A Similarity-based Abstract Argumentation Approach to Extractive Text Summarization

Stefano Ferilli, Andrea Pazienza, Sergio Angelastro and Alessandro Suglia*

University of Bari Aldo Moro, Bari, Italy

*Interaction Lab, Heriot-Watt University, Edinburgh Centre for Robotics

What is Automatic Text Summarization?

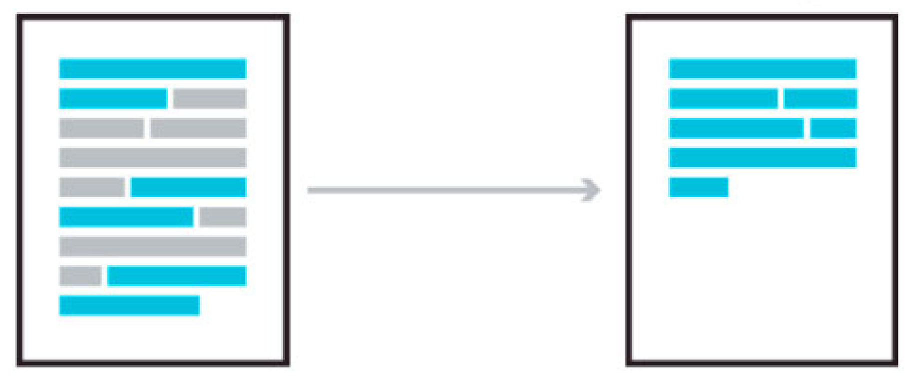
Text summarization is the process of automatically creating a shorter version of one or more text documents.

Traditional techniques can be classified in:

- *Extractive:* work by selecting sentences from the input document(s) according to some criterion [1]
- *Abstractive:* may produce summaries containing sentences that were not present in the input document(s) [2]

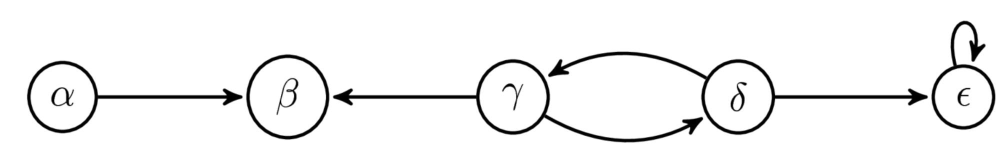
Why Automatic Text Summarization?

- In the era of Internet 2.0 every Internet user suffer from the *Information overload problem*
- Automatic Text Summarization can help users to analyze huge number of documents in an efficient way



Abstract Argumentation Theory

- *Abstract Argumentation* is a non-monotonic inferential strategy that aims at selecting reliable items in a set of conflicting claims
- *Abstract Argumentation Theory* [6] works on *Argumentation Frameworks* (AF), i.e. graphs composed by *arguments* (nodes) and *attacks* (edges) which are able to understand which subsets of arguments in the graph are mutually compatible



Abstract Argumentation approach

NLP pipeline → Weighted Graph Building → Semantics Evaluation → Solutions scoring strategy

NLP pipeline

1. Sentence splitting and tokenization
2. Lemmatization
3. Stopwords removal
4. Embedding-based *Simplified Lesk Word sense disambiguation* [22]

Weighted Graph Building

1. Words similarity computation evaluated as a linear combination ϕ of three similarity functions:
   1. Syntactic similarity sim_{syn}
   2. Sematic similarity sim_{sem}
   3. Word embedding similarity sim_{emb}
2. Arguments relationships weights evaluated as:

$$sim(s_i, s_j) = \frac{1}{|U(\widetilde{s_i}, \widetilde{s_j})|} \sum_{w_p \in U(\widetilde{s_i}, \widetilde{s_j})|} \max_{w_q \in U(\widetilde{s_i}, \widetilde{s_j}) \setminus w_p} \phi(w_p, w_q)$$

Semantics Evaluation

Inspired by the notion of *Inconsistency Budget* of *Weighted AF* [7], we derived a *Bipolar AF,* which includes also support relations*:*

1. Attack relations are retained if the attack weight is above the *attack threshold* α
2. Support relations are retained if the support weight is below the *support threshold* β

We exploit different semantics to evaluate the acceptability of the arguments such as: *d-/s-admissible, d-/s-preferred* and *conflict-free sets*

Solution scoring function

1. Iterative argument ranking procedure inspired to *Maximal Marginal Relevance* [2] applied on the evaluated solutions set
2. Authoritative arguments scoring function:

$$auth(a) = [1 + \log(\sup(a))] * \left[\log\left(\frac{|R_{att}|}{att(a) + 1}\right)\right]$$

3. Two alternative implementations:
   1. **AUTH-MMR:** selects the best subset of arguments from the semantic solutions
   2. **AUTH:** selects the solution which approximates the target length and which maximizes the intra-cohesion coefficient of the arguments

Results

Evaluation conducted on the single-document text summarization task of the English version of the *MultiLing 2015* dataset [11].

We are convinced that summarization is not appropriately evaluated because:

1. No exact match between the sentences in the input text and those in the summary
2. Summaries may contain words that are not contained in the input text
3. The summarization process is incredibly subjective and it is possible that multiple summaries are appropriate for the given input text despite only one of them is provided in the test data

We defined a compound indicator to assess the goodness of the generated summaries:

$$Quality(s) = \frac{ROUGE - 1(s)}{length(s)}$$

Techniques based on Argumentation frameworks only

Step	Semantics	β	α	length (%)	Quality	Rouge-1		Rouge-2	
						Recall	Precision	Recall	Precision
2	s/d-admissible	0.1	0.4	4365 (17%)	1.13E-04	49.28%	30.32%	15.49%	7.22%
3	stable, complete	0.1	0.5	8544 (33%)	7.07E-05	60.44%	24.44%	23.98%	7.43%
4	s/d-admissible	0.1	0.5	9826 (38%)	7.33E-05	72.09%	26.65%	27.26%	6.16%

Approaches based on re-ranking procedures

Semantics	β	α	length	ROUGE-1	ROUGE-2
conflict-free	0.65	0.80	1966	46.30%	10.58%
s/d-preferred	0.25	0.75	1942	44.12%	10.54%

References

[1]: Banerjee, S., Mitra, P., Sugiyama, K.: Multi-document abstractive summarization using ILP based multi-sentence compression. In: IJCAI'15 (2015)

[2]: Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for re-ordering documents and producing summaries. In: ACM SIGIR. pp. 335–336. ACM (1998)

[6]: Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artificial intelligence 77(2), 321–357 (1995)

[7]: Dunne, P.E., et al.: Weighted argument systems: Basic definitions, algorithms, and complexity results. Artificial Intelligence 175(2), 457 – 486 (2011)

[11]: Giannakopoulos, G., et al.: Multiling 2015: multilingual summarization of single and multi-documents, on-line fora, and call-center conversations. SIGDIAL pp. 270–274 (2015)

[14]: Lloret, E., Palomar, M.: Text summarisation in progress: a literature review. Artificial Intelligence Review 37(1), 1–41 (2012)

[22]: Vasilescu, F., Langlais, P., Lapalme, G.: Evaluating variants of the Lesk approach for disambiguating words. In: Lrec (2004)