Semantic-Based Opinion Summarization

Marcio Lima Inácio and **Thiago Alexandre Salgueiro Pardo**
Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematics and Computer Science (ICMC)
University of São Paulo (USP)
Avenida Trabalhador São-Carlense, 400, 13566-590 São Carlos, SP, Brazil
marcio.lima.inacio@gmail.com, taspardo@icmc.usp.br

Abstract

The amount of information available online can be overwhelming for users to digest, specially when dealing with other users' comments when making a decision about buying a product or service. In this context, opinion summarization systems are of great value, extracting important information from the texts and presenting them to the user in a more understandable manner. It is also known that the usage of semantic representations can benefit the quality of the generated summaries. This paper aims at developing opinion summarization methods based on Abstract Meaning Representation of texts in the Brazilian Portuguese language. Four different methods have been investigated, alongside some literature approaches. The results show that a Machine Learning-based method produced summaries of higher quality, outperforming other literature techniques on manually constructed semantic graphs. We also show that using parsed graphs over manually annotated ones harmed the output. Finally, an analysis of how important different types of information are for the summarization process suggests that using Sentiment Analysis features did not improve summary quality.

1 Introduction

With the advance of web technologies during the last decades, a numerous amount of textual data is produced constantly, especially User Generated Content within social media and e-commerce domains, which include opinions towards many entities, such as products, organizations, and others. Such quantity of information is virtually impossible for other users to assimilate by their own, which creates a need for automated content selection and summarization methods, originating the Opinion Summarization research area, which integrates both the Text Summarization and Sentiment Analysis areas. (Kim et al., 2011; Liu, 2012).

There are two main types of summaries: extracts, which are composed by copying excerpts from the original texts literally, and abstracts, consisting of new text constructions derived out of the information obtained from the input data. Some authors in the literature call attention to the need of incorporating semantic knowledge into the summarization process (Mani, 2001; Li, 2015; Huang et al., 2020), particularly for producing abstracts, since humans usually create summaries by rewriting, paraphrasing and, mainly, interpreting texts (See et al., 2017; Dohare et al., 2018). This kind of information can be explicitly formalized as semantic representations, such as Predicate Argument Structure (PAS) (Khan et al., 2016, 2018), Abstract Meaning Representation (AMR) (Liu et al., 2015; Liao et al., 2018; Dohare et al., 2018) and others.

There are also arguments promoting the usage of explicit semantic representations to deal with texts from the opinions genre – in which the sentiment expressed by the author is an important information – as this type of linguistic knowledge is able to associate correlated information scattered throughout the texts and also to deal with many different semantic phenomena (such as negations and multiword expressions) that can alter the interpretation that would be obtained in a word-by-word basis (Cambria, 2013). However, there are, to the best of our knowledge, no works that focus on integrating these representations for opinion summarization.

In this context, this paper focuses on exploring techniques of Text Summarization based on explicit semantic representations upon opinions. As a base to instantiate the semantic knowledge, we used the Abstract Meaning Representation (AMR) (Banarescu et al., 2013), which has already been used in previous research in Text Summarization (Liu et al., 2015; Liao et al., 2018; Dohare et al., 2018) and has also been adapted for the Portuguese language in other former works (Anchiêta and Pardo,

Proceedings of Recent Advances in Natural Language Processing, pages 619–628
Sep 1–3, 2021.
https://doi.org/10.26615/978-954-452-072-4_070

2018b; Cabezudo and Pardo, 2019).

As a result, we present one of the first applications of methods based on explicit semantic representations for Opinion Summarization, showing that the Random Forest machine learning algorithm produced the best summary graphs, according to the metrics applied, outperforming other AMR-based approaches from the literature for manually annotated graphs. The results also indicate, in the same manner as previous works (Liu et al., 2015), that the quality of the AMR graphs used have an impact on the output. The experiments also suggest that using Sentiment Analysis features do not improve summary quality.

In short, the main contributions of this paper are the following:

- Development of four new summarization methods based on the AMR representation with results compatible to the current state of the art on using Semantic Representations for Text Summarization;
- Investigation of how other methods of the literature, not developed initially with opinions in mind, behave in a sentiment analysis context;
- Implementation of a semantic-based opinion summarization tool, which is publicly available[1];
- Analysis of the importance of proposed features for text summarization.

The remaining of the text is organized as follows: in section 2, we present a general notion of the AMR. Later, in section 3, previous works in Text Summarization are outlined with main focus on the ones based on the chosen semantic representation (AMR). Our methods are shown in section 4 followed in section 5 by some important notes about the experiments executed, whose results are then introduced in section 6. Finally, some conclusions and further discussing are made in section 7.

2 Abstract Meaning Representation

Sentences in AMR are represented as rooted directed graphs in which nodes correspond to concepts and edges to relations between them. An example can be seen in Figure 1. Concepts can be of different kinds, e.g framesets (`possible-01`, `recommend-01`, `ler-01`) originated from separate repositories containing predicates and their corresponding arguments. For the English language, this repository is called Propbank[2] (Kingsbury and Palmer, 2002), whilst for Brazilian Portuguese, there is VerboBrasil[3] (Duran and Aluísio, 2015).

The concepts can also be words from the base language, extracted directly, as they are, from the sentence (`e`, `eu`, `coisa`) or derived by some annotation procedure provided in the guidelines, such as the transformation of the comparative "melhor" (*better*) into a degree relation between "mais" (*more*) and "bom" (*good*) (`bom`, `mais`). There are also concepts that are specific of the representation formalism (e.g., `have-degree-91`).

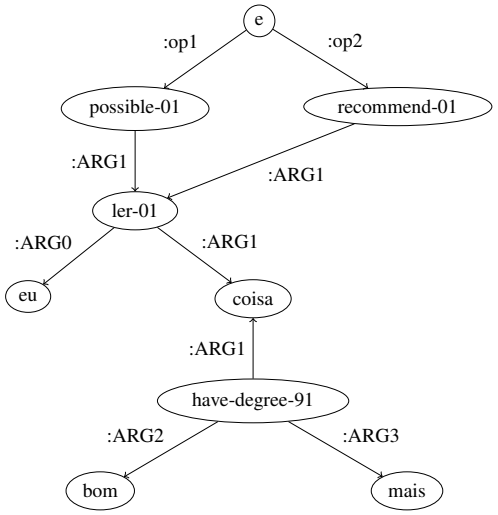

Figure 1: Example of AMR graph corresponding to the sentence "Eu poderia e deveria ter lido coisa melhor" (*I could and should have read something better*).

All edges are of two types: argument relations defined for each predicate (`ARG0`, `ARG1`, `ARG2`, `ARG3`) and the ones determined by the AMR project (`op1`, `op2`).

As the original representation was developed specifically for the English language, there is need for adaptations of the guidelines[4] towards other languages. For Brazilian Portuguese, this initiative was made by Anchiêta and Pardo (2018b), followed by Cabezudo and Pardo (2019), dealing with some phenomena that are not so common in English,

[1]The code for each method, as well as for the experiments performed for this paper, is available at: `https://github.com/Superar/SemOpinionS`.

[2]Available at: `http://verbs.colorado.edu/verb-index/index.php`.

[3]Available at: `http://143.107.183.175:21380/verbobrasil/`.

[4]The original guidelines are available at: `https://github.com/amrisi/amr-guidelines`.

such as indeterminate and hidden subjects.

3 Related Work

There are two main areas related to our work: Summarization based on Semantic Representations and Opinion Summarization, each with a long history in the literature. Thus, we present the works most related to ours, those using the Abstract Meaning Representation and other Opinion Summarization methods.

3.1 Summarization Based On Abstract Meaning Representation

The usage of semantic representations for Text Summarization can be traced back until the 1980s, with the TOPIC system (Reimer and Hahn, 1988). Since then, numerous techniques upon various representations have been developed, such as UNL (Martins and Rino, 2001; Mangairkarasi and Gunasundari, 2012), LNS (Rusu et al., 2009), RSG (Moawad and Aref, 2012) and many others. With a special focus on the AMR representation, used in this paper, there are some works – exclusively for the English language – that are going to be explained more thoroughly below.

The first incorporation of AMR into the Summarization process was conducted by Liu et al. (2015). Their method uses several numerical features to represent nodes and edges, used to calculate a score for the given graph; each attribute has a weight trained through AdaGrad (Duchi et al., 2011), so that graphs similar to gold standard handmade summary graphs are assigned to higher scores. Afterwards, Integer Linear Optimization (ILP) is used upon the original text AMR graph in order to select the subgraph which maximizes the scoring. The authors report a maximum F-score for nodes prediction of 58.7% and of 39% regarding edges.

Later, this technique was enhanced by Liao et al. (2018), as they included new attributes, and a previous sentence clustering step using Spectral Clustering (von Luxburg, 2007), which leverages similarity metrics between each clause, in order to identify the multiple topics in the input. Then, a small amount of sentences are selected from each cluster to proceed with the ILP summarization. They obtained a prediction F-Score of 30.1% for nodes and 9.8% for edges, it is important to note that they used exclusively parsed graphs, which have already been indicated as a factor that harms the results (Liu et al., 2015).

In parallel to these techniques, an unsupervised approach for Summarization using AMR has been developed by Dohare et al. (2018). This method uses the TF-IDF score of each concept in order to select a subset of relevant nodes to serve as a basis for the summary graph. Then, a few rules are used to determine the most important path between each pair of nodes chosen and to, finally, expand these paths by adding edges and nodes according to OpenIE (Banko et al., 2008) triples. This unsupervised method obtained a node prediction F-score of 60.4%, however there is no report for edges.

3.2 Opinion Summarization Approaches

This paper is also included within the Opinion Summarization area, which, to the extent of our knowledge, has no initiative of using explicit semantic representations during this process. Therefore, we focus here on two main works related to ours.

The Opinosis method (Ganesan et al., 2010) uses a graph representation obtained from the text to perform the summarization of opinions. It is important to note that this graph does not convey the semantics of the input, but rather the order with which the tokens occur. A single graph encloses multiple sentences, so that identical tokens are merged into a single node, capturing redundancy. Then, a set of rules are applied upon this graph to find those relevant valid paths, prioritizing nodes which occur in more input sentences. Finally the summary text is composed from these selected paths of tokens with some other additional rules for combining them through conjunctions.

More recently, the Opizer system (Condori and Pardo, 2017) was developed as one of the few projects of Opinion Summarization with a specific focus on the Portuguese language. It comprises two main methods: an extractive (Opizer-E) and an abstractive one (Opizer-A). In Opizer-E, the sentences are first clustered together with respect to their general sentiment polarity and also to which aspect of the product they describe (e.g. the camera, battery, durability, etc.). Then, they are ranked according to an importance score which leverages the sentence position within the document and how the aspects and their corresponding sentiment words are related. The most important sentences are then selected to compose the summary.

In its turn, Opizer-A focuses on the selection of salient n-grams rather than full sentences, clustering them according to the polarity, aspect and con-

tent. Afterwards, a set of representative n-grams is chosen from each cluster with respect to their TF-IDF scores. Lastly, they are used to fill in a predefined textual pattern, resulting in the final abstract.

From this overview of the literature, some methods can be derived and developed, as we present in the following.

4 Methods

As discussed before, the usage of explicit semantic information represents an interesting path for both Text Summarization and Sentiment Analysis, consequently for Opinion Summarization. Thus, we chose the Abstract Meaning Representation as an instance of semantic knowledge. The starting point of our research is, therefore, all the previous works which exploited this representation in the process of summarizing texts (Liu et al., 2015; Liao et al., 2018; Dohare et al., 2018), even if for other genres rather than opinions.

First, some preprocessing is necessary. As AMR is a sentence-level representation, we need to combine multiple sentences into a multi-document graph. This has been done, in this work, similarly to Liu et al. (2015) by merging every node that contains the same concept. Some semantic units – such as named entities or dates – are represented in multiple nodes, which are collapsed before the merging, only allowing the combination of entities with the same overall information (same name or date). This was applied upon all comments about each product, resulting in a multi-document semantic representation of all opinions about some given book or electronic device.

4.1 Rule-based Method

As discussed by Dohare et al. (2018), all steps of their method can be directly applied to multi-document summarization. In this paper, we argue that the usage of the traditional TF-IDF scoring may not be suitable for a multi-document scenario, since it penalizes concepts that occur in a larger range of texts, potentially penalizing important information that is common to a large number of comments. For example, if a lot of comments state that the camera quality of a given smartphone is poor, this information should not be penalized, but rather rewarded. Therefore, we propose a simple variation of this method, by substituting the ranking score form TF-IDF to TF only, i.e. the number of occurrences of each concept across the texts.

4.2 Score Optimization Method

Following the objective of improving the original Dohare et al. (2018) method, we decided to explore the usage of multiple different information in order to enhance the ranking of important nodes.

Similarly to Liu et al. (2015), various features may be applied to represent each node or edge in the graph and then weights can be used to combine these attributes into a salience score for the given element. This idea has been transposed into our work, however we do not use the same features as Liu et al. (2015), as they turn each one into multiple boolean values without any specification of how the thresholds were defined. Thus, we decided to develop our own feature set based on the work of Leskovec et al. (2005), which used graph-based and linguistic features to perform the classification of relevant graph nodes for summarization. Our feature set is presented in Table 1.

As can be seen, our features are almost exclusively focused on the graph structure, excluding four last ones (TF, TF-IDF, Sentiment polarity and Aspect). We decided to keep both TF and TF-IDF, as they can convey some differences, as discussed before, and a combination of both may be interesting. Both Sentiment polarity and Aspect features are common within the Opinion Summarization area (Condori and Pardo, 2017). It is important to report that Leskovec et al. (2005) also states that they used another 118 linguistic features that did enhance the summaries quality, however there is no indication of which these features were, so we decided to not include any other features, as it would be necessary to do a more focused feature engineering research in order to determine which other linguistic information is important and how they should be represented.

After the definition of the feature set, they should be combined into a score for each node. This is accomplished through a linear combination of all attributes with an optimized weight vector, so that important nodes have higher scores. The weights are optimized through the Simulated Annealing method, as it is designed to be capable of escaping from local minima or maxima. To this extent, we used the algorithm developed by Ludermir et al. (2006) for weight optimization in Neural Networks. To obtain the target class of each node (if it should be considered relevant or not), the extractive sum-

Feature	Brief description
Incoming degree	Number of edges pointing towards the given node.
Average neighbor degree	Average of the incoming degree values of all neighbors of the given node.
Degree centrality	Proportion of nodes from the graph that are connected to the given node.
Eigenvector centrality	Recursive method that assigns higher values to nodes that have a higher incoming degree, taking into account that more important nodes have a stronger "voting power". (Newman, 2010)
Pagerank	Similar to the Eigenvector Centrality, but also considering the amount of outcoming edges of an important node. (Brin and Page, 1998)
HITS	This method results into two coefficients: hub and authority. "Hub" nodes are those which point out to multiple "authority" nodes and vice versa. The scores are calculated jointly. (Kleinberg, 1999)
Closeness	The importance of a given node is computed as the inverse of the average distance of the given node to all others in the graph. (Newman, 2010, p.181)
Betweeness	This is calculated as the fraction of shortest paths between each pair of vertices that go through the given node. (Newman, 2010, p.186)
Local clustering coefficient	Defined as the proportion of triangles formed between the given node and its neighbors among all triangles possible. (Newman, 2010, p.202)
Depth	Defined as the minimum distance between the root of the graph and the given node.
TF	Term frequency. Indicates the number of occurrences of a concept over all input comments.
TF-IDF	The original scoring used by Dohare et al. (2018). It rewards concepts that occur with a high TF value, but penalizes those that are too common throughout a given corpus.
Sentiment polarity	Three boolean attributes, indicating if the sentiment of the given concept is either positive, negative or neutral according to the OpLexicon (Souza et al., 2011).
Aspect	Manual aspect annotation provided by Condori et al. (2015) within the OpiSums-PT corpus. This boolean feature indicates if a given concept represents an aspect of the main product or not.

Table 1: Features used to represent AMR graph nodes

maries provided within the OpiSums-PT corpus are used, i.e. if a sentence was selected to compose an extract, all nodes from its representation are considered important.

After the ranking of each node in the graph via the score described above, the summarization process is carried out in the same way as the original work of Dohare et al. (2018).

4.3 Machine Learning Methods

As there is a new whole set of features developed (Table 1), they can be used as a training input for some Machine Learning algorithms aiming at classifying nodes as relevant or not, similarly to the original work of Leskovec et al. (2005). The algorithms explored in this work are the following: SVM (originally used by Leskovec et al. (2005)); Decision Trees and Random Forest, which are straightforward interpretable algorithms; and Multilayer Percepton, a simple Neural Network method, which is a popular research area currently, specially through Deep Learning[5]. From the selected nodes, the same original rules can be used upon them to

[5]We, however, do not explore Deep Learning methods, as they usually require high quantities of data, which we do not have available.

create the final summary graph.

We also explored another machine learning based approach in order to overcome the limitations of the set of rules used, as they prioritize the first sentences of the corpus, which may not be ideal in a multi-document setting. Therefore, one can use the concept of Levi Graphs to also include the relations into the classification problem. Levi Graphs are AMR representations within which the relations (edges) are turned into nodes connected to each original extreme of the edge (Beck et al., 2018). In this way, the relations can also be represented as a feature vector and, consequently, they can be classified directly by the algorithms, dismissing the use of rules for this purpose.

As the classification of edges can lead to disconnected graphs, which would be against the representation guidelines, the largest connected component is selected as the final summary graph.

5 Experiments Setup

As a source of opinions, we used the OpiSums-PT corpus[6] (Condori et al., 2015), comprising the opinions from ReLi (Freitas et al., 2014) (concerning 13 books) and also comments about 4 electronic products, obtained from the Buscapé e-commerce website[7]. This corpus also comprehends 10 summaries for each product, 5 extractive and 5 abstractive, created by 14 human specialists. Thus, the corpus contains 171 opinionative documents with a total of 1,502 sentences.

Following the AMR guidelines for the Portuguese language, 404 sentences from the corpus were manually annotated[8]. Two products of the OpiSums-PT corpus have been fully annotated into AMR to this date. Thus, in this work, they are used as a gold standard test set. In order to compare how the quality of the graphs used with each method interfere in the results, we used an automatic AMR parser (Anchiêta and Pardo, 2018a) to annotate two similar products[9], resulting in a total of 4 products in the test set.

We also made use of the parser in order to create an artificial training corpus for the proposed supervised approaches (via score optimization or Machine Learning) from the 13 remaining products of the OpiSums-PT, resulting in 65 examples (pairs of opinions graphs and their corresponding summary graph), with a total of 3,290 AMR sentences, from which 388 are considered relevant to compose the summaries.

Another important resource used is the B2W-Reviews01 corpus[10] (Real et al., 2019), which was used, in conjunction with the ReLi (Freitas et al., 2014) corpus (excluding the two products used for the testing) to calculate the Document Frequency term in the TF-IDF counts.

For evaluation, we adopted two metrics from the AMR research community: Smatch (Cai and Knight, 2013) and SEMA (Anchiêta et al., 2019), which are used to compare two AMR graphs (the automatic summary graph and its corresponding gold standard manually annotated graph). As the OpiSums-PT corpus provides 5 summaries for each product, all of them have been used to evaluate the summarization output, however, due to lack of space, we report only the average of these values.

We also would like to call attention to the fact that the Simulated Annealing optimization is sensitive to its random initialization, so it was evaluated by running the same experiment 10 times with different starting weights. The values reported in this paper are the average upon all experiments.

6 Results

The results for every experiment executed are shown in Table 2. For fairness, we compare our methods to other AMR-based summarization approaches, since there is no robust Natural Language Generation tool from AMR to text in Portuguese.

As can be seen, the Random Forest algorithm using the Machine Learning method without Levi graphs, proposed in this work, had the best results among the manual annotated comments, resulting in a Smatch improvement of 15% from the best literature method (Liao et al., 2018). Meanwhile, the Integer Linear Programming based ones (Liu et al., 2015; Liao et al., 2018) were the best for the parsed sentences, indicating that their method is more robust with regards to the quality of the input graphs. The quality of the graphs is, how-

[6] Available at: https://sites.google.com/icmc.usp.br/opinando.

[7] Available at: https://www.buscape.com.br.

[8] Available at: https://github.com/nilc-nlp/AMR-BP.

[9] The manual annotated products are the Iphone 5 smartphone and the "O Apanhador no Campo de Centeio" book, similarly we automatically parsed the comments for the Galaxy SIII smartphone and the "O Outro Lado da Meia Noite" book.

[10] Available at: https://github.com/b2wdigital/b2w-reviews01.

Method	Smatch		SEMA	
	Manual	Parser	Manual	Parser
Liu et al. (2015)[11]	0.32	0.22	0.28	**0.19**
Liao et al. (2018)[12]	0.40	**0.32**	0.28	0.18
Dohare et al. (2018)[13]	0.38	0.29	0.27	0.12
Rule-based method	0.34	0.24	0.26	0.10
Score optimization method[14]	0.38	0.26	0.30	0.13
Machine Learning: SVM[15]	0.38	0.27	0.24	0.05
Machine Learning: Decision Tree[16]	0.37	0.24	0.32	0.18
Machine Learning: Random Forest[17]	**0.46**	0.27	**0.37**	0.10
Machine Learning: MLP[18]	0.38	0.29	0.24	0.10
Machine Learning with Levi graphs: SVM	0.33	0.21	0.32	0.13
Machine Learning with Levi graphs: Decision Tree	0.32	0.21	0.31	0.17
Machine Learning with Levi graphs: Random Forest	0.32	0.22	0.32	0.16
Machine Learning with Levi graphs: MLP	0.33	0.20	0.30	0.13

Table 2: Summarization results

ever, still an important factor to have satisfactory summaries, since the usage of parsed AMR representations caused a drop on the metrics for each approach, even among the supervised ones, which were trained over an artificial parsed corpus.

It is also important to notice that the usage of Levi graphs to select important relations did not actually improve the results, suggesting that the set of rules for edge selection and path expansion, originally developed by Dohare et al. (2018), are also suitable for a multi-document purpose.

Comparing our rule-based method with the original technique of Dohare et al. (2018), the adoption of the TF counts over TF-IDF decreased the quality of summaries, refuting our initial hypothesis that the IDF term would harm the content selection.

Regarding the score optimization method, it did not vary much from its original counterpart – the one from Dohare et al. (2018) – indicating that the scoring did not improve from one to the other. We do, however, highlight that this does mean that the attributes set is not fit to the problem, since it worked well with the Random Forest algorithm. Thus, the research of other types of weight optimization techniques, such as Genetic Algorithms as used in the work by Khan et al. (2018), may lead to better results. It is also expected that the inclusion of other linguistic features may improve the outcome (Leskovec et al., 2005).

From the optimized weights, however, we can see how important each of the developed features are. This is accomplished by calculating the correlation of the absolute value for the weights with the evaluation metrics used. This result is shown in Figure 2, in which the more intense the green color is, the more important its corresponding feature is.

As can be seen in the heatmap, the most important features to be noticed are the average neighbor degree, the eigenvector centrality and the depth, with some attention to the clustering coefficient feature. On the other hand, the most detrimental features are those which are assigned to deeper shades of pink. There are not much features which can be considered effectively harmful to the quality of the summaries – i.e. there are no features which have such a high negative correlation to the evaluation metrics – but it can be observed that both the TF-IDF and the Hubs features do have a stronger negative correlation.

The lighter colors are related to the features that

[11] Using Ramp loss. For training: $\eta = 1.0, \epsilon = 1.0$, as used by the original authors.

[12] Using the Longest Common Subsequence similarity metric, $N = M = 5$.

[13] Selecting 10 relevant nodes.

[14] $I_{\max} = 1000, K = 5, T_0 = 1.0, r = 0.9, I_T = 10$ (Ludermir et al., 2006).

[15] Using RBF Kernel.

[16] Using the Gini index as a node splitting criterion.

[17] With 100 trees.

[18] Using 100 hidden units. Activation function: ReLu. Learning rate: 10^{-4}.

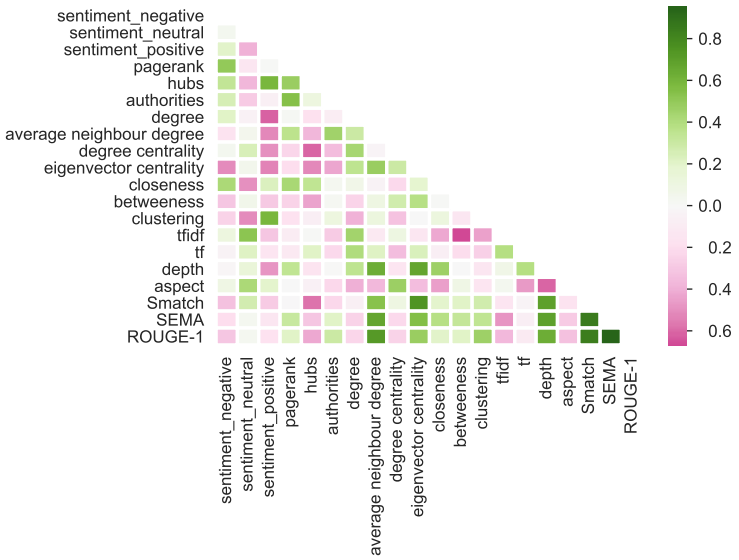

Figure 2: Correlation heatmap of the optimized weights including the evaluation metrics

may not have had an impact over the output quality. The sentiment analysis features (sentiment and aspect, commonly used in Opinion Summarization approaches) do not appear to have a role of importance during the summarization process, even though we are dealing with opinions.

The feature analysis also highlights our previous observation that the feature set comprises other types of important information. Although there was not much difference (using Smatch or SEMA) between the score optimization method and the original one from Dohare et al. (2018), the results were obtained taking into account different features (e.g., eigenvector centrality vs. TF-IDF). This indicates that there is indeed room for improvement by combining these forms of knowledge.

7 Conclusion

The incorporation of semantic knowledge via explicit representations into the summarization process is a fruitful field of research. In this context, this paper focused on the development of methods for opinion summarization that took advantage of the semantics encoded within AMR graphs.

The experiments indicate that our Machine Learning approach, using the Random Forest algorithm, produced better summary graphs compared to other AMR-based summarization techniques. We also confirm the findings of Liu et al. (2015), stating that the quality of the input graphs used is important to ensure the creation of good summaries. An analysis of the feature set also suggests that using sentiment and aspect information does not actually improve the results.

As future work, we recommend the further development of the feature set, since it has been shown to be effective for a Machine Learning approach, including mainly new linguistic features, as originally proposed by Leskovec et al. (2005). We also advocate that exploring other optimization techniques can be an important path of investigation, such as the Genetic Algorithm approach used by Khan et al. (2018).

There is also a significant improvement to be done during the merging of sentence graphs into a single document representation. Currently, only the concept labels are taken into account, however the usage of coreference resolution can be a valuable enhancement towards the creation of a consistent multi-sentence representation (Liao et al., 2018; Dohare et al., 2018; O'Gorman et al., 2018).

Acknowledgments

This research was carried out with the support of Coordination for the Improvement of Higher Education Personnel (CAPES #88887.355305/2019-00). We also had the support of the OPINANDO project[19] (PRP #668) and the POeTiSA project[20] of the Center for Artificial Intelligence (C4AI)[21] (IBM and FAPESP #2019/07665-4).

[19] https://sites.google.com/icmc.usp.br/opinando/

[20] https://sites.google.com/icmc.usp.br/poetisa

[21] https://c4ai.inova.usp.br/

References

Rafael Torres Anchiêta, Marco Antonio Sobrevilla Cabezudo, and Thiago Alexandre Salgueiro Pardo. 2019. SEMA: An extended semantic evaluation for AMR. In *Proceedings of the 20th Computational Linguistics and Intelligent Text Processing*. Springer International Publishg.

Rafael Torres Anchiêta and Thiago Alexandre Salgueiro Pardo. 2018a. A Rule-Based AMR Parser for Portuguese. In *Advances in Artificial Intelligence - IBERAMIA 2018*, volume 11238, pages 341–353, Trujillo. Springer International Publishing.

Rafael Torres Anchiêta and Thiago Alexandre Salgueiro Pardo. 2018b. Towards AMR-BR: A SemBank for brazilian portuguese language. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pages 974–979, Miyazaki, Japan. European Languages Resources Association.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Boardhead, and Oren Etzioni. 2008. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74.

Daniel Beck, Gholamreza Haffari, and Trevor Cohn. 2018. Graph-to-Sequence Learning using Gated Graph Neural Networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 273–283, Melbourne, Australia. Association for Computational Linguistics.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117.

Marco Antonio Sobrevilla Cabezudo and Thiago Pardo. 2019. Towards a general Abstract Meaning Representation corpus for Brazilian Portuguese. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 236–244, Florence, Italy. Association for Computational Linguistics.

Shu Cai and Kevin Knight. 2013. Smatch: An evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 748–752.

Erik Cambria. 2013. An introduction to concept-level sentiment analysis. In *Advances in Soft Computing and Its Applications. MICAI 2013.*, pages 478–483, Mexico City, Mexico. Springer, Berlin, Heidelberg.

Roque Enrique López Condori and Thiago Alexandre Salgueiro Pardo. 2017. Opinion summarization methods: Comparing and extending extractive and abstractive approaches. *Expert Systems With Applications*, 78:124–134.

Roque Enrique López Condori, Thiago Alexandre Salgueiro Pardo, Lucas Vinicius Avanço, Pedro Filho, Alessandro Bokan, Paula Cardoso, Márcio Dias, Fernando Nóbrega, Marco Cabezudo, Jackson Souza, Andressa Zacarias, Eloize Seno, and Ariani Di Felippo. 2015. A qualitative analysis of a corpus of opinion summaries based on aspects. In *Proceedings of the 9th Linguistic Annotation Workshop*, pages 62–71, Denver, Colorado, USA. Association for Computational Linguistics.

Shibhansh Dohare, Vivek Gupta, and Harish Karnick. 2018. Unsupervised Semantic Abstractive Summarization. In *Proceedings of ACL 2018, Student Research Workshop*, pages 74–83, Melbourne, Australia. Association for Computational Linguistics.

John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159.

Magali Sanches Duran and Sandra Maria Aluísio. 2015. Automatic generation of a lexical resource to support semantic role labeling in portuguese. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 216–221, Denver, Colorado. Association for Computational Linguistics.

Cláudia Freitas, Eduardo Motta, Ruy Luiz Milidiú, and Juliana César. 2014. Sparkling vampire... LOL! Annotating opinions in a book review corpus. In *New Language Technologies and Linguistic Research*, pages 128–146, Newcastle upon Tyne. Corpus Linguistics Symposium, Cambridge Scholars Publishing.

Kavita Ganesan, ChengXiang Zhai, and Jiawei Han. 2010. Opinosis: A graph based approach to abstractive summarization of highly redundant opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 340–348, Beijing, China. Coling 2010 Organizing Committee.

Luyang Huang, Lingfei Wu, and Lu Wang. 2020. Knowledge Graph-Augmented Abstractive Summarization with Semantic-Driven Cloze Reward. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5094–5107, Online. Association for Computational Linguistics.

Atif Khan, Naomie Salim, and Haleem Farman. 2016. Clustered genetic semantic graph approach for multi-document abstractive summarization. In *2016 International Conference on Intelligent Systems Engineering (ICISE)*, pages 63–70.

Atif Khan, Naomie Salim, Haleem Farman, Murad Khan, Bilal Jan, Awais Ahmad, Imran Ahmed, and Anand Paul. 2018. Abstractive text summarization based on improved semantic graph approach. *International Journal of Parallel Programming*, 46(5):992–1016.

Hyun Duk Kim, Kavita Ganesan, Parikshit Sondhi, and ChengXiang Zhai. 2011. Comprehensive review of opinion summarization. Technical report, University of Illinois at Urbana-Champaign, Urbana, IL, EUA.

Paul Kingsbury and Martha Palmer. 2002. From Treebank to PropBank. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 1989–1993, Las Palmas. European Language Resources Association.

Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632.

Jure Leskovec, Natasa Milic-Frayling, and Marko Grobelnik. 2005. *Extracting Summary Sentences Based on the Document Semantic Graph*. Microsoft Research.

Wei Li. 2015. Abstractive multi-document summarization with semantic information extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1908–1913, Lisbon, Portugal. Association for Computational Linguistics.

Kexin Liao, Logan Lebanoff, and Fei Liu. 2018. Abstract Meaning Representation for Multi-Document Summarization. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1178–1190, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Bing Liu. 2012. Sentiment analysis and opinion mining. In Graeme Hirst, editor, *Synthesis Lectures on Human Language Technologies*, volume 5, pages 1–167. Morgan & Claypool Publishers.

Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A. Smith. 2015. Toward Abstractive Summarization Using Semantic Representations. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1077–1086, Denver, Colorado. Association for Computational Linguistics.

Teresa Bernarda Ludermir, Akio Yamazaki, and Cleber Zanchettin. 2006. An optimization methodology for neural network weights and architectures. *IEEE Transactions on Neural Networks*, 17(6):1452–1459.

S Mangairkarasi and S Gunasundari. 2012. Semantic based Text Summarization using Universal Networking Language. *International Journal of Applied Information Systems*, 3(8):18–23.

Inderjeet Mani. 2001. *Automatic Summarization*, volume 3 of *Natural Language Processing*. John Benjamins Publishing Company, Amsterdam.

Camilla Brandel Martins and Lucia Helena Machado Rino. 2001. Pruning UNL texts for Summarizing Purposes. In *NLPRS*, pages 539–544.

Ibrahim F. Moawad and Mostafa Aref. 2012. Semantic graph reduction approach for abstractive Text Summarization. In *2012 Seventh International Conference on Computer Engineering Systems (ICCES)*, pages 132–138.

M. E. J. Newman. 2010. *Networks: An Introduction*. Oxford University Press, Oxford ; New York.

Tim O'Gorman, Michael Regan, Kira Griffitt, Ulf Hermjakob, Kevin Knight, and Martha Palmer. 2018. AMR beyond the sentence: The multi-sentence AMR corpus. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3693–3702, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Livy Real, Marcio Oshiro, and Alexandre Mafra. 2019. B2W-Reviews01: An open product reviews corpus. In *Proceedings of the 12th Brazilian Symposium in Infomation and Human Language Technology*, pages 200–208, Salvador. Brazilian Symposium in Information and Human Language Technology, Sociedade Brasileira de Computação.

Ulrich Reimer and Udo Hahn. 1988. Text condensation as knowledge base abstraction. In *Proceedings of the Fourth Conference on Artificial Intelligence Applications*, pages 338–344, San Diego, California. Conference on Artificial Intelligence Applications, Institute of Electrical and Electronics Engineers.

Delia Rusu, Blaž Fortuna, Marko Grobelnik, and Dunja Mladeniæ. 2009. Semantic graphs derived from triplets with application in document summarization. *Informatica*, 33(3):357–362.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1073–1083, Vancouver, Canada. Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics.

Marlo Souza, Renata Vieira, Débora Busetti, Rove Chishman, Isa Mara Alves, and Faculdade De Letras Unisinos. 2011. Construction of a portuguese opinion lexicon from multiple resources. In *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology*, Cuiabá, Mato Grosso, Brazil.

Ulrike von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.