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.

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

3. Two-Layer Mutually Reinforced Random Walk

- **Basic Idea:** high importance means
  - Utterances with higher original score
  - Utterances topically/lexically similar to the indicative utterances
  - 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)**
  - Utterance node \( U \) can get higher score when
    1) Higher original importance
    2) More speaker nodes similar to utterance \( U \)
  - Scores propagated from speaker-layer then propagated within utterance-layer.

- **Two-Layer MRRW-WBP (Within- and Between-Layer Propagation)**
  - 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 \( U \)
  - Scores propagated from utterance-layer then propagated within speaker-layer.

4. Experiments

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

5. 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.