A Study of Abstractive Summarization Using Semantic Representations and Discourse Level Information

Gregory César Valderrama Vilca[1] and Marco Antonio Sobrevilla Cabezudo[2(✉)]

[1] Escuela de Posgrado, Maestría en Informática,
Pontificia Universidad Católica del Perú, Lima, Perú
a20133303@pucp.edu.pe
[2] Grupo de Reconocimiento de Patrones e Inteligencia Artificial Aplicada,
Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima, Perú
msobrevilla@pucp.edu.pe

Abstract. The present work proposes an exploratory study of abstractive summarization integrating semantic analysis and discursive information. Firstly, we built a conceptual graph using some lexical resources and Abstract Meaning Representation (AMR). Secondly, we applied PageRank algorithm to get the most relevant concepts. Also, we incorporated discursive information of Rethorical Structure Theory (RST) into the PageRank to improve the relevant concepts identification. Finally, we made some rules over the relevant concepts and applied SimpleNLG to make the summaries. This study was performed on the corpus of DUC 2002 and the results showed a F1-measure of 24% in Rouge-1 when AMR and RST were used, proving their usefulness in this task.

Keywords: Abstractive summarization · Abstract Meaning Representation · Rethorical Structure Theory

1 Introduction

The web is a giant resource of data and information that has great utility for people. However, getting an abstract about one or many documents is an expensive labor, which with manual process might be impossible to complete due to the huge amount of data.

Automatic Summarization [12] is a challenging task, because it involves analysis and comprehension of the written text in non-structural natural language and it is dependent of a context that must describe an event synthesis or knowledge in a simple form, becoming natural for any reader. There are diverse approaches to summarize text and categorize into extractive or abstractive.

Abstractive summaries regenerate the content extracted from source text by terms fusion, compression or suppression processes. Thus, paraphrased sentences are obtained and these are not in the original text. This approach has a major

probability to reach coherence and smoothness like one generated or made by a human beings.

Previous work has shown progress using semantic representations such as Abstract Meaning Representation (AMR) presented in [11], Discursive Analysis with Rhetorical Structure Theory (RST) present in [6] and conceptual models using linguistic resources such as WordNet present in [14].

This work presents an exploratory study of how to integrate semantic (AMR annotator) and discursive (RST annotator) information into an abstractive summarization method produces better results. In a first phase, the method generated a conceptual graph using AMR parsing and other lexical resources like WordNet and PropBank [16]. Thus, to find the most relevant concepts we use PageRank, considering all discursive information given by the O'Donell method application. Then, sentence candidates are built with the most important concepts and semantic roles information. Finally, an abstractive summary is generated using SimpleNLG, as Natural Language Generation tool, over the sentence candidates. This shows that using these techniques are workable and even more profitable, recommended configurations and useful tools for this task.

This paper organization is: first, the Sect. 2 presents the related works, Sect. 3 presents the proposed method, Sect. 4 presents experiments and results. Finally, Sect. 5 presents some conclusions and future works.

2 Related Works

The performance of extractive and abstract techniques was tested in [2], not only the automatic methods but also the summaries made by people. It conclude that in the linguistic-grammatical aspect, and in the quality of the content, summaries generated by humans are far superior to those generated automatically, and the abstractive methods have more possibilities to achieve results more similar to their human counterparts.

In [14] we can observe an intermediate representation models and the use of knowledge sources presented on the Web. The authors generated summaries of a single document using a semantic representation of texts through conceptual graphs, in which, the weights are associated with the edges linking concept nodes, creating a flow called "semantic flow". A semantic flow is the weight accumulated by the nodes and that transmit to other nodes increasing or decreasing its value when passing through any conceptual relation. For the graph generation, the authors used the semantic information from external sources as *WordNet* [8] and *VerbNet* [3] that rule the structural coherence of the graphs. In the synthesis stage, the graphs were reduced according to a set of generalization, union, weighting and pruning operations shown in [7]. In [6], the authors present opinion summarization by an abstractive method based on the analysis of the structures and relations of the discourse, and also they proposed a method to generate new sentences that uses the PageRank algorithm to identify the most important content.

In [11], the authors used AMR [10] for the representation and generation of abstract summaries for a single document. The authors generated an AMR

graph for each document sentence using JAMR parser [4]. The AMR graphs are merged based on the concepts common between them. Thus, generate an unique graph for a document that reduced its concept redundancy.

3   Abstractive Summarization Method

We used the architecture proposed by [12] which comprises three stages: (1) in the analysis phase, input text are interpreted and represented in a computational format; (2) in the transformation phase, representation mentioned in first phase, is processed to identify and select the content more relevant and as a result a condensed computational representation of texts is got, and (3) in the synthesis phase, a natural language text is generated. In the Fig. 1, we may see the pipeline of the proposed Abstractive Summarization method.
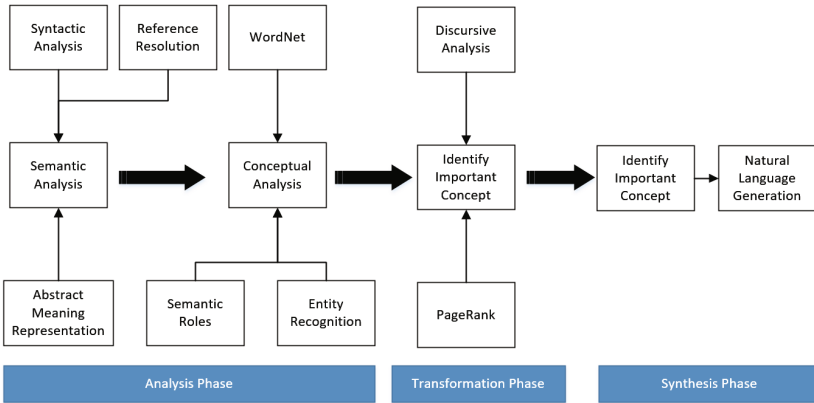

Fig. 1. Pipeline of the abstractive summarization method

3.1   Analysis Phase

This phase aimed at building the representation of a text as a graph. Given the abstractive approach, we had to change the original text using techniques of reference resolution to expand it and increase the amount of information in each sentence. This process helped the conceptual analysis.

In this work, we used the Natural Language Processing tools of theStanford[1], that includes also the syntactic tree generation, the part-of-speech tagging, delimitation of sentences and reference resolution among other tasks.

Because of the complexity of the reference resolution task, we only considered to exploit the references of pronouns to entities recognized by the annotations (*NN*, *NNS*, *NNP*, *NNPS*) into the *Part-of-Speech* Tagger. For example, in the

---

[1] Available at https://nlp.stanford.edu/software/. Accessed on February 2017.

following sentences we may see how this process increases the information contained when replacing the pronoun "it" with the full text of the organization that references:

"***The United Nations Food and Agriculture*** *organization said hot and dry conditions in January and February were expected to reduce the total cereal harvest in 11 southern African countries to 16 m tonnes, 25% down on the average. [* ***It (PRP)—The United Nations Food and Agriculture (NNP)*** *]said Zimbabwe and South Africa, which normally offset shortages in the area with their own surpluses, would themselves have to import food*"

After these steps, we generated the knowledge graph that represents the document. To do this, we used an Abstract Meaning Representation parser called CAMR parser [17]. This parser has taken part in SemEval-2016[2] reaching an average F1 of 66.5% over the corpus of the competition.

Once generated the AMR graph for each sentence in the document, we needed to join all sentences via some analysis to generate a knowledge graph.

In the same line of work as used in [14], a model was necessary to take the analysis to a higher level of abstraction, which we called "Conceptual", due we needed to abstract the concepts to merge them and generate new sentences.

Unlike the work presented in [14], where VerbNet [9] was used through manual work to align concepts and semantic relationships, we generated conceptual graphs automatically based on the AMR output (and its features) that is already aligned with a linguistic resource such as Propbank[3], a corpus annotated with information related to syntactic and semantics of verbs.

In order to generate the conceptual graph, we used some and procedures and criteria to merge terms or expressions into a concept, which are shown as below:

– Semantic Roles: In AMR, the relationships between concepts have identifiers like `Arg0...Arg5` which are associated with a semantic role such as agent, patience, among others. In our work we used the relationship that exists between Propbank and VerbNet to identify the semantic relationships and semantic roles of each concept. Thus, we may find that `Arg0` usually represents the "Agent" for a verb. However, in case of ambiguity or when no exists information in PropBank, we associated semantic roles's PropBank with semantic role's VerbNet by default. This association may be seen in Table 1.
– Fusion by entities: AMR has ability to recognize entities like *Person, Organization, Location*, among others, which contains several subtypes. For example, *organization* contains *company, government, military, criminal organization*, among others. In our work, we merged entities (with the same name) which are considered "Agent", "Patient", "Goal" and "Theme" in the Semantic Role Labeling avoiding fusion of graphs by verbs because this generates confusion and ambiguity in the graph.

---

[2] Available at http://alt.qcri.org/semeval2016/. Accessed in February 2017.
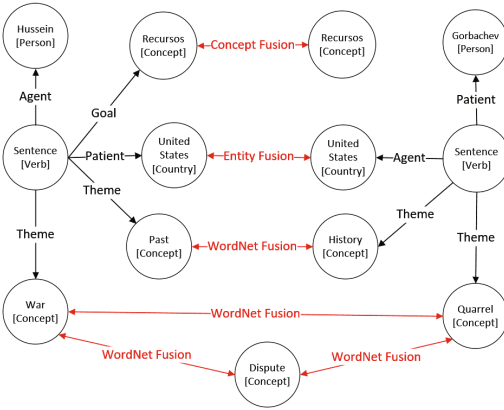[3] Proposition Bank Available at https://verbs.colorado.edu/mpalmer/projects/ace.html. Accessed on March 2017.

Fig. 2. Semantic graph fusion sample

Table 1. Default relation between AMR and semantic roles

AMR relation	Semantic role
Arg0	Agent
Arg1	Patient
Arg2	Goal
Arg3	Start
Arg4	End

– Fusion by WordNet concepts: Other criterion to merge terms/expressions into concepts was related to the measure got between two terms in the WordNet. To merge terms, we used the similarity measure Wu, proposed in [18]. In experiments, we determined that similarity measure must be greater than 0.9 to merge two terms into a concept.

Figure 2 shows an example of the fusion method using where we may appreciate that (1) some concepts that have been identified as *Agent* or *Patient* in different sentences, (2) entities recognized such as countries or persons, and (3) similar concepts in the WordNet (such as *Past* and *History*) may be merged.

3.2   Transformation Phase

In this phase, we needed to identify the most relevant concepts in the graph to create a summarization graph which includes them. To perform this, we executed the PageRank algorithm [1] over the conceptual graph. This algorithm is useful to identify relevant concepts considering the number of relations between different concepts and a possibility to do a random jump in a concept. In Eq. 1 we may see the formula where "M" represents the transition matrix (related to the number of relations), "v" represents the random jump vector, "c" represents a dumping factor and "Pr" represents the PageRank vector. In PageRank execution, the best results were obtained using a damping factor value of 0.65 and 30 iterations.

$$Pr = cMPr + (1 - c)v \tag{1}$$

Once the PageRank was executed, we perceived that some nodes with many relations received higher weights (although the related nodes were less important) generating noise. To solve this problem, we incorporated discourse-level information into the PageRank, since this information has proven to be useful

in extractive automatic summarization task [15]. We decided for this algorithm because use the nuclear-satellite information and take in consideration the relation type between the EDUs[4] to assign importance.

Thus, we applied a method proposed in [15] (called O'Donell method), which calculates the importance of each EDU according to the relations found in the Rhetorical Structure Theory [13]. The results got from O'Donell method were incorporated into the random jump vector in the PageRank algorithm. This made that concepts with in high score in O'Donell method benefit others around itself and unimportant concepts with many relations have low scores.

3.3 Synthesis Phase

Once the concepts in the graph have been weighted, the model iterated the conceptual graph to extract information about the actions done (*Verbs*), who has made those actions (*Agents*), who is affected by them (*Patients*), what is the theme (*Themes*) and what is the aim (*Goals*).

Then, our algorithm started in the verb nodes and as from there attempted to extract the nodes attached to it with the semantic relationship of Agent and thus for the semantic roles of *Patient*, *Theme* and *Goal*. Once these subgraphs were identified, it was the basis of a new sentence whose importance was given by:

$$Sentence\_Relevance = Sum(P(Agents) + P(Verbs) + P(Themes) + P(Goals))$$

These total values represented the final relevance of the expression. Then, we applied a descendant sorting over the sentence relevance to generate a summary with the most important expressions until up to a compression rate. To generate a sentence that has the synthesis of the document ideas a similar form as human production, we used `SimpleNLG` [5][5] as a tool for Natural Language Generation.

4 Evaluation

To evaluate the use of Abstract Meaning Representation and Discourse-level information into Automatic Summarization, we conducted experiments for each case. All experiments were performed on *Document Understanding Conference* (DUC) corpus[6].

Table 2 shows results of each experiment in the training corpus and test corpus, i.e., when only used the expanded conceptual graph with the reference resolution (conceptual + RR), when used a conceptual graph with reference resolution and includes discursive information (RST) (Conceptual + RR + RST) and when used a conceptual graph with reference resolution, includes discursive information (RST) and Natural Language Generation (Conceptual + RR + RST + NLG). Furthermore, the results improved in each experiment, i.e., Conceptual + RR + RST

---

[4] Elementary Discourse Unit is the basic unit in discourse-level.

[5] https://github.com/simplenlg/simplenlg. Last visited in February 2017.

[6] http://duc.nist.gov/data.html. Last visited in February 2017.

model was better than Conceptual + RR model and Conceptual + RR + RST + NLG model was better than Conceptual + RR + RST model.

We noted in our experiments with NLG that best combination was the use of *"with objective of"* like connector when we detected the *goal* semantic relation. For example, the sentence *"We agree possible international peaceful order devour large state and Gorbachev neighbor"* was transformed in *"We agreed with objective of possible international peaceful order devour large state and Gorbachev neighbor"*.

Another point to note is generated sentences had a correct use of the pronoun *We*, also we can identify the verb and expression goal. Table 2 shows a significant improvement in Rouge-1 and Rouge-L metrics and an important enhancement in Rouge-SU4 metric. This means a much better coherence in the generated text. In particular, the use of connectors like *And*and the correctness in the person and number over the generated expression improve the result.

In relation to incorporating discursive information into the original method, we may note an increment between conceptual and conceptual with discursive information. It based on the myopia of pure conceptual model to include additional concepts (*Agents, Patients, Goal, Themes*), because it only uses semantic relations at sentence level. For example, in a specific document the application of Conceptual + RR model produces 6 sentences, where four of them talk about the same subject. The discursive information incremented the possibility to detect expressions that can produce more valuable sentences to the summary. For example, in a same document, when applying the conceptual + RR + RST, it got 9 sentences where only two of them mentioned the same subject and the F1 Rouge-1 score was increment by 15%.

Table 2. F1 metric between Conceptual, Conceptual-RST and Conceptual-RST-NLG

Corpus	Training corpus			Test corpus		
F1/Rouge	R-1	R-L	R-SU4	R-1	R-L	R-SU4
Conceptual + RR	0.199	0.187	0.024	0.224	0.211	0.029
Conceptual + RR + RST	0.212	0.200	0.027	0.228	0.217	0.029
Conceptual + RR + RST + NLG	**0.230**	**0.216**	**0.031**	**0.244**	**0.231**	**0.033**

Finally, we may highlight that our experiments used none algorithm that may present an over-fitting to the specific data, so the goodness of the method only depends on the text in a document. Also, is important to note, the model never show a negative effect. However, the increment of the performance was not statistically significant according to the Wilcoxon Test.

5   Conclusions and Future Work

This work presented an automatic abstractive summarization using semantic representations and discourse-level information. The analysis phase used

information from semantic analysis, got from use AMR parser for each sentence in a document. Then, we generated a conceptual graph by merging concepts with help of WordNet and Semantic Roles got from AMR. During the transformation phase, Discourse-level information was incorporated into PageRank algorithm to identify the most important concepts, resulting in an improvement on the concept identification.

In the synthesis phase, we implemented a navigation method to generate expressions from the ranked conceptual graph using hand-crafted rules based on semantic roles. With these rules we extracted many expressions that have a final score equal to the amount of their parts. After that, we sorted these expressions based on the amount score and take the most valuables for the natural language generation task, in our experiment we have worked with a compression rate of 20% more and less 100 words.

At last, the got expressions were used with SimpleNLG to generate a much natural expressions. In this work, we configured the tool to generate the sentence in a past form to get a coherent expression in tempo and number. The proposed method was evaluated on Document Understanding Conference (DUC) 2002 Corpus showing a F1 score of 24% on the Rouge-1 metric and outperformed the other variations of our method.

One limit related with the abstraction model is related with AMR. Although AMR is an important player in Semantic Analysis, in its current form is not enough to support the discovery and manipulation of the principal concepts, because it is too influenced by the syntax. We found evidence that different representations of the same idea are got, depending if these are written in active or passive voice.

One future work is related to the way of navigation or iteration over the Ranked Conceptual Graph with score information on its nodes to generate the candidate sentences. As a future work, we would like to explore other ways to navigate this graph to improve selection of concepts and generate better sentences.

References

1. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. **30**(1), 107–117 (1998)
2. Carenini, G., Ng, R., Pauls, A.: Multi-document summarization of evaluative text. In: Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (2006)
3. Dang, H.T., Kipper, K., Palmer, M.: Integrating compositional semantics into a verb lexicon. In: Proceedings of the 18th conference on Computational linguistics, vol. 2, pp. 1011–1015. Association for Computational Linguistics (2000)
4. Flanigan, J., Thomson, S., Carbonell, J., Dyer, C., Smith, N.A.: A Discriminative Graph-Based Parser for the Abstract Meaning Representation, pp. 1426–1436. ACL (2014)
5. Gatt, A., Reiter, E.: SimpleNLG: a realisation engine for practical applications. In: Proceedings of the 12th European Workshop on Natural Language Generation, pp. 90–93. Association for Computational Linguistics (2009)

6. Gerani, S., Mehdad, Y., Carenini, G., Ng, R.T., Nejat, B.: Abstractive summarization of product reviews using discourse structure. In: EMNLP, pp. 1602–1613 (2014)
7. Montes-y-Gómez, M., Gelbukh, A., López-López, A., Baeza-Yates, R.: Flexible comparison of conceptual graphs. In: Mayr, H.C., Lazansky, J., Quirchmayr, G., Vogel, P. (eds.) DEXA 2001. LNCS, vol. 2113, pp. 102–111. Springer, Heidelberg (2001). doi:10.1007/3-540-44759-8_12
8. Kilgarriff, A., Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge (2000)
9. Kipper, K., Dang, H.T., Palmer, M., et al.: Class-based construction of a verb lexicon. In: AAAI/IAAI, pp. 691–696 (2000)
10. Knight, K., Baranescu, L., Bonial, C., Georgescu, M., Griffitt, K., Hermjakob, U., Marcu, D., Palmer, M., Schneifer, N.: Abstract Meaning Representation (AMR) Annotation Release 1.0. Web download (2014)
11. Liu, F., Flanigan, J., Thomson, S., Sadeh, N., Smith, N.A.: Toward Abstractive Summarization Using Semantic Representations (2015)
12. Mani, I.: Automatic Summarization, vol. 3. John Benjamins Publishing, Amsterdam (2001)
13. Mann, W.C., Thompson, S.A.: Rhetorical structure theory: toward a functional theory of text organization. Text-Interdiscipl. J. Study Discourse **8**(3), 243–281 (1988)
14. Miranda-Jiménez, S., Gelbukh, A., Sidorov, G.: Conceptual graphs as framework for summarizing short texts. Int. J. Concept. Struct. Smart Appl. (IJCSSA) **2**(2), 55–75 (2014)
15. O Donnell, M.: Variable-length on-line document generation. In: the Proceedings of the 6th European Workshop on Natural Language Generation, Gerhard-Mercator University, Duisburg, Germany (1997)
16. Palmer, M., Gildea, D., Kingsbury, P.: The proposition bank: an annotated corpus of semantic roles. Comput. Linguist. **31**(1), 71–106 (2005)
17. Wang, C., Pradhan, S., Pan, X., Ji, H., Xue, N.: CAMR at SemEval-2016 task 8: an extended transition-based AMR parser. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 1173–1178. Association for Computational Linguistics, San Diego (June 2016)
18. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics, pp. 133–138. Association for Computational Linguistics (1994)