# Unsupervised Abstractive Meeting Summarization with Multi-Sentence Compression and Budgeted Submodular Maximization
## *Supplementary Material*

**Guokan Shang**[1,2], **Wensi Ding**[1*], **Zekun Zhang**[1*], **Antoine J.-P. Tixier**[1],
**Polykarpos Meladianos**[1,3], **Michalis Vazirgiannis**[1,3], **Jean-Pierre Lorré**[2]

[1]École Polytechnique, [2]Linagora, [3]AUEB

## Appendices

## A    Use of WordNet

When the word to be mapped to the MSCG is a **non-stopword**, and if there is no node in the graph that has the same lowercased form and the same part-of-speech tag, we try to perform the mapping by using WordNet in the following order:

(i) there is a node which is a synonym of the word (e.g., "price" and "costs"). The word is mapped to that node, and the node is relabeled with the word if the latter has a higher TW-IDF score.

(ii) there is a node which is a hypernym of the word (e.g., "diamond" and "gemstone"). The word is mapped to that node, and the node is relabeled with the word if the latter has a higher TW-IDF score.

(iii) there is a node which shares a common hypernym with the word (e.g., "red","blue" → "color"). If the product of the WordNet path distance similarities of the common hypernym with the node and the word exceeds a certain threshold, the word is mapped to that node and the node is relabeled with the hypernym. A completely new word might thus be introduced. We set its TW-IDF score as the highest TW-IDF of the two words it replaces. When multiple nodes are eligible for mapping, we select the one with greatest path distance similarity product.

(iv) there is a node which is in an entailment relation with the word (e.g., "look" is entailed by "see"). The word is mapped to that node, and the node is relabeled with the word if the latter has a higher TW-IDF score.

In attempts **i**, **ii**, and **iv** above, if there is more than one candidate node, we select the one with highest TW-IDF score. If all attempts above are unsuccessful, a new node is created for the word.

## B    Baseline Details

- **Random**. A basic baseline recommended by (Riedhammer et al., 2008) to ease cross-study comparison. This system randomly selects utterances without replacement from the transcription until the budget is violated. To account for stochasticity, we report scores averaged over 30 runs.

- **Longest Greedy**. A basic baseline recommended by (Riedhammer et al., 2008) to ease cross-study comparison. The longest remaining utterance is selected at each step from the transcription until the summary size constraint is satisfied.

- **TextRank** (Mihalcea and Tarau, 2004). Utterances within the transcription are represented as nodes in an undirected complete graph, and edge weights are assigned based on lexical similarity between utterances. To provide a summary, the top nodes according to the weighted PageRank algorithm (Page et al., 1999) are selected. We used a publicly available implementation[1].

---

*Work done as part of 3rd year project, with equal contribution.

[1] https://github.com/summanlp/textrank

- **ClusterRank** (Garg et al., 2009). This system is an extension of TextRank to meeting summarization. Firstly, utterances are segmented into clusters. A complete graph is built from the clusters. Then, a score is assigned to each utterance based on both the PageRank score of the cluster it belongs to and its cosine similarity with the cluster centroid. In the end, a greedy selection strategy is applied to build the summary out of the highest scoring utterances. Since the authors did not make their code publicly available and were not able to share it privately, we wrote our own implementation.

- **CoreRank submodular & PageRank submodular** (Tixier et al., 2017). These two *extractive* baselines implement the last step of our pipeline (see Section 4.4). That is, budgeted submodular maximization is applied directly on the full list of utterances. As can be inferred from their names, the only difference between those two baselines is that the first uses PageRank scores, whereas the second uses CoreRank scores.

- **Oracle**. This system is the same as the Random baseline, but instead of sampling utterances from the ASR transcription, it draws from the human extractive summaries as input. Annotators put those summaries together by selecting the best utterances from the entire manual transcription. Scores were averaged over 30 runs due to the randomness of the procedure.

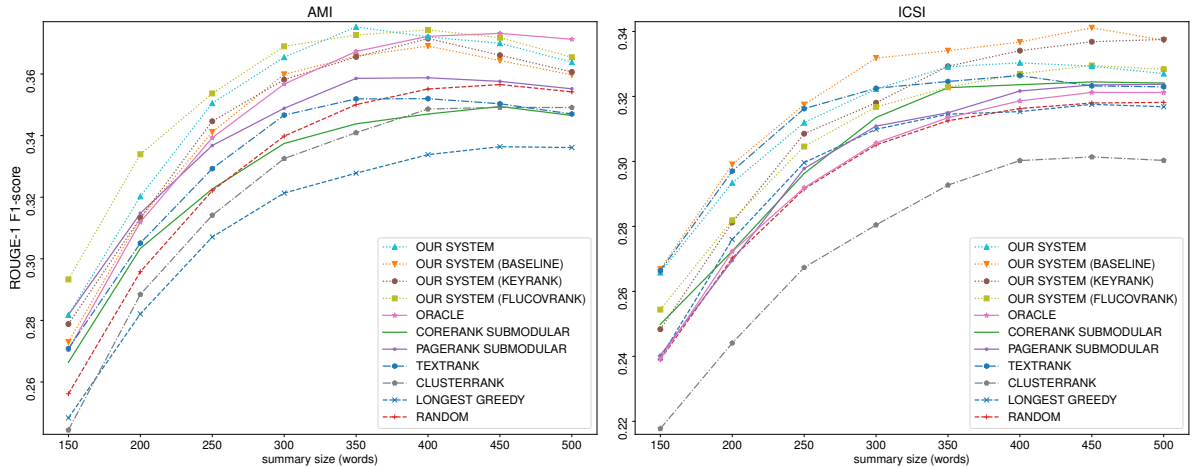## C   Results for Manual Transcriptions



Figure 1: ROUGE-1 F-1 scores for various budgets (manual transcriptions).

| | AMI ROUGE-1 | | | AMI ROUGE-2 | | | AMI ROUGE-SU4 | | | ICSI ROUGE-1 | | | ICSI ROUGE-2 | | | ICSI ROUGE-SU4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | F-1 | R | P | F-1 | R | P | F-1 | R | P | F-1 | R | P | F-1 | R | P | F-1 |
| Our System | 42.03 | 34.77 | **37.53** | 8.87 | 7.56 | 8.06 | 15.92 | 14.08 | **14.76** | 38.57 | 29.30 | 32.93 | 5.80 | 4.74 | 5.14 | 13.92 | 10.79 | 12.04 |
| Our System (Baseline) | 40.88 | 33.96 | 36.58 | 8.13 | 6.95 | 7.39 | 15.17 | 13.25 | 13.97 | 40.03 | 30.20 | **34.11** | 6.65 | 5.51 | **5.98** | 14.65 | 11.37 | **12.70** |
| Our System (KeyRank) | 40.87 | 33.91 | 36.56 | 8.42 | 7.12 | 7.62 | 15.50 | 13.48 | 14.25 | 39.55 | 29.79 | 33.68 | 6.32 | 5.19 | 5.64 | 14.63 | 10.99 | 12.47 |
| Our System (FluCovRank) | 41.73 | 34.50 | 37.27 | 8.45 | 7.05 | 7.60 | 16.08 | 13.47 | 14.49 | 38.57 | 29.21 | 32.95 | 6.38 | 5.08 | 5.60 | 14.38 | 10.62 | 12.13 |
| Oracle | 40.49 | 34.65 | **36.73** | 8.07 | 7.35 | **7.55** | 15.00 | 14.03 | **14.26** | 37.91 | 28.39 | **32.12** | 5.73 | 4.82 | **5.18** | 13.35 | 10.73 | **11.80** |
| CoreRank Submodular | 38.95 | 31.49 | 34.38 | 7.85 | 6.81 | 7.20 | 14.08 | 13.55 | 13.61 | 37.31 | 29.51 | 32.45 | 5.59 | 5.05 | 5.24 | 13.19 | 11.08 | 11.87 |
| PageRank Submodular | 40.58 | 32.87 | 35.86 | 9.20 | 7.77 | **8.32** | 15.59 | 14.14 | 14.64 | 37.72 | 28.86 | 32.35 | 6.35 | 5.46 | 5.82 | 13.35 | 11.60 | 12.30 |
| TextRank | 39.47 | 32.57 | 35.19 | 7.74 | 6.62 | 7.05 | 14.80 | 13.03 | 13.69 | 37.60 | 28.79 | 32.32 | 6.63 | 5.53 | **5.98** | 14.18 | 11.18 | 12.41 |
| ClusterRank | 38.32 | 31.51 | 34.10 | 6.93 | 5.95 | 6.31 | 13.69 | 12.40 | 12.84 | 35.66 | 26.58 | 30.14 | 4.53 | 3.99 | 4.21 | 12.10 | 9.71 | 10.69 |
| Longest Greedy | 36.73 | 30.39 | 32.78 | 5.52 | 4.58 | 4.93 | 13.52 | 10.91 | 11.93 | 37.15 | 28.21 | 31.76 | 5.50 | 4.60 | 4.98 | 13.59 | 10.03 | 11.46 |
| Random | 39.29 | 32.38 | 35.01 | 7.14 | 6.16 | 6.52 | 14.16 | 12.95 | 13.35 | 37.48 | 28.10 | 31.80 | 5.41 | 4.65 | 4.95 | 12.97 | 10.67 | 11.61 |

Table 1: Macro-averaged results for 350 and 450 word summaries (manual transcriptions).

## D   Example Summaries

Examples were generated from the manual transcriptions of meeting AMI TS3003c. Note that our system can also be interactively tested at http://datascience.open-paas.org/abs_summ_app.

**Reference Summary (254 words)**

The project manager opened the meeting and recapped the decisions made in the previous meeting.
The marketing expert discussed his personal preferences for the design of the remote and presented the results of trend-watching reports, which indicated that there is a need for products which are fancy, innovative, easy to use, in dark colors, in recognizable shapes, and in a familiar material like wood.
The user interface designer discussed the option to include speech recognition and which functions to include on the remote.
The industrial designer discussed which options he preferred for the remote in terms of energy sources, casing, case supplements, buttons, and chips.
The team then discussed and made decisions regarding energy sources, speech recognition, LCD screens, chips, case materials and colors, case shape and orientation, and button orientation.
The team members will look at the corporate website.
The user interface designer will continue with what he has been working on.
The industrial designer and user interface designer will work together.
The remote will have a docking station.
The remote will use a conventional battery and a docking station which recharges the battery.
The remote will use an advanced chip.
The remote will have changeable case covers.
The case covers will be available in wood or plastic.
The case will be single curved.
Whether to use kinetic energy or a conventional battery with a docking station which recharges the remote.
Whether to implement an LCD screen on the remote.
Choosing between an LCD screen or speech recognition.
Using wood for the case.

**Our System (250 words)**

attract elderly people can use the remote control
changing channels button on the right side that would certainly yield great options for the design of the remote
personally i dont think that older people like to shake your remote control
imagine that the remote control and the docking station
remote control have to lay in your hand and right hand users
finding an attractive way to control the remote control
casing the manufacturing department can deliver a flat casing single or double curved casing
top of that the lcd screen would help in making the remote control easier
increase the price for which were selling our remote control
remote controls are using a onoff button still on the top
apply remote control on which you can apply different case covers
button on your docking station which you can push and then it starts beeping
surveys have indicated that especially wood is the material for older people
mobile phones so like the nokia mobile phones when you can change the case
greyblack colour for people prefer dark colours
brings us to the discussion about our concepts
docking station and small screen would be our main points of interest
industrial designer and user interface designer are going to work
innovativeness was about half of half as important as the fancy design
efficient and cheaper to put it in the docking station
case supplement and the buttons it really depends on the designer
start by choosing a case
deployed some trendwatchers to milan

**Our System (Baseline) (250 words)**

apply remote controls on which you can apply different case for his remote control
changing channels and changing volume button on both sides that would certainly yield great options for the design of the remote
personally i dont think that older people like to shake their remote control
finding an attractive way to control the remote control the i found some something about speech recognition
imagine that the remote control and the docking station should be telephoneshaped
casing the manufacturing department can deliver a flat casing single or double curved casing
remote control have to lay in your hand and right hand users
remote controls are using a onoff button over in this corner
woodlike for the more exclusive people can use the remote control
heard our industrial designer talk about flat single curved and double curved
innovativeness this means functions which are not featured in other remote control
button on your docking station which you can push and then it starts beeping
greyblack colour for people prefer dark colours
docking station and small screen would be our main points of interest

special button for subtitles for people which c f who cant read small subtitles
pretty big influence on production price and image unless we would start two product lines
surveys have indicated that especially wood is the material for older people
mobile phones so like the nokia mobile phones when you can change the case
case the supplement and the buttons it really depends on the designer
buttons

### Our System (KeyRank) (250 words)

changing case covers
prefer a design where the remote control and the docking station
greyblack colour for people prefer dark colours
remote controls are using a onoff button over in this corner
requirements are teletext docking station and small screen with some extras that button information
apply remote controls on which you can apply different case covers
woodlike for the more exclusive people can use the remote control
casing the manufacturing department can deliver a flat casing single or double curved casing
remote control have to lay in your hand and right hand users
asked if w they would if people would pay more for speech recognition function would not make the remote control
start by choosing a case
innovativeness this means functions which are not featured in other remote controls
top of that the lcd screen would help in making the remote control easier
changing channels and changing volume button on both sides that would certainly yield great options for the design of the remote
personally i dont think that older remotes are flat board smartboard
button on your docking station which you can push and then it starts beeping
case supplement and the buttons it really depends on the designer
surveys have indicated that especially wood is the material for older people will recognise the button
speak speech recognition and a special button for subtitles for people which c f who cant read small subtitles
innovativeness was about half as important as the fancy design
pretty big influence

### Our System (FluCovRank) (250 words)

elderly people can use the remote control
remote controls are using a onoff button still on the top
general idea of the concepts and the material for older people like to shake your remote control
docking station and small screen would be our main points of interest
industrial designer and user interface designer are going to work
casing the manufacturing department can deliver single curved
changing channels and changing volume button on both side that would certainly yield great options for the design of the remote
button on your docking station which you can push and then it starts beeping
imagine that the remote control will be standing up straight in the docking station will help them give the remote
asked if w they would if people would pay more for speech recognition in a remote control you can call it and it gives an sig signal
research about bi large lcd sh display for for displaying the the functions of the buttons
case the supplement and the buttons it really depends on the designer
pointed out earlier that a lot of remotes rsi
innovativeness was about half of half as important as the fancy design
push on the button for subtitles for people which c f who cant read small subtitles
efficient and cheaper to put it in the docking station could be one of the marketing issues
difficult to handle and to get in the right shape to older people
talk about the energy source is rather fancy

## References

Nikhil Garg, Benoit Favre, Korbinian Reidhammer, and Dilek Hakkani-Tür. 2009. Clusterrank: a graph based method for meeting summarization. In *Tenth Annual Conference of the International Speech Communication Association*.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. http://aclweb.org/anthology/W04-3252.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Korbinian Riedhammer, Dan Gillick, Benoit Favre, and Dilek Hakkani-Tür. 2008. Packing the meeting summarization knapsack. In *Ninth Annual Conference of the International Speech Communication Association*.

Antoine Tixier, Polykarpos Meladianos, and Michalis Vazirgiannis. 2017. Combining graph degeneracy and submodularity for unsupervised extractive summarization. In *Proceedings of the Workshop on New Frontiers in Summarization*. Association for Computational Linguistics, pages 48–58. http://aclweb.org/anthology/W17-4507.