variants of the word-to-word semantic similarity are evaluated. Different from the explanatoriness scoring conception, this approach allows us to directly put the sentence-relevance scoring process in alignment with the news context, with the independence of the topic to the one belongs.

3.7. Summary Construction

Once the relevance of the sentences is computed in the previous phase, the summary-construction process is carried out by selecting the N opinionated sentences with a higher relevance score from each contextualized relevant topic. The N value depends on the predefined compression rate (summary size). However, we set $N = 3$ when evaluating our proposal.

4. Experimental Results

4.1. Description of Datasets

To evaluate the effectiveness of our proposed model, two datasets with real information in the Spanish language, regarding two different domains, namely telecommunications services (TelecomServ dataset) and COVID-19 pandemic (COVID-19), were created. These datasets were manually constructed

recovering information (news and opinions) from Cubadebate (www.cubadebate.cu), which is one of the most important and visited digital news platforms available in Cuba. For both datasets, the news-selection task was carried out while considering two fundamental requirements:

- The news should have an interest in national scope;
- The news should have more than 50 associated opinions or comments.

The TelecomServ dataset consists of 80 news and its associated opinions. Selected news are related to the Cuban Telecommunication Enterprise S.A. (ETECSA) and published in the last three years. The gathered information is one of the information sources that the enterprise may consider for measuring the customer's satisfaction regarding its services. On the other hand, the COVID-19 dataset consists of 85 news, along with their associated opinions, related to the battle against the one SARS-CoV2 coronavirus pandemic in Cuba. This dataset mostly gathers news related to information emitted by government authorities that were published in six months of the pandemic (March–August 2020). In this case, the gathered information and its processing/summarizing could be of great value for monitoring the social impact of the government actions for breaking the pandemic growth and the events that emerge in this difficult situation. The characterization of these datasets is shown in Table 1.

Table 1. Dataset characterization.

Datasets/Characteristics	#News	#Opinions	#Opinions/News	#Sentences	#Sentences/Opinion	#Terms	#Terms/Opinion
TelecomServ	80	15,776	197.2	34,665	2.2	917,674	58.2
COVID-19	85	21,707	255.4	55,447	2.5	1,587,813	73.1

4.2. Evaluation Metrics

Evaluation in text summarization can be extrinsic or intrinsic. In an extrinsic evaluation, summaries are assessed in the context of a specific task a human or machine has to carry out. In an intrinsic evaluation, summaries are evaluated about some ideal model. An intrinsic evaluation has been the most adopted paradigm, and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) measures [43] are the most widely used metrics for evaluating automatic summaries. However, these content-based evaluation metrics require us to compare the automatic summary with a human summary model; this is a problem when this human summary is not available.

The effectiveness of our proposal was evaluated in a real context where the human summary model is not available; therefore, the ROUGE measures would be discarded. To address this problem, we use Jensen–Shannon divergence [16] as the quality evaluation metric for assessing our automatic summary from different perspectives. The adoption of this metric is mainly motivated by two reasons: (1) good summaries to be characterized by a low divergence between probability distributions of words in the input and summary would be expected [44] and (2) several reported studies demonstrate the existence of a strong correlation among measures that use human models (e.g., ROUGE, Pyramids, and others) and the Jensen–Shannon metric [44,45]. These studies and their experiments were developed in the context of generic multi-document summarization, topic-based multi-document summarization [44], and opinion summarization tasks [45].

Jensen–Shannon divergence (*JSD*) is an Information-Theoretic measure of divergence between two probability distributions and is defined as shown in Equations (8)–(10) [45]:

$$JSD(P \parallel D) = \frac{1}{2}\sum_{w} P_w \log_2 \frac{2P_w}{P_w + Q_w} + Q_w \log_2 \frac{2Q_w}{P_w + Q_w} \tag{8}$$

$$P_w = \frac{C_w^T}{N} \tag{9}$$

$$Q_w = \begin{cases} \frac{C_w^S}{N_S} & if\ w \in S \\ \frac{C_w^T + \delta}{N + \delta * B} & otherwise \end{cases} \tag{10}$$

where P is the probability distribution of a word, w, in the text, T, and Q is the probability distribution of a word, w, in a summary, S; N, defined as $N = N_T + N_S$, is the number of words in the text (N_T) and the summary (N_S); B is equal to 1.5 $|V|$, where V is the vocabulary extracted from the text and the summary; C_w^T is the number of words, w, in the text; and C_w^S is the number of words, w, in the summary. For smoothing the summary's probabilities, we used $\delta = 0.005$. The *JSD* measure values are in the range [0, 1], where a lower value indicates a low divergence between the compared two probability distributions, resulting in a better quality of the automatic summary in our context. This measure can be applied to the distribution of units in system summaries P and reference summaries Q, and the value obtained would be used as a score for the system summary [45]. Nevertheless, in our evaluation framework, this measure was applied according to Reference [44], using the input (text news and opinions set) as a reference, through comparing the distribution of words in full input documents with the distribution of words in automatic summaries.

Topic detection constitutes another key piece in our summarization framework; therefore, its evaluation is also very important. The proposed topic-detection process was conceived through a clustering approach, applying a HAC algorithm, which suggests that, the higher quality the clustering process has, the higher quality the topic detection has. According to this supposition, we decide to apply the Silhouette measure [15]. Silhouette, a clustering validity measure, is conceived to select the optimal number of clusters with ratio scale data (as in the case of Euclidean distances) that are suitable for a separated cluster. It is important to point out that Silhouette values range from −1 to +1, where a high value indicates that the object is well matched to its cluster and poorly matched to neighboring clusters, therefore resulting in a better quality of the clustering process.

4.3. Experimental Setup

In this section, we describe the experimental setup that was considered for both datasets and used to evaluate the effectiveness of the proposed news focused opinion summarization model. In our experiments, several solutions based on our model were developed and evaluated, to identify the best alternatives. The characterization of the evaluated approaches is shown in Table 2. For each piece of processed news and automatically generated summary with each of these solutions, we computed the averaged Silhouette and *JSD* measures. The *JSD* measure was computed from two perspectives:

- To measure the divergence between the automatic summary and the news content (*JSD focused on the news*), intending to know the correspondence level of the generated summary concerning the news.
- To measure the divergence between the automatic summary and the content of all opinions (*JSD focused on opinions*), intending to know the correspondence level of the generated summary concerning all opinions. The generated summary not only should be relevant to the news, but it should also be a good synthesis of the opinion set.

Table 2. Characterization and identification of the evaluated solutions.

Topic Detection Approaches	Semantic Processing Based on WordNet			Semantic Processing Based on Word Embeddings		
	Relevance Scoring			Relevance Scoring		
	Explanatoriness Scoring	TextRank Scoring (Baseline)	Sentence-to -News Scoring	Explanatoriness Scoring	TextRank Scoring (Baseline)	Sentence-to -News Scoring
Term clustering	OS1-WN	OS3-WN	OS4-WN	OS1-we	OS3-we	OS4-we
Sentence Clustering	OS2-WN	OS5-WN	OS6-WN	OS2-we	OS5-we	OS6-we

The following experimental tasks were performed:

1. Evaluating two topic detection approaches by using both term and sentence based granularities in the clustering process and comparing them by applying both WordNet

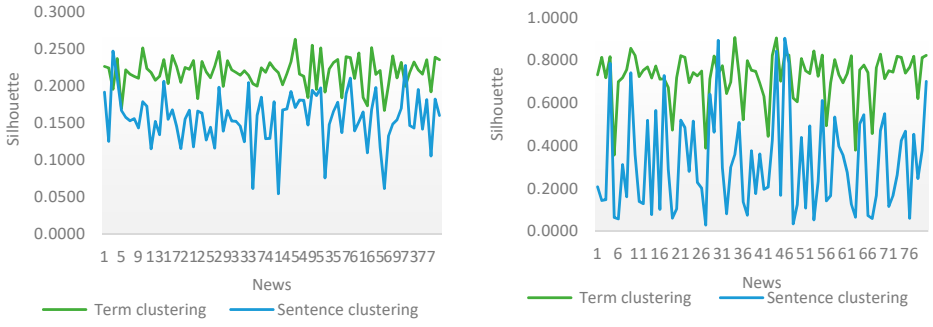
and word-embedding-based semantic-processing approaches. Selecting the clustering and semantic-processing approaches that provide the best results for topic detection.

2. Evaluating the automatically generated summaries from each solution in Table 2 according to *JSD focused on the news* (JSD_{News}) and *JSD focused on opinions* ($JSD_{Opinions}$), considering both WordNet and word-embeddings-based semantic-processing approaches. The obtained results would provide more details to the evaluation of the different configurations of the proposed model.
3. Comparing the results obtained by each solution in the previous tasks, identifying the best alternative for news-focused opinion summarization. TextRank-based [41] solutions are adopted as a baseline to evaluate the generated summaries according to the *JSD* measure. The best solution based on our model should work better than this popular and standard text summarization method.

Wilcoxon's Statistics Test was performed to validate the obtained results and to find significant differences between the evaluated solutions. From each dataset, 100% news and opinions were selected to constitute the sample group. In each test, the statistical significance was 95%, which means that the null hypothesis (H_0) will be rejected when the p-value ≤ 0.05.

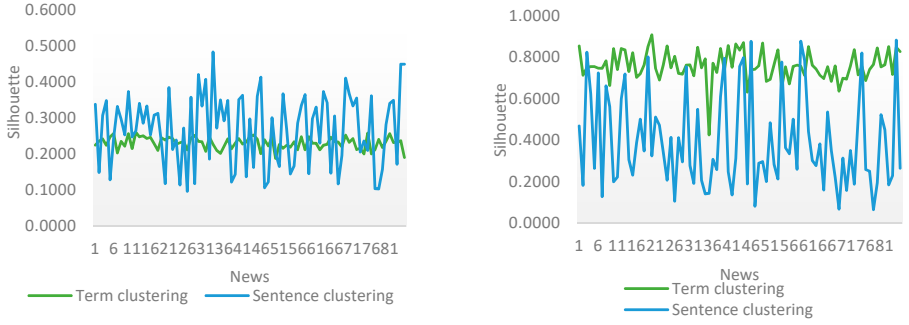
4.4. Results and Discussion

Figures 2 and 3 show detailed results of the first experimental task, where the evaluated solutions are grouped by the clustering approaches (term and sentence clustering), and the semantic processing (WordNet or word embeddings). This experimental task is focused on the Silhouette measure. Figures 4 and 5 show a comparative summary of the averaged Silhouette values for both datasets.



(**a**) (**b**)

Figure 2. Results of the Silhouette measure for the two clustering approaches in the topic detection on the TelecomServ dataset by applying (**a**) WordNet and (**b**) word embeddings based semantic processing approaches.



(**a**) (**b**)

Figure 3. Results of the Silhouette measure for the two clustering approaches in the topic detection on COVID-19 dataset by applying (**a**) WordNet and (**b**) word embeddings based semantic processing approaches.

As shown in Figures 2 and 3, Silhouette values are generally better when terms are clustered, regardless of the used semantic processing technique. Only in the case of the COVID-19 dataset, when WordNet is used (Figure 3a), do Silhouette values show better performance when sentences are clustered. It is important to point out that Silhouette values associated with each news show less dispersion when term clustering is applied, which is very positive behavior, because that means it is less sensitive to the diversity of news length and the number of associated opinions. Besides, term clustering represents a more stable clustering quality behavior. According to Figures 4 and 5, applying word embedding representation reaches best-averaged Silhouette values, those that are significantly higher when terms are clustered. These results allow us to conclude that term clustering, combined with word embeddings, is a more promising and effective setting of the topic modeling in our model. This combination guarantees good quality in the clustering-based topic detection, under the assumption that the quality of the detected topics is proportional to the clustering quality.

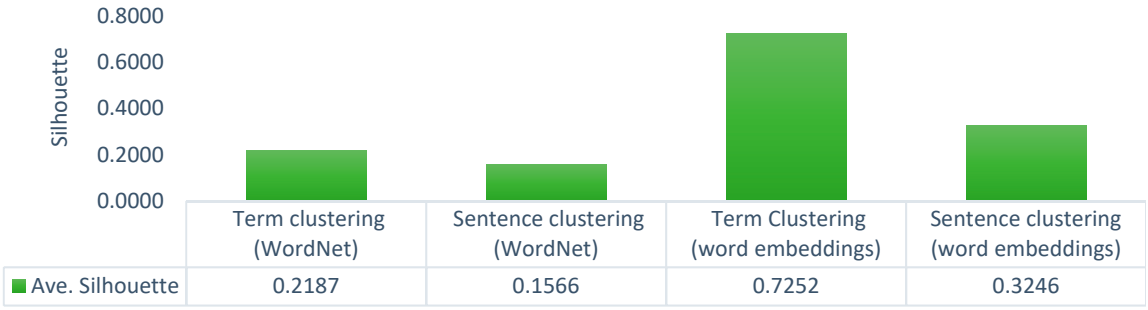


Figure 4. Averaged Silhouette values of compared topic detection approaches applied to the TelecomServ dataset.

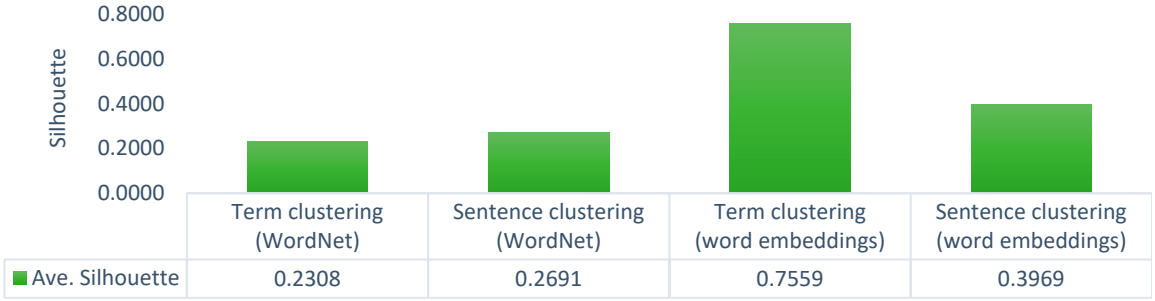
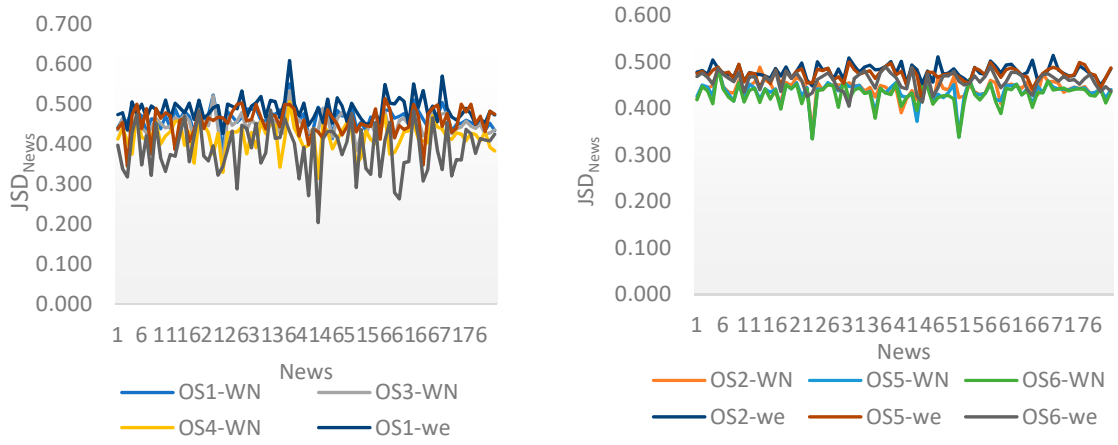


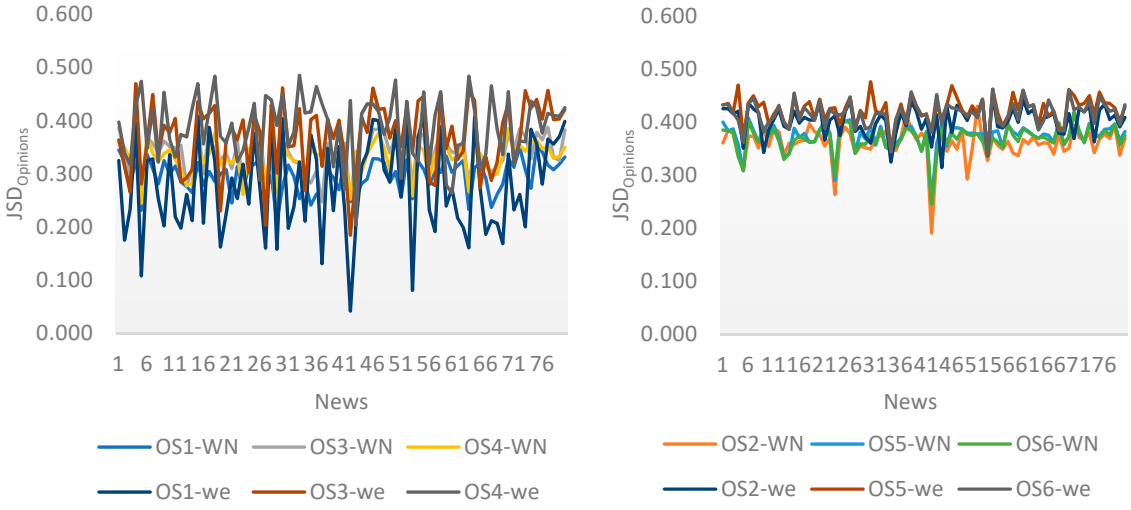
Figure 5. Average Silhouette values of compared topic detection approaches applied to the COVID-19 dataset.

Figures 6–9 show the detailed results associated with the second experimental task, which is based on the *JSD* measure. The evaluated and compared solutions are grouped according to the *JSD* scope focused on news or all opinions, as well as both term and sentence clustering. The semantic processing approach is specified in the identification of each solution (according to Table 2), which allows for an integral analysis of all developed model instances. As shown in Figures 6–9, OS4-WN and OS4-we are solutions that obtained the best results from JSD_{News} in both datasets, concerning the use of WordNet (OS4-WN) or word embeddings (OS4-we). These results indicate that combining topic modeling based on term clustering with the proposed *Sentence-to-news_scoring* for the sentence ranking is the setting of our model that allows us to generate automatic summaries more aligned to the main topics in the news, regardless of the semantic processing approach adopted.



(**a**) (**b**)

Figure 6. Results of JSD_{News} (Jensen–Shannon divergence focused on the news) applying (**a**) term and (**b**) sentence clustering, using WordNet and word embeddings on the TelecomServ dataset.



(**a**) (**b**)

Figure 7. Results of $JSD_{Opinions}$ (Jensen–Shannon divergence focused on the opinions) applying (**a**) term and (**b**) sentence clustering, using WordNet and word embeddings on the TelecomServ dataset.

On the other hand, OS1-WN and OS1-WN are solutions that reach the best results from $JSD_{Opinions}$ in both datasets, which means that *Explanatoriness_scoring* reaches better effectiveness to summarize the most important ideas of all opinions. These solutions do not ensure that the generated summaries have higher alignment with the news, concerning other solutions. Nevertheless, *JSD* focused on news obtained by these solutions, and their comparison with the rest of the solutions (see Tables 3 and 4) suggests that the inclusion of the topic-contextualization phase in the proposed model improves news-focused opinion summarization. Unlike the results shown in the first experiment, sentence clustering shows less sensitive behavior concerning the diversity of news length and the number of associated opinions.

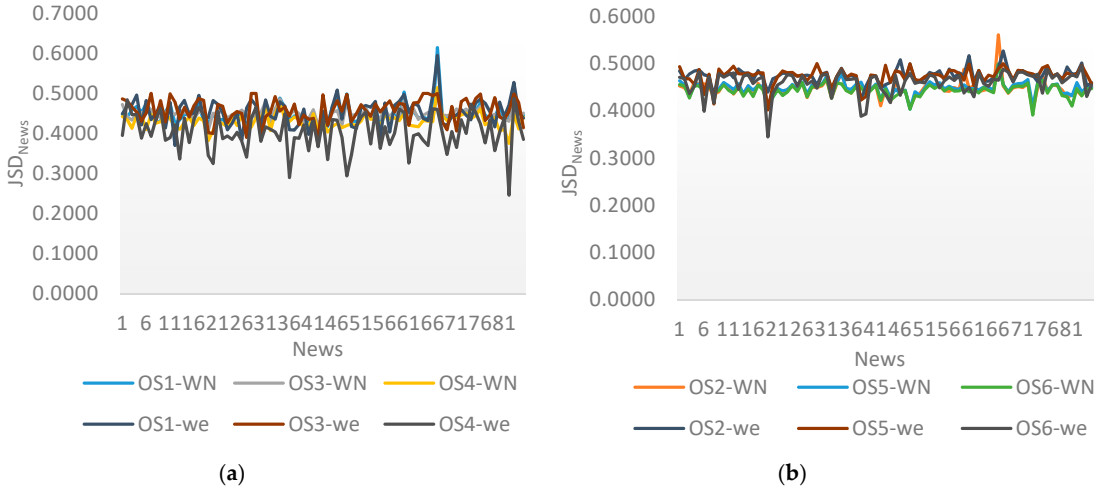


Figure 8. Results of JSD_{News} applying (**a**) term and (**b**) sentence clustering, using WordNet and word embeddings on the COVID-19 dataset.

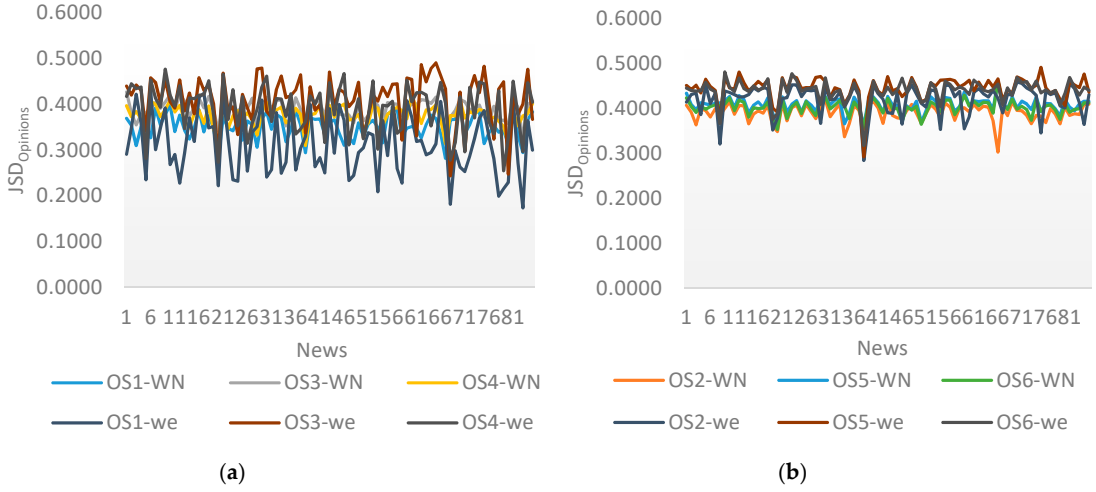


Figure 9. Results of $JSD_{Opinions}$ applying (**a**) term and (**b**) sentence clustering, using WordNet and word embeddings on the COVID-19 dataset.

Results shown in Tables 3 and 4, as well as in Figures 6–9, signify that the combination of term clustering and the word embedding representation model is also the more promising and effective setting of our model for reaching news-focused automatic summaries. Tables 3 and 4 show the averaged results of the JSD_{News} and $JSD_{Opinions}$ metrics, allowing them to complete the objective of the third task. Results of the WordNet-based semantic processing approaches are shown in Table 3, where OS3-WN was adopted as baseline 1. Results of the word-embedding-based semantic processing approaches are shown in Table 4, where OS3-we was adopted as baseline 2. These baselines were selected because the previous evaluation task concludes that the term clustering is the more promising and effective setting for topic modeling in our proposal. Thus, it allows us to evaluate the performance of the different approaches of our model and to compare them with notable summarizers as TextRank [41] (a similar decision is adopted in References [46,47]).

All solutions are compared according to the *JSD* scope for both datasets, and the best results are highlighted in bold. This comparison allows us to have a better understanding of the behavior of each approach. In general, the obtained results also showed that OS4-we is the best setting of our proposed model, according to JSD_{News} in both datasets. Furthermore, OS4-we is one of those solutions with best results from $JSD_{Opinions}$ when the word embedding representation is applied. This result allows us to conclude that the integration of term clustering, word embeddings, and the similarity-based

sentence-to-news scoring turned out to be the more promising and effective setting of our model. The automatic summaries obtained with OS4-we are more focused on the news content; they also cover the main topics in the opinion set, reaching an appropriate balance among these targets.

Table 3. Summary of averaged results of the JSD_{News} and $JSD_{Opinions}$ metrics considering WordNet-based semantic processing.

Compared Solutions	TelecomServ		COVID-19	
	$JSD_{Opinions}$	JSD_{News}	$JSD_{Opinions}$	JSD_{News}
OS1-WN	0.296	0.465	0.351	0.448
OS2-WN	0.361	0.443	0.390	0.447
OS4-WN	0.331	0.418	0.376	0.431
OS5-WN	0.374	0.435	0.408	0.449
OS6-WN	0.369	0.430	0.403	0.443
Baseline 1: OS3-WN TextRank [41]	0.335	0.449	0.390	0.449

Table 4. Summary of averaged results of the JSD_{News} and $JSD_{Opinions}$ metrics considering word-embedding-based semantic processing.

Compared Solutions	TelecomServ		COVID-19	
	$JSD_{Opinions}$	JSD_{News}	$JSD_{Opinions}$	JSD_{News}
OS1-we	0.278	0.487	0.314	0.453
OS2-we	0.403	0.479	0.420	0.474
OS4-we	0.388	0.388	0.392	0.404
OS5-we	0.424	0.473	0.445	0.474
OS6-we	0.416	0.459	0.439	0.460
Baseline 2: OS3-we TextRank [41]	0.370	0.457	0.411	0.458

The previous results were validated through statistical tests. Wilcoxon's test was applied to find significant differences between the OS4-we results and those obtained by the rest of the evaluated solutions, using JSD_{News} as quality metrics, as shown in Table 5. The statistical results show that there are significant differences between OS4-we and the compared solutions, since the obtained p-value is less than 0.05; thus, the null hypothesis in all compared cases is rejected. On the other hand, according to the #items-best values, OS4-we obtains best results for 87% of news (as average) in the TelecomServ dataset and the 85% of news (as average) in the COVID-19 dataset. Therefore, OS4-we is the best configuration of our proposed model for news-focused opinion summarization.

Table 5. Statistical results of Wilcoxon's test from OS4-we vs evaluated solutions.

Compared Solutions	TelecomServ (80 News)				COVID-19 (85 News)			
	Statistics Variables				Statistics Variables			
	Mean-Difference	z-Value	p-Value	#Items-Best	Mean-Difference	z-Value	p-Value	#Items-Best
OS1-WN	−0.07	−7.5094	<0.00001	76	−0.04	−6.6506	<0.00001	69
OS2-WN	−0.06	−6.052	<0.00001	67	−0.04	−6.6377	<0.00001	67
OS3-WN	−0.07	−6.4606	<0.00001	66	−0.04	−6.7884	<0.00001	72
OS4-WN	−0.05	−3.6639	<0.00043	52	−0.03	−4.8421	<0.00001	55
OS5-WN	−0.06	−5.1576	<0.00001	61	−0.05	−6.8702	<0.00001	72
OS6-WN	−0.05	−4.4902	<0.00001	61	−0.05	−6.375	<0.00001	67
OS1-we	−0.09	−7.9135	<0.00001	80	−0.06	−7.4688	<0.00001	80
OS2-we	−0.09	−7.809	<0.00001	79	−0.06	−7.9639	<0.00001	81
OS3-we	−0.06	−7.5053	<000001	76	−0.08	−7.9553	<0.00001	81
OS5-we	−0.09	−7.6592	<0.00001	76	−0.06	−7.9209	<0.00001	82
OS6-we	−0.08	−7.2742	<0.00001	73	−0.06	−7.3869	<0.00001	73
Average	−0.07	-	-	70	−0.05	-	-	73

4.5. Illustrative Examples

Examples 1 and 2 were selected to illustrate the summaries generated by applying OS4-we on opinions about two news articles related to COVID-19, which facilitates a better understanding of how our proposal works.

Example 1. Excerpt from the summary generated regarding opinions related to the news "VALIENTES: Cuatro heroínas en la batalla contra la COVID-19" by applying OS4-we.

Context	News title: VALIENTES: Cuatro heroínas en la batalla contra la COVID-19		
	URL: http://www.cubadebate.cu/noticias/2020/03/30/cuatro-heroinas-en-la-batalla-contra-la-covid-19-fotos/		
	News fragment: A Celeste, Claudia, Esther y Melisa solo se les puede ver a través de un cristal en el Instituto de Medicina Tropical "Pedro Kourí" (IPK Cuba) y después de someterse a un complejo protocolo de seguridad (. . .) Ellas comparten 24 horas seguidas con la COVID-19 y necesitan una alta concentración, pues el virus pasa por sus manos y no se pueden equivocar (. . .) Gracias a ese arriesgado trabajo, cada día se sabe si una persona en Cuba padece o no de una pandemia que amenaza a toda la humanidad. Lo mismo ocurre en otros dos laboratorios en Villa Clara y Santiago de Cuba.		
Terms topic	'agradecerles', 'salud', 'héroe'		
Opinions	Total: 171; Sentences: 347	Pos. Score	Neg. Score
Summary	Felicitaciones a todos los que están trabajando en la epidemia del coronavirus.	1.25	1.0
	Gracias, respeto, admiración, se merecen todo nuestros médicos, todo el personal de la salud y fuera de ella que esta dando todo para erredicar este virus.	7.1	1.9
	Combatientes por la humanidad¡.	1.9	1.4
$JSD_{Opinions}$	0.374		
JSD_{News}	0.382		

Example 2. Excerpt from the summary generated regarding opinions related to the news "Cuba frente a la COVID-19, día 100: Últimas noticias" by applying OS4-we.

Context	News title: Cuba frente a la COVID-19, día 100: Últimas noticias		
	URL: http://www.cubadebate.cu/noticias/2020/06/18/cuba-frente-a-la-covid-19-dia-100-ultimas-noticias/		
	News fragment: Cuba entra hoy, excepto La Habana y Matanzas, en la primera fase de la recuperación de la COVID-19. El presidente Miguel Díaz-Canel subrayó este miércoles la necesidad de intensificar en ambas provincias el trabajo para que, en el menor tiempo posible, también puedan pasar a la etapa pospandemia (. . .) Cuando se ha dispuesto el tránsito a la primera fase de la primera etapa pos-COVID-19, en 13 provincias de la Isla y el Municipio Especial Isla de la Juventud, Matanzas y La Habana figuran como las dos únicas dolorosas excepciones que por ahora no podrán retornar a la normalidad (. . .) Eliminar o mantener las restricciones (tránsito paulatino de una etapa a otras) responde a criterios sanitarios y no políticos, ha explicado Torres Iríbar (. . .) La tasa de incidencia acumulada es de 57,5 por 100 000 habitantes, con siete municipios por encima de la media provincial: Cotorro, Centro Habana, Cerro, Regla, La Habana del Este, La Lisa y La Habana Vieja (. . .)		
Terms topic	'habanero', 'provincia', 'fase', 'etapa', 'indisciplina'		
Opinions	Total: 70; Sentences: 225	Pos. Score	Neg. Score
Summary	Como habanero, me siento muy apenado de que el epicentro actual y cola de la epidemia de covid 19 en cuba sea debido al comportamiento de los pobladores en mi provincia.	4.5	12.4
	Soy habanero y siento lo que diré, lo que es una pena, pero con el anuncio de que matanzas y la habana son las únicas provincias que no entran en la fase 1 de la etapa recuperativa parece que esperan compulsar a los pobladores de la habana a disciplinarse para poder llegar a esa etapa cunado la tendencia de los últimos tiempos es exactamente lo contrario de cada vez mas indisciplina.	12.1	14.5
	Veo como va en aumento las personas en las calles y la indisciplina en general como no uso o el mal uso del nasobuco, las aglomeraciones, las personas en las calles	14.7	15.6
$JSD_{Opinions}$	0.255		
JSD_{News}	0.357		

In these examples, some fragments of the news and generated summaries were included to avoid further extension. These examples show summaries constituted by negatives and positives sentences, as well as the terms related to the most relevant opinion topics. Terms that more contribute to compute the polarity ratings (according to the SpanishSentiWordNet lexicon) are highlighted. Selected examples illustrate that the generated summaries are strongly related to the general meaning of the news content, still when the terminology used in both information units is different. The semantic relatedness with the most relevant identified topics is also appreciated. These results are achieved due to the semantic processing conceived in our model, which is carried out by integrating a semantic representation model (word2vec [14]) and two semantic similarity measures (Wu and Palmer [36] and the sentence-to-sentence similarity measure reported in Reference [37]).

Some sentences in the generated summaries are slightly extensive, which is fundamentally due to the opinion size is not restricted in the news platform used as opinion source—being another challenge to determine the relevance of the sentences with effectiveness. The longest sentences have more probability of obtaining higher relevance scores, since they can contain a higher number of terms semantically related to the news' content. Therefore, this suggests considering other sentence features, such as tf-idf and sentence length, and their integration to the sentence relevance assessment [48].

5. Conclusions and Future Works

In this paper, we have presented a news-focused opinion summarization approach that was designed according to the conception of extractive and topic-based text summarization methods. The proposed model can retrieve relevant sentences for the essential aspects of the news (context of interest), as well as cover the main topics of the opinionated texts in the generated summary. Our proposal integrates topic modeling, sentiment analysis, news-focused relevance scoring, and semantic analysis techniques. Several techniques and settings of our model were developed and evaluated with Spanish news and opinions regarding two different domains. The selected texts come from a real digital news platform.

The proposed model outperforms both adopted baselines, which are based on the classical text summarization method TextRank, obtaining automatic summaries more relevant to the news content, as well as covering the main topics in the opinionated texts well. The integration of term clustering, word embeddings, and similarity-based-sentence-to-news scoring turned out to be the more promising and effective setting of our model, due to its reaching the best values of Jensen–Shannon divergence concerning the news and very good values for all opinions. The use of semantic representation of words for applying similarity metrics was especially effective, resulting in the best option when the word embedding representation is used. Filtering the topics non-related with the news was a crucial step for generating automatic summaries aligned with the news, as well as the calculation of the semantic similarities of the sentences with the news to extract relevant sentences. The application of the explanatoriness-scoring technique in the sentences-ranking phase reached summaries that best cover the main topics in the opinionated texts. Nevertheless, it is necessary to point out that an important factor to achieve those good results was the integration of the topic-contextualization process, where the news is used to refine the identified topics from opinions. These results give us an idea that generally the topics treated in opinions are, in fact, closely related to a context that originates them (e.g., the news).

Despite promising results, several tasks could be considered as future works. Studying the effects of applying other clustering algorithms and similarity measures could contribute to obtaining better results. In the case that there are too-short sentences, to explore opinion and sentence augmentation could improve the opinion summarization process. Besides, it would be necessary to address the problem of the inverse polarity caused by the negation and integrate several sentiment lexicons in the sentiment analysis process. The use of other sentence features and the aggregation of their results for improving the relevance scoring should also be studied.

Author Contributions: Conceptualization, A.S.-C., A.R.-H. and M.M.G.L.; methodology, A.S.-C. and M.M.G.L.; software, A.R.-H.; validation, A.R.-H., A.S.-C. and M.M.G.L.; formal analysis, A.S.-C., A.R.-H. and M.M.G.L.; investigation, A.R.-H., A.S.-C., M.M.G.L., L.A. and J.S.-G.; resources, A.R.-H.; data curation, A.S.-C. and A.R.-H.; writing—original draft preparation, A.S.-C. and A.R.-H.; writing—review and editing, A.S.-C., A.R.-H., M.M.G.L., L.A. and J.S.-G.; visualization, A.S.-C.; supervision, A.S.-C. and M.M.G.L.; project administration, A.S.-C.; funding acquisition, A.S.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The work within the projects SAFER—PID2019-104735RB-C42 (AEI/FEDER, UE) and MERINET—TIN2016-76843-C4-2-R (AEI/FEDER, UE) supported by the Spanish Government.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Balahur, A.; Kabadjov, M.; Steinberger, J.; Steinberger, R.; Montoyo, A. Challenges and solutions in the opinion summarization of user-generated content. *J. Intell. Inf. Syst.* **2012**, *39*, 375–398. [CrossRef]
2. Kumar, A.; Sharma, A. Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence. *Webology* **2017**, *14*, 6–47.
3. Zhao, J.; Liu, K.; Xu, L. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. *Comput. Linguist.* **2016**, *42*, 595–598. [CrossRef]
4. Sun, S.; Luo, C.; Chen, J. A review of natural language processing techniques for opinion mining systems. *Inf. Fusion* **2017**, *36*, 10–25. [CrossRef]
5. Ravi, K.; Ravi, V. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowl. Based Syst.* **2015**, *84*, 14–46. [CrossRef]
6. Moussa, M.E.; Mohamed, E.H.; Haggag, M.H. A survey on opinion summarization techniques for social media. *Futur. Comput. Inform. J.* **2018**, *3*, 82–109. [CrossRef]
7. Condori, R.E.L.; Pardo, T.A.S. Opinion summarization methods: Comparing and extending extractive and abstractive approaches. *Expert Syst. Appl.* **2017**, *78*, 124–134. [CrossRef]
8. Li, P.; Huang, L.; Ren, G.-J. Topic Detection and Summarization of User Reviews. *arXiv* **2020**, arXiv:2006.00148.
9. Rossetti, M.; Stella, F.; Zanker, M. Analyzing user reviews in tourism with topic models. *Inf. Technol. Tour.* **2015**, *16*, 5–21. [CrossRef]
10. Chakraborty, R.; Bhavsar, M.; Dandapat, S.K.; Chandra, J. Tweet Summarization of News Articles: An Objective Ordering-Based Perspective. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 761–777. [CrossRef]
11. Kilgarriff, A.; Fellbaum, C. WordNet: An Electronic Lexical Database. *Language* **2000**, *76*, 706. [CrossRef]
12. Kamath, U.; Liu, J.; Whitaker, J. *Deep Learning for NLP and Speech Recognition*; Springer Nature Switzerland: Cham, Switzerland, 2019.
13. Yang, H.; Luo, L.; Chueng, L.P.; Ling, D.; Chin, F. Deep Learning and Its Applications to Natural Language Processing. In *Deep Learning: Fundamentals, Theory and Applications*; Huang, K., Hussain, A., Wang, Q.-F., Zhang, R., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2019; pp. 89–109.
14. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. In Proceedings of the 1st International Conference on Learning Representations (ICLR 2013), Scottsdale, AZ, USA, 2–4 May 2013.
15. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
16. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [CrossRef]
17. Allahyari, M.; Pouriyeh, S.; Assefi, M.; Safaei, S.; Trippe, E.D.; Gutierrez, J.; Kochut, K. Text Summarization Techniques: A Brief Survey. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 397–405. [CrossRef]
18. Abualigah, L.M.; Bashabsheh, M.Q.; Alabool, H.; Shehab, M. Text Summarization: A Brief Review. In *Recent Advances in NLP: The Case of Arabic Language, Studies in Computational Intelligence*; Abd El Aziz, M., Al-qaness, M.A.A., Ewees, A.A., Dahou, A., Eds.; Springer: Cham, Switzerland, 2020; pp. 1–15.
19. Gambhir, M.; Gupta, V. Recent automatic text summarization techniques: A survey. *Artif. Intell. Rev.* **2017**, *47*, 1–66. [CrossRef]
20. Amplayo, R.K.; Lapata, M. Informative and Controllable Opinion Summarization. *arXiv* **2019**, arXiv:1909.02322.

21. Lloret, E.; Boldrini, E.; Vodolazova, T.; Martínez-Barco, P.; Muñoz, R.; Palomar, M. A novel concept-level approach for ultra-concise opinion summarization. *Expert Syst. Appl.* **2015**, *42*, 7148–7156. [CrossRef]
22. Mukherjee, R.; Peruri, H.C.; Vishnu, U.; Goyal, P.; Bhattacharya, S.; Ganguly, N. Read what you need: Controllable Aspect-based Opinion Summarization of Tourist Reviews. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Xi'an, China, 25–30 July 2020; pp. 1825–1828. [CrossRef]
23. Jiang, Y.; Meng, W.; Yu, C. Topic Sentiment Change Analysis. In Proceedings of the Machine Learning and Data Mining in Pattern Recognition, MLDM 2011, New York, NY, USA, 30 August–3 September 2011; LNCS 6871. Springer: Berlin/Heidelberg, Germany, 2011; pp. 443–457. [CrossRef]
24. Ali, S.M.; Noorian, Z.; Bagheri, E.; Ding, C.; Al-Obeidat, F. Topic and sentiment aware microblog summarization for twitter. *J. Intell. Inf. Syst.* **2018**, *54*, 129–156. [CrossRef]
25. Rohit, S.V.K.; Shrivastava, M. Using Argumentative Semantic Feature for Summarization. In Proceedings of the 2019 IEEE 13th International Conference on Semantic Computing (ICSC), Newport Beach, CA, USA, 30 January–1 February 2019; pp. 456–461.
26. Abdi, A.; Shamsuddin, S.M.; Aliguliyev, R.M. QMOS: Query-based multi-documents opinion-oriented summarization. *Inf. Process. Manag.* **2018**, *54*, 318–338. [CrossRef]
27. Wang, L.; Raghavan, H.; Cardie, C.; Castelli, V. Query-Focused Opinion Summarization for User-Generated Content. In Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics, Dublin, Ireland, 23–29 August 2014; Dublin City University and Association for Computational Linguistics. pp. 1660–1669.
28. Conrad, J.G.; Leidner, J.L.; Schilder, F.; Kondadadi, R. Query-based opinion summarization for legal blog entries. In Proceedings of the 12th International Conference on Extending Database Technology Advances in Database Technology-EDBT '09, New York, NY, USA, 8–12 June 2009; pp. 167–176.
29. Luo, W.; Zhuang, F.; He, Q.; Shi, Z. Exploiting relevance, coverage, and novelty for query-focused multi-document summarization. *Knowl. Based Syst.* **2013**, *46*, 33–42. [CrossRef]
30. Ramón Hernández, A.; García Lorenzo, M.M.; Simón-Cuevas, A.; Arco, L.; Serrano-Guerrero, J. A semantic polarity detection approach: A case study applied to a Spanish corpus. *Procedia Comput. Sci.* **2019**, *162*, 849–856. [CrossRef]
31. Pennington, J.; Socher, R.; Manning, C. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Association for Computational Linguistics (ACL), Doha, Qatar,, 25–29 October 2014; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1532–1543. [CrossRef]
32. Verberne, S.; Krahmer, E.; Wubben, S.; Bosch, A.V.D. Query-based summarization of discussion threads. *Nat. Lang. Eng.* **2019**, *26*, 3–29. [CrossRef]
33. Angioni, M.; Devola, A.; Locci, M.; Tuveri, M.L.A.F. An Opinion Mining Model Based on User Preferences. In Proceedings of the 18th International Conference on WWW (Internet 2019), IADIS-International Association for the Development of the Information Society, Cagliari, Italy, 7–8 November 2019; pp. 183–185. [CrossRef]
34. Dalal, M.K.; Zaveri, M.A. Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews. *Appl. Comput. Intell. Soft Comput.* **2013**, *2013*, 1–8. [CrossRef]
35. Manning, C.; Prabhakar, R.; Schütze, H. *An Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008.
36. Pedersen, T.; Patwardhan, S.; Michelizzi, J. WordNet:Similarity-Measuring the Relatedness of Concepts. In Proceedings of the 19th National Conference on Artificial Intelligence (AAAI-04), San Jose, CA, USA, 25–29 July 2004; pp. 1024–1025.
37. Mihalcea, R.; Corley, C.; Strapparava, C. Corpus-based and Knowledge-based Measures of Text Semantic Similarity. In Proceedings of the 21st National Conference on Artificial Intelligence (AAAI'06), Boston, MA, USA, 16–20 July 2006; pp. 775–780.
38. Baccianella, S.; Esuli, A.; Sebastiani, F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the 7th International Conference on Language Resources and Evaluation, Valleta, Malta, 17–23 May 2010; pp. 2200–2204.
39. Amores, M.; Arco, L.; Borroto, C. Unsupervised Opinion Polarity Detection based on New Lexical Resources. *Comput. Sist.* **2016**, *20*, 263–277. [CrossRef]

40. Kim, H.D.; Castellanos, M.G.; Hsu, M.; Zhai, C.; Dayal, U.; Ghosh, R. Ranking explanatory sentences for opinion summarization. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR '13, Dublin, Ireland, 28 July–1 August 2013; p. 1069.
41. Mihalcea, R.; Tarau, P. TextRank: Bringing Order into Texts. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP'04), Barcelona, Spain, 25–26 July 2004; pp. 404–411.
42. Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. *Comput. Networks ISDN Syst.* **1998**, *30*, 107–117. [CrossRef]
43. Lin, C.-Y. Rouge: A Package for Automatic Evaluation of Summaries. In Proceedings of the Text Summarization Branches Out, Barcelona, Spain, 25–26 July 2004; pp. 74–81.
44. Louis, A.; Nenkova, A. Automatically evaluating content selection in summarization without human models. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–7 August 2009; pp. 306–314.
45. Saggion, H.; Torres-Moreno, J.M.; da Cunha, I.; SanJuan, E. Multilingual Summarization Evaluation without Human Models. In *Proceedings of the Coling 2010: Poster*; Beijing, China, 23–27 August 2010, Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 1059–1067.
46. Coavoux, M.; Elsahar, H.; Gallé, M. Unsupervised Aspect-Based Multi-Document Abstractive Summarization. In Proceedings of the 2nd Workshop on New Frontiers in Summarization, Hong Kong, China, 3–4 November 2019; pp. 42–47.
47. Elsahar, H.; Coavoux, M.; Gallé, M.; Rozen, J. Self-Supervised and Controlled Multi-Document Opinion Summarization Hady. *arXiv* **2020**, arXiv:2004.14754.
48. Valladares-Valdés, E.; Simón-Cuevas, A.; Olivas, J.A.; Romero, F.P. A Fuzzy Approach for Sentences Relevance Assessment in Multi-document Summarization. In *International Workshop on Soft Computing Models in Industrial and Environmental Applications*; Springer: Cham, Switzerland, 2019; pp. 57–67.