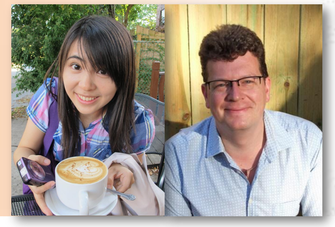
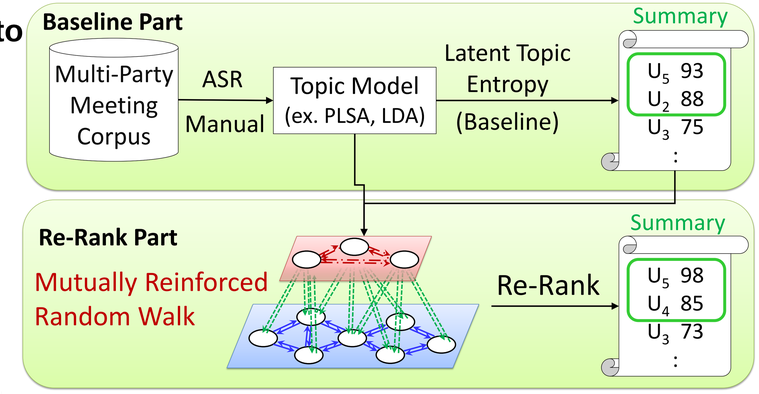
Two-Layer Mutually Reinforced Random Walk for Improved Multi-Party Meeting Summarization

Yun-Nung (Vivian) Chen and Florian Metze



1. Summary

- **Idea:**
  - **Important utterances are topically similar to each other**
  - **Utterances similar to important speakers should be more important**
- **Approach for extractive summary**
  - **Construct a two-layer graph to represent**
  1) **the utterance nodes in utterance-layer**
  2) **the speaker nodes in speaker-layer**
  - **Mutually propagate importance scores via within-layer edges and between-layer edges**



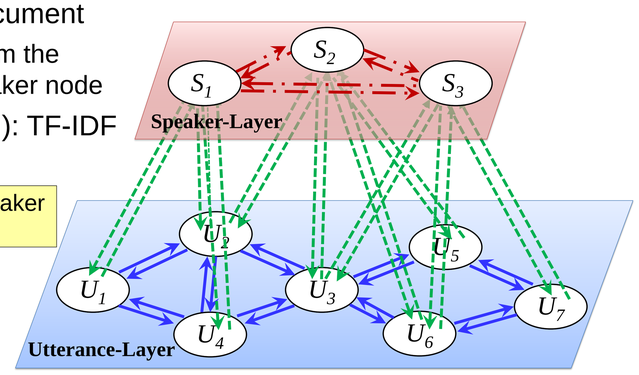
Flowchart of proposed approach

2. Graph Construction

- Speaker-Layer
  - Node: speakers in a document combine all utterances from the same speaker as the speaker node
  - Edge weight (red, green): TF-IDF cosine similarity

➢ The utterances from the similar speaker can partially share the importance

- Utterance-Layer
  - Node: utterances in a document
  - Edge weight (blue): topical/lexical similarity



➢ The utterances topically similar to more important utterances should be more important

3. Two-Layer Mutually Reinforced Random Walk

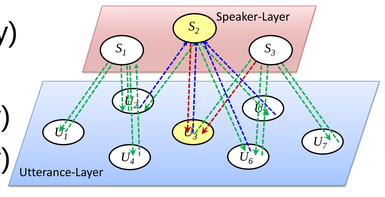
- **Basic Idea: high importance means**
  - **① Utterances with higher original score**
  - **② Utterances topically/lexically similar to the indicative utterances**
  - **③ Utterances similar important speakers' utterances**
- Similarity Matrix

L_{UU}: utterance-to-utterance relation (topical/lexical similarity)
L_{SS}: speaker-to-speaker relation (TF-IDF cosine similarity)
L_{US}: utterance-to-speaker relation (TF-IDF cosine similarity)
L_{SU}: speaker-to-utterance relation (TF-IDF cosine similarity)

- Two-Layer MRRW-BP (Between-Layer Propagation)

$$\begin{cases} F_U^{(t+1)} = (1-\alpha)F_U^{(0)} + \alpha \cdot L_{US}F_S^{(t)} \\ F_S^{(t+1)} = (1-\alpha)F_S^{(0)} + \alpha \cdot L_{SU}F_U^{(t)} \end{cases}$$

$F_U^{(t+1)}$: utterance score at (t+1)-th iteration
$F_S^{(t+1)}$: speaker score at (t+1)-th iteration
$F_U^{(0)}$: original importance of utterance
$F_S^{(0)}$: equal weight
$L_{US}F_S^{(t)}$: scores propagated from speaker-layer
$L_{SU}F_U^{(t)}$: scores propagated from utterance-layer

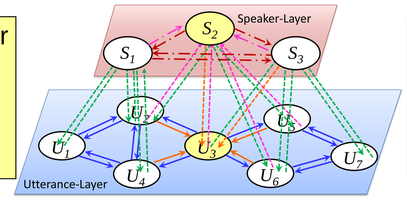


➢ Utterance node U can get higher score when
① Higher original importance
② More speaker nodes similar to utterance U

- Two-Layer MRRW-WBP (Within- and Between-Layer Propagation)

$$\begin{cases} F_U^{(t+1)} = (1-\alpha)F_U^{(0)} + \alpha \cdot L_{UU}^T L_{US}F_S^{(t)} \\ F_S^{(t+1)} = (1-\alpha)F_S^{(0)} + \alpha \cdot L_{SS}^T L_{SU}F_U^{(t)} \end{cases}$$

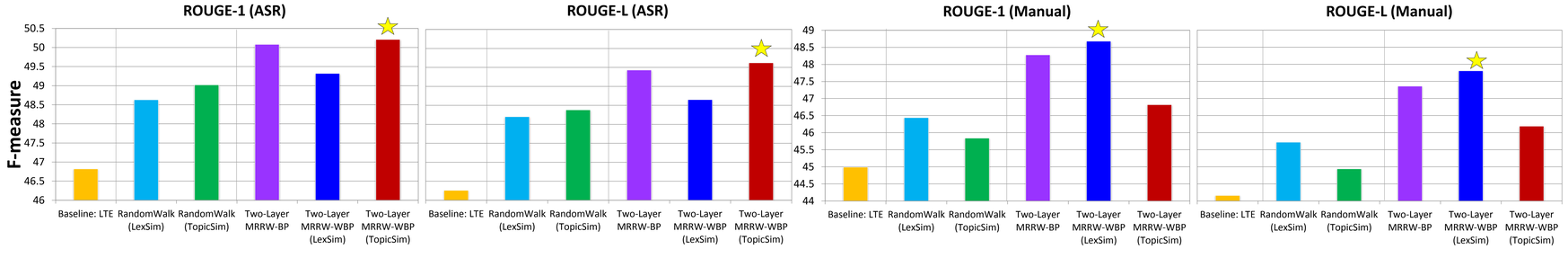
$L_{UU}^T L_{US}F_S^{(t)}$: scores propagated from speaker-layer then propagated within utterance-layer
$L_{SS}^T L_{SU}F_U^{(t)}$: scores propagated from utterance-layer then propagated within speaker-layer



➢ Utterance node U can get higher score when
① Higher original importance
② More speaker nodes similar to utterance U
③ More important utterances similar to utterance U

5. Experiments

- Dataset: 10 meetings from CMU Speech Group, #Speaker: 6 (total), 2-4 (each), WER = 44%
- Parameter setting: α = 0.9, summary ratio = 30%



6. Conclusions

- Graph-based approaches can improve speech summarization performance
- Two-layer approaches involving speaker information can get further improvement
- Topical similarity is more robust to recognition errors
  → better for ASR transcripts
- Lexical similarity is more accurate when absence of errors
  → better for manual transcripts
- Our proposed approaches achieve more than 7% relative improvement compared to the baseline