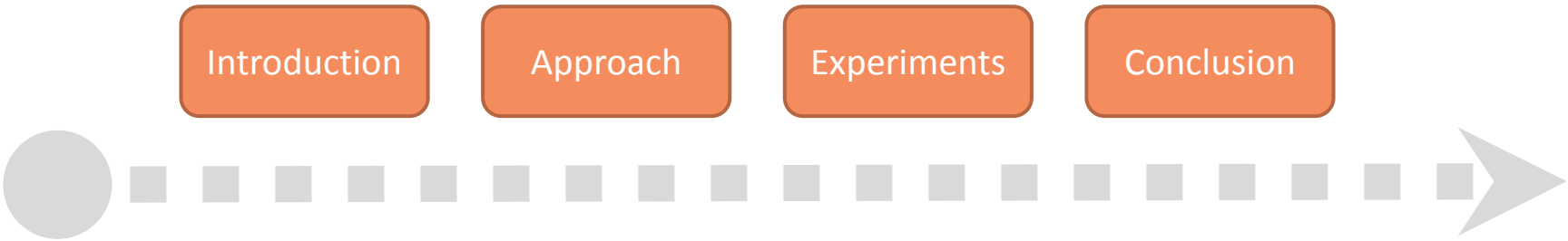
Multi-Layer Mutually Reinforced Random Walk with Hidden Parameters for Improved Multi-Party Meeting Summarization

Yun-Nung (Vivian) Chen and Florian Metze
Language Technologies Institute
School of Computer Science
Carnegie Mellon University

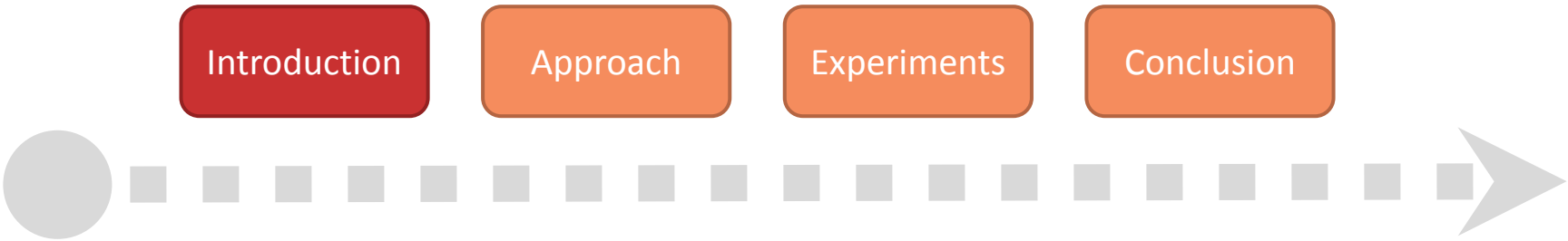
{yvchen, fmetze}@cs.cmu.edu

The 14th Annual Conference of the International Speech Communication Association – Aug. 25-29, 2013

Outline

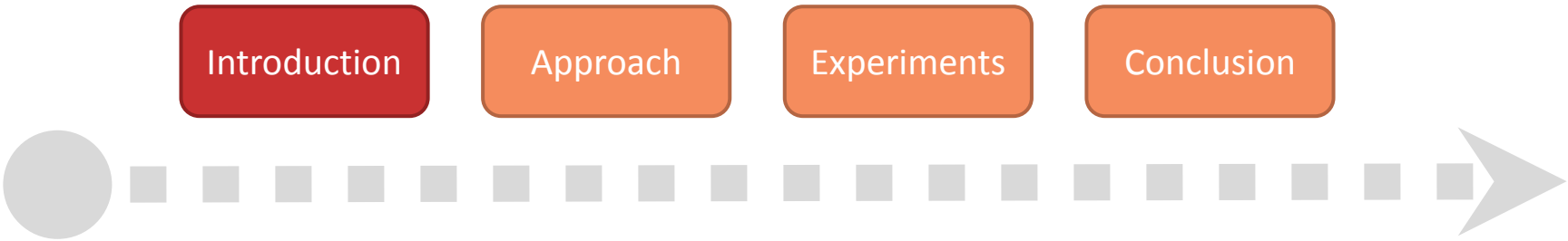


Outline



- Motivation
- Extractive Summarization

Outline



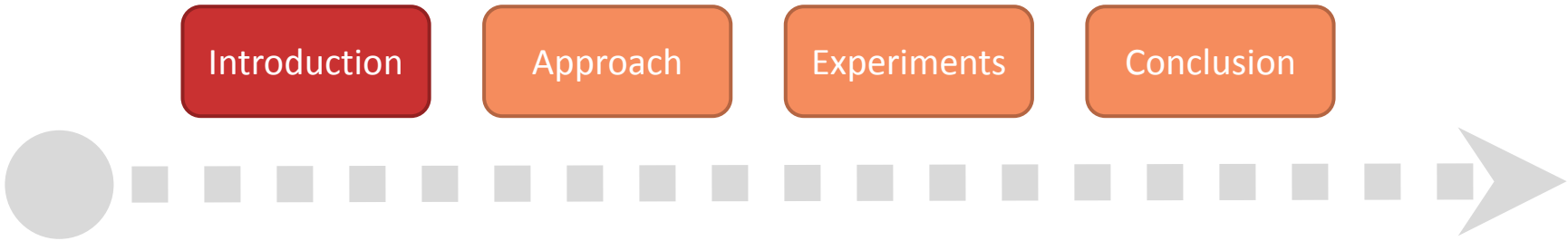
- Motivation
- Extractive Summarization

Motivation

- Speech Summarization
  - Spoken documents are more difficult to browse than texts

    → easy to browse, save time, easily get the key points
- Multi-Party Corpus
  - Speaker information may help summarization

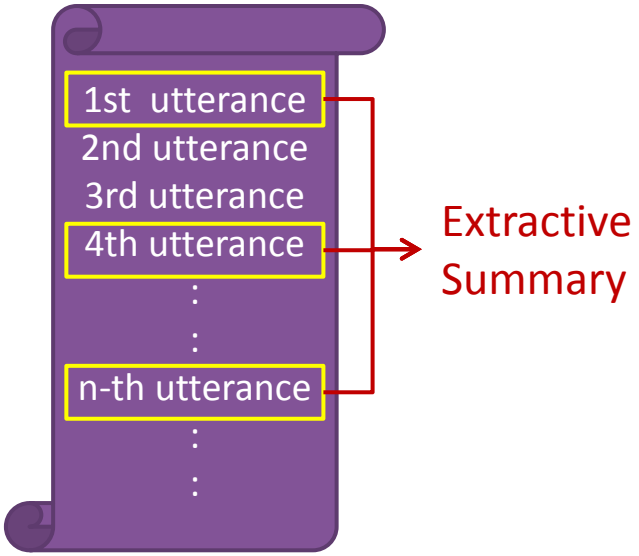
Outline



- Motivation
- Extractive Summarization

Extractive Summarization (1/2)

- Extractive Speech Summarization
  - Select the indicative utterances in a spoken document
  - Cascade the utterances to form a summary



How to select indicative utterances?

Extractive Summarization (2/2)

- Selection of Indicative Utterances
  - Each utterance U in a spoken document d is given an *importance score* I(U, d)
  - Select the indicative utterances based on I(U,d)
  - The number of utterances selected as summary is decided by a predefined ratio

utterance term

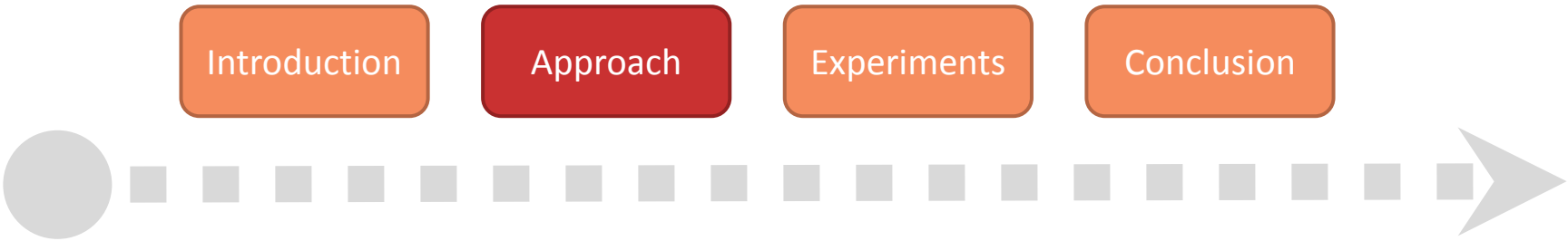
$$U = t_1 t_2 \ldots\ldots t_i \ldots\ldots t_n$$

$$\mathrm{I}(U,d) = \sum_{i=1}^{n} [s(t_i,d) \cdot \ldots\ldots] + \ldots\ldots$$

Importance score

term statistical measure (ex. TF-IDF)

Outline



- Graph Construction
  - Between-Layer Relation via Hidden Parameters
  - Within-Layer Relation via Similarity
- Multi-Layer Mutually Reinforced Random Walk

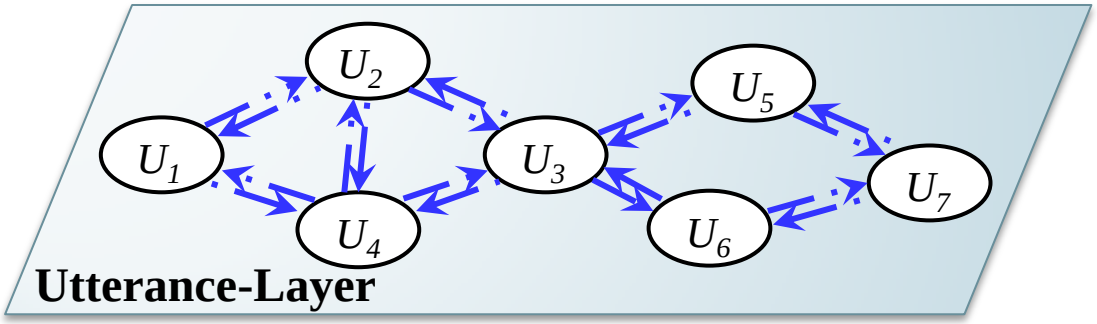
Outline

Introduction | Approach | Experiments | Conclusion

- Graph Construction
  - Between-Layer Relation via Hidden Parameters
  - Within-Layer Relation via Similarity
- Multi-Layer Mutually Reinforced Random Walk

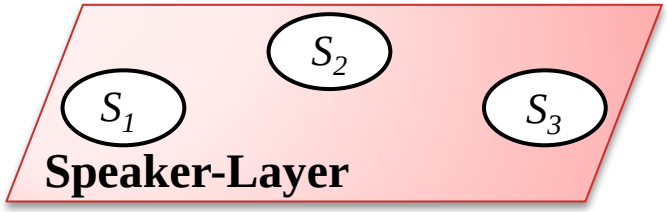
Graph Construction (1/3)

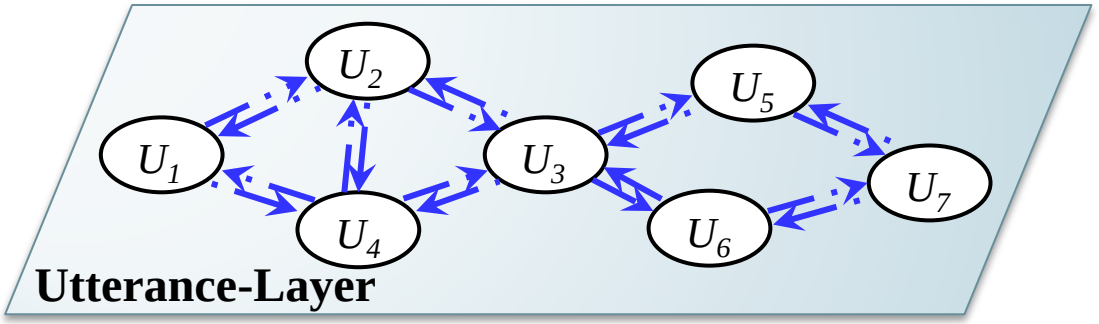
- Utterance-Layer
  - Each node is the utterance in the meeting document
  - The edge is weighted by topical similarity between two utterances



Graph Construction (2/3)

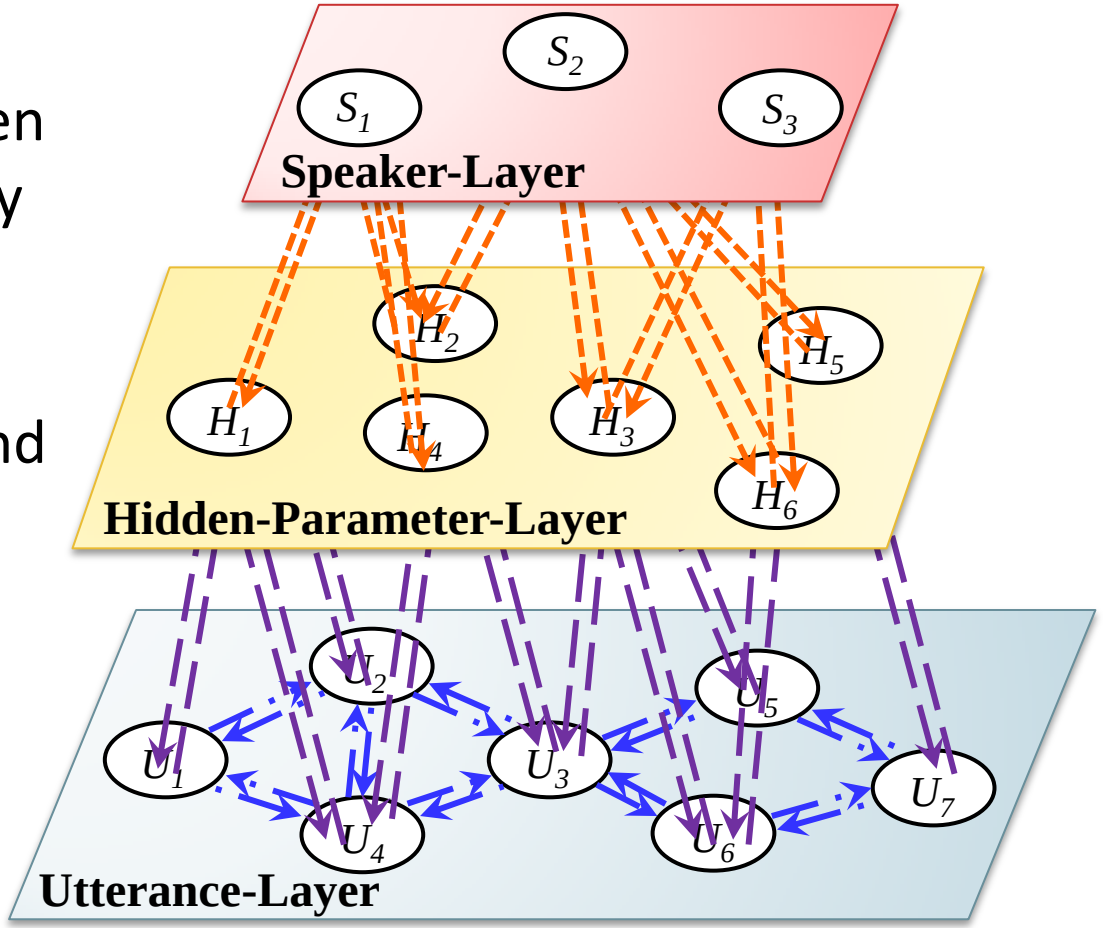
- Speaker-Layer
  - Each node is the speaker in the meeting document
    - Combine all utterances from the same speaker as the speaker node





Graph Construction (3/3)

- Hidden-Parameter-Layer
  - Each node is a hidden parameter shared by speakers and utterances
  - The edge (orange and purple) is weighted by the parameter weight of the speaker/utterance



Outline

Introduction | Approach | Experiments | Conclusion

- Graph Construction
  - Between-Layer Relation via Hidden Parameters
  - Within-Layer Relation via Similarity
- Multi-Layer Mutually Reinforced Random Walk

Between-Layer Relation via Hidden Parameters

① Hidden-Parameter-Layer → Term-Layer

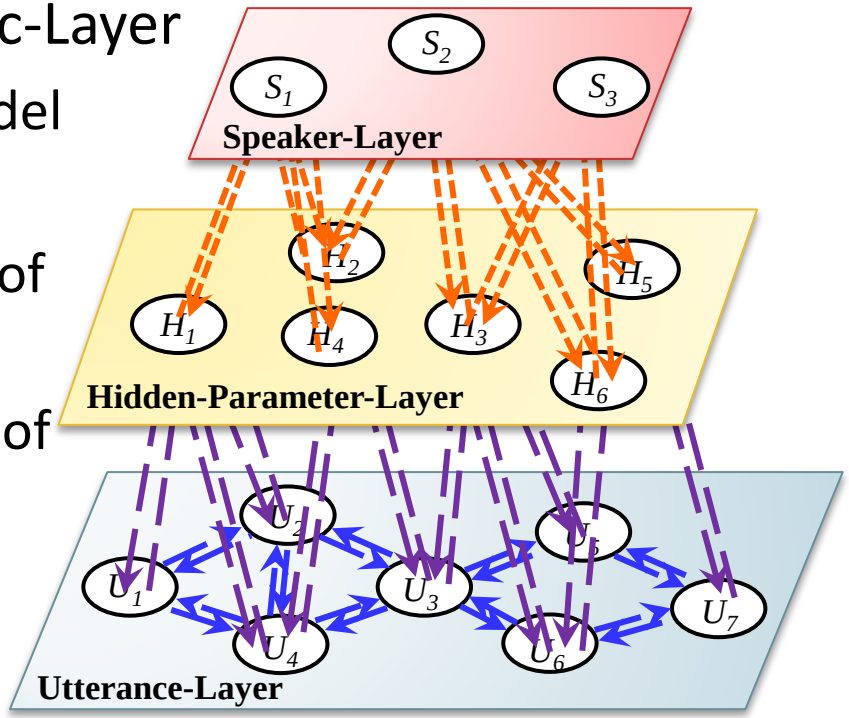
terms from the vocabulary to represent hidden parameters

- L_{SH}: avg TFIDF of the terms in the speakers
- L_{UH}: avg TFIDF of the terms in the utterances

② Hidden-Parameter-Layer → Topic-Layer

latent topics from trained topic model as hidden parameters

- L_{SH}: avg Latent Topic Significance of the topics in the speakers
- L_{UH}: avg Latent Topic Significance of the topics in the utterances



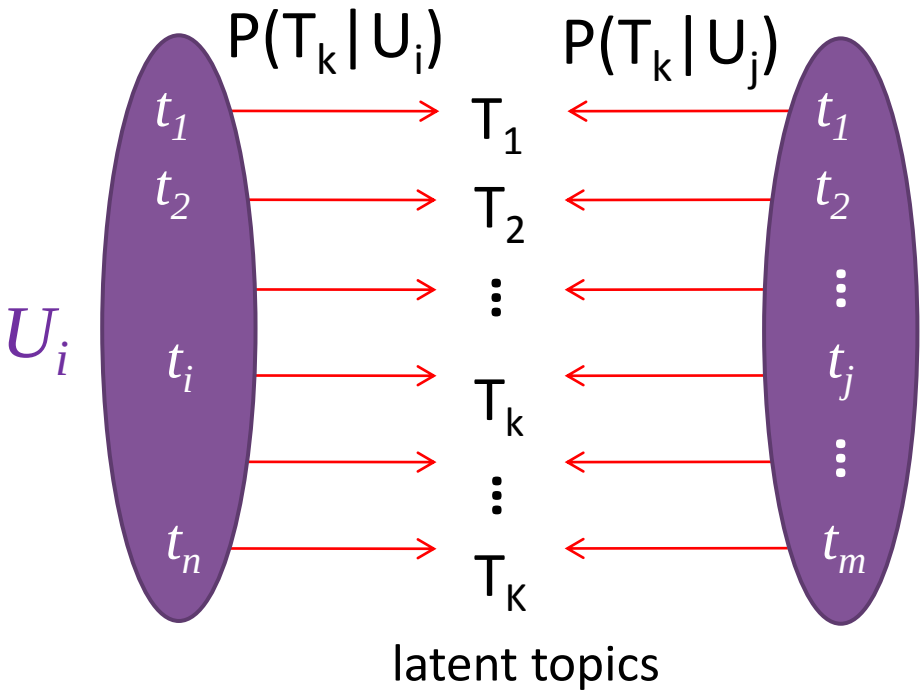
Outline

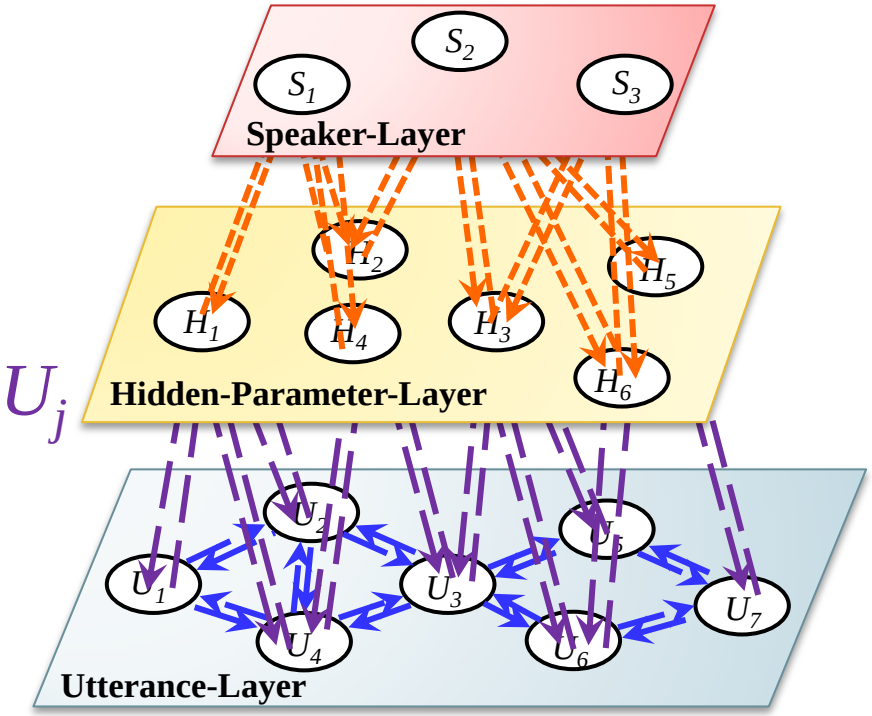
Introduction | Approach | Experiments | Conclusion

- Graph Construction
  - Between-Layer Relation via Hidden Parameters
  - Within-Layer Relation via Similarity
- Multi-Layer Mutually Reinforced Random Walk

Within-Layer Relation via Similarity

- Similarity Matrix
  - L_{UU}: utterance-to-utterance relation (topical similarity)
    edge weight TopicSim(U_i, U_j) (utterance U_i → utterance U_j)





Outline

Introduction | Approach | Experiments | Conclusion

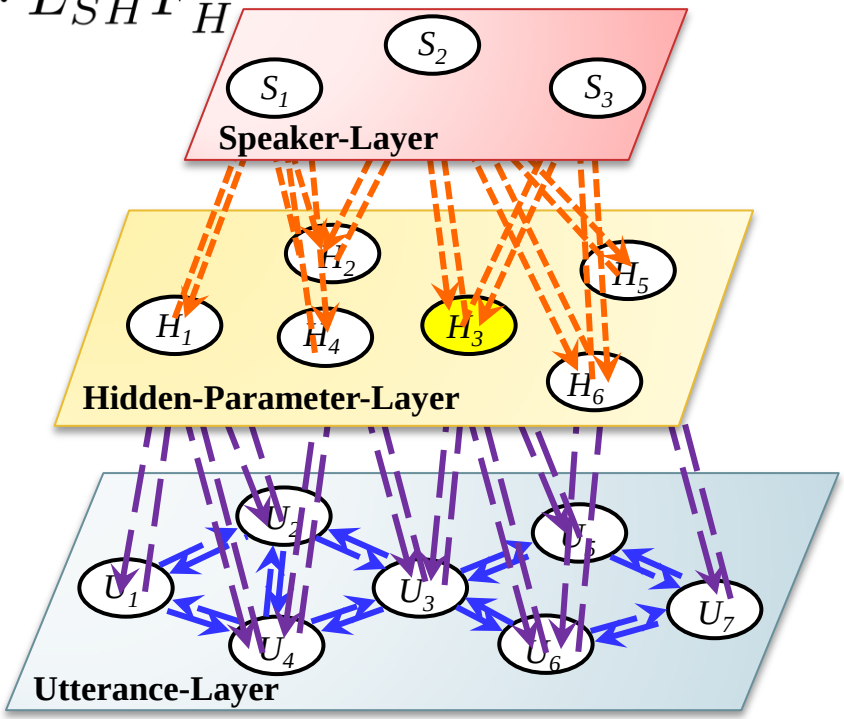
- Graph Construction
  - Between-Layer Relation via Hidden Parameters
  - Within-Layer Relation via Similarity
- Multi-Layer Mutually Reinforced Random Walk

Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

hidden parameter scores at (t+1)-th iteration

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$



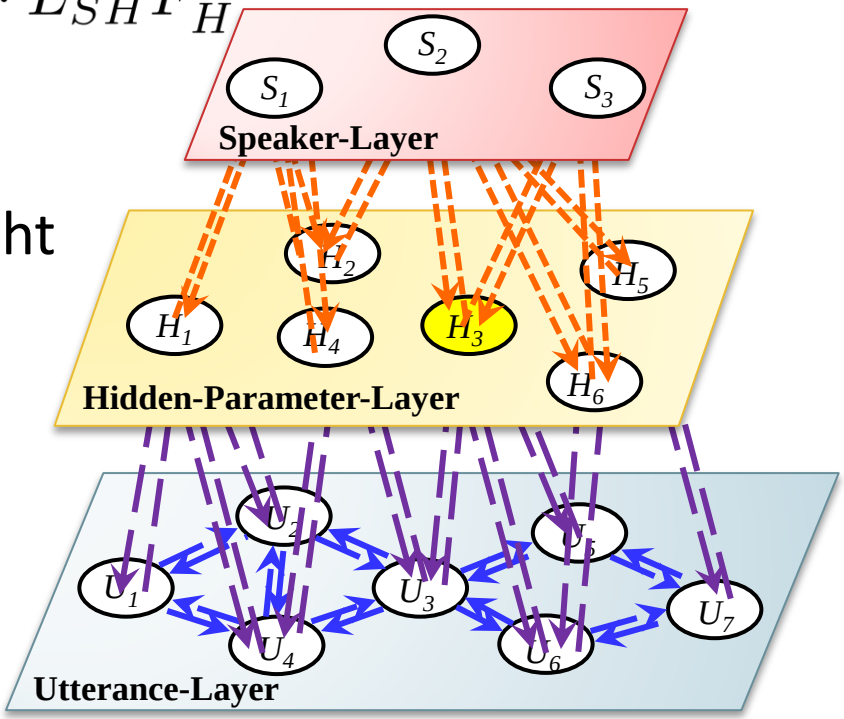
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

original importance of hidden parameters

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)\boxed{F_H^{(0)}} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha) F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha) F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight



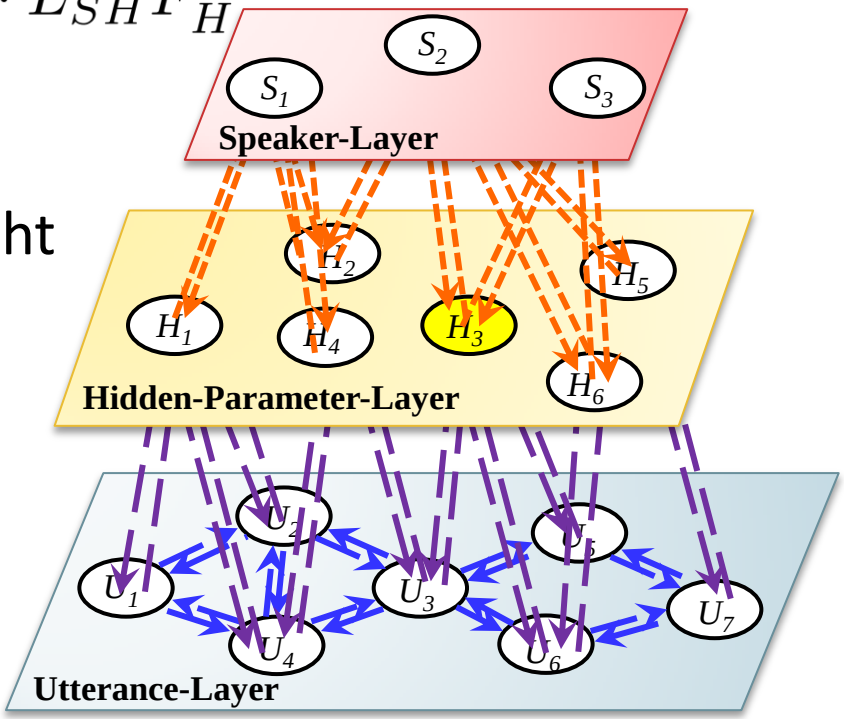
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from utterances weighted by parameter values

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot \boxed{L_{UH}^T F_U^{(t)}} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight



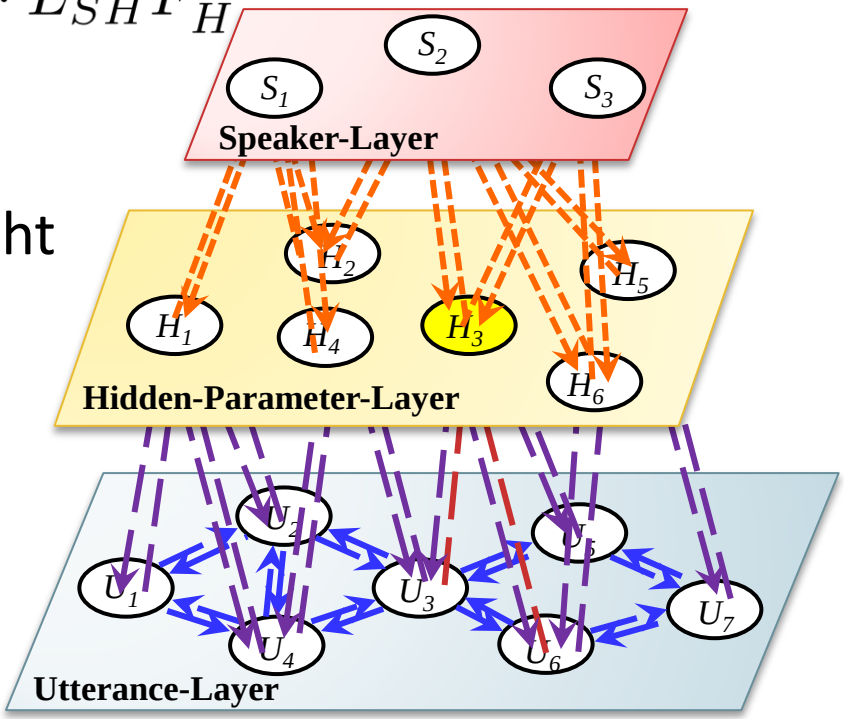
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from utterances weighted by parameter values

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot \boxed{L_{UH}^T F_U^{(t)}} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight



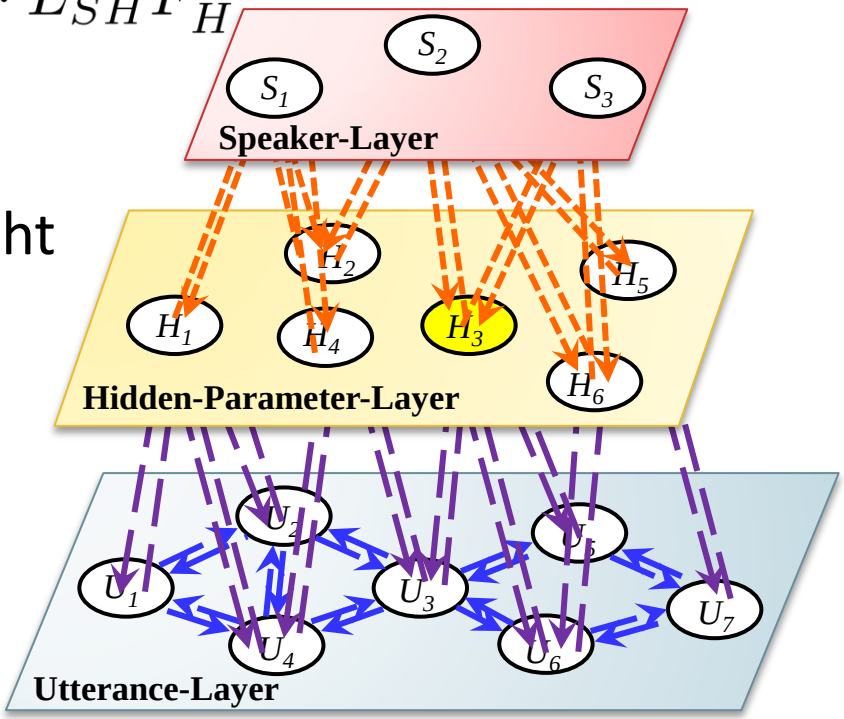
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from speakers weighted by parameter values

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot \boxed{L_{SH}^T F_S^{(t)}} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight



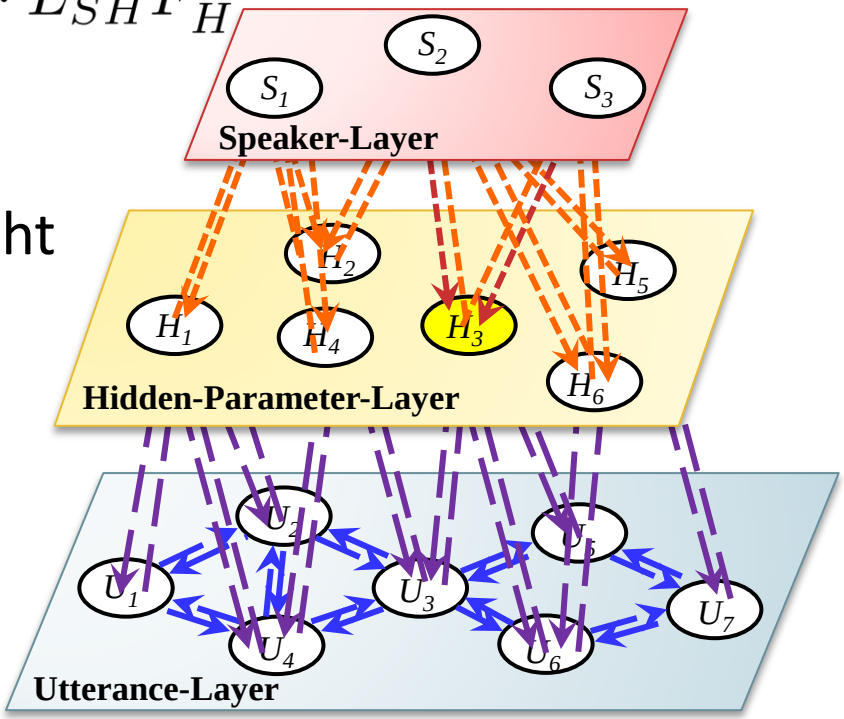
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from speakers weighted by parameter values

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot \boxed{L_{SH}^T F_S^{(t)}} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight



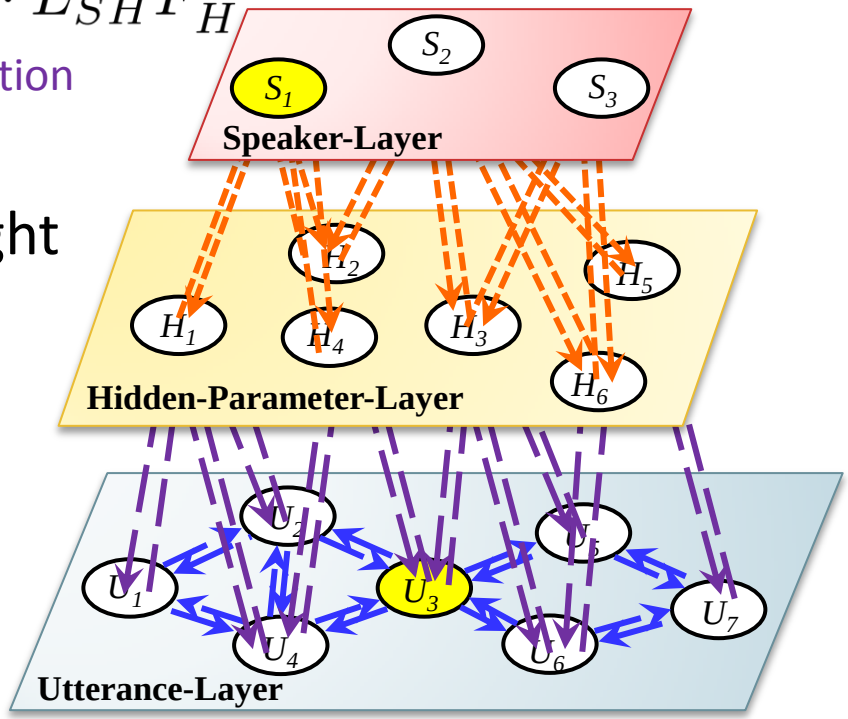
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

utterance/speaker scores at (t+1)-th iteration

- Original importance
  - Hidden parameter: equal weight



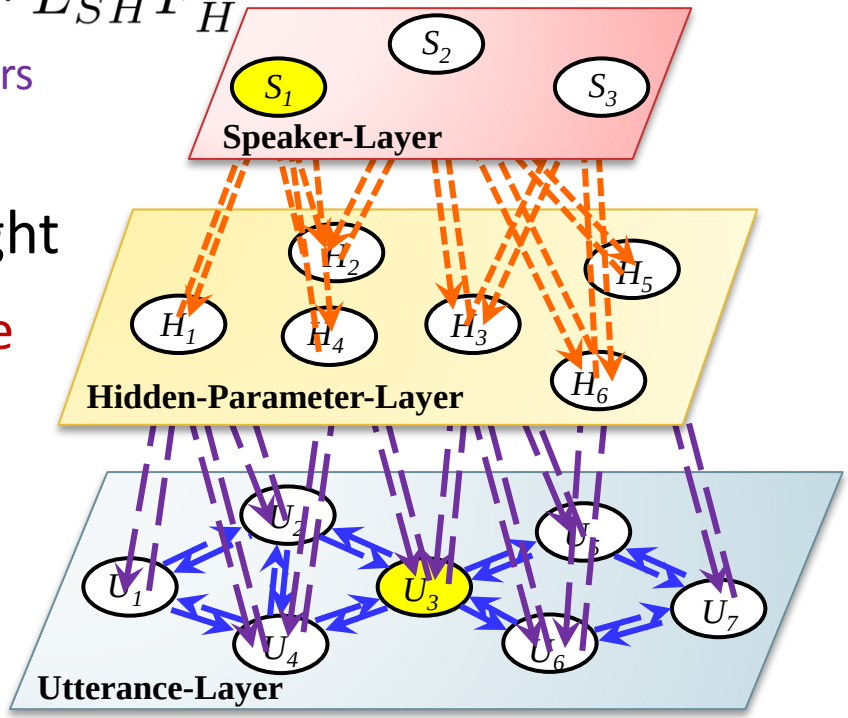
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

original importance of utterances/speakers

- Original importance
  - Hidden parameter: equal weight
  - Utterance: $I(U, d)$ → baseline
  - Speaker: equal weight



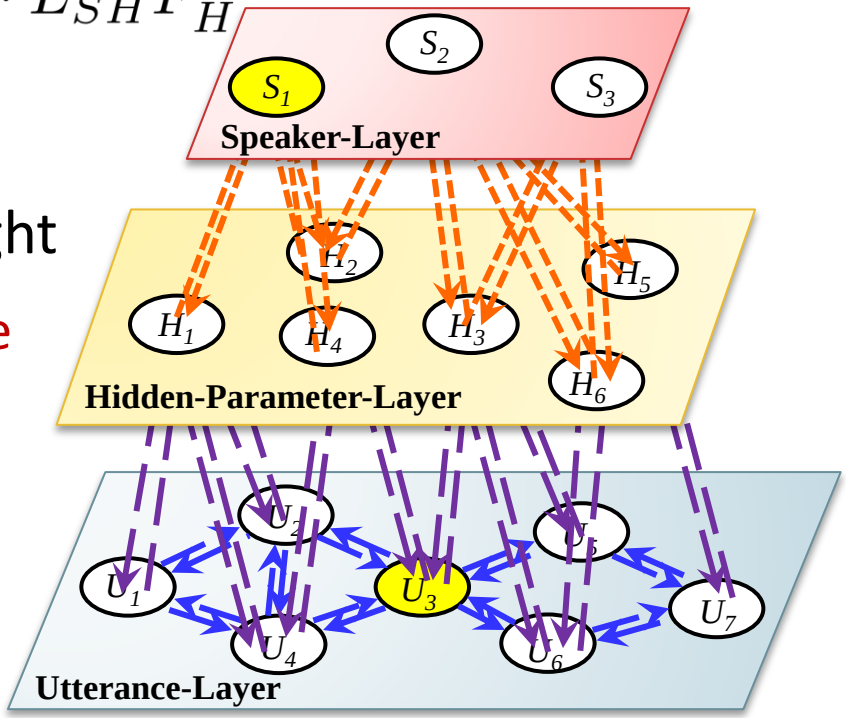
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from parameters weighted by between-layer relation and then within-layer relation

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight
  - Utterance: $I(U, d)$ → baseline
  - Speaker: equal weight



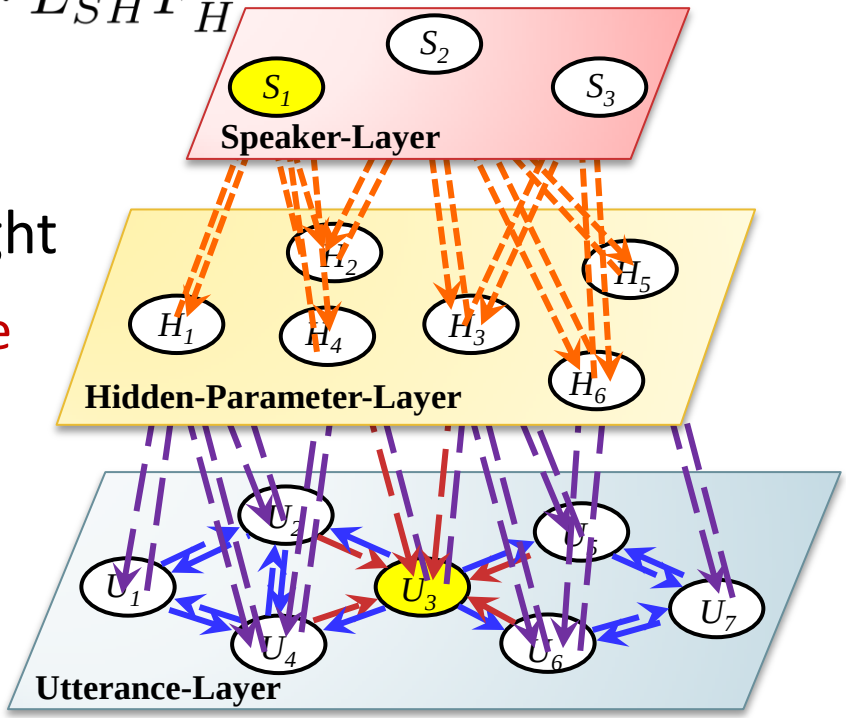
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from parameters weighted by between-layer relation and then within-layer relation

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight
  - Utterance: $I(U, d)$ → baseline
  - Speaker: equal weight



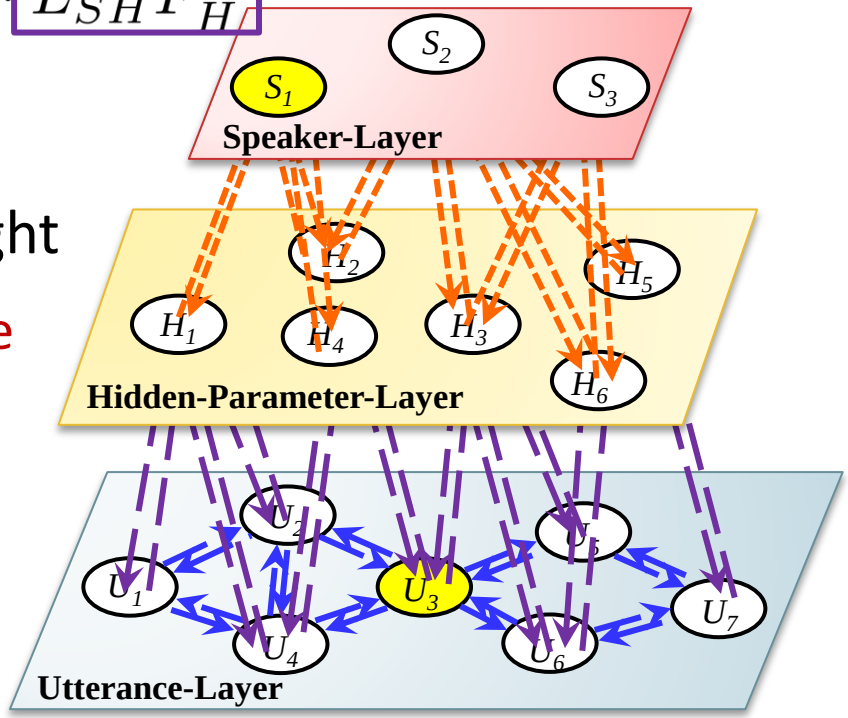
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from parameters weighted by between-layer relation

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight
  - Utterance: $I(U, d)$ → baseline
  - Speaker: equal weight



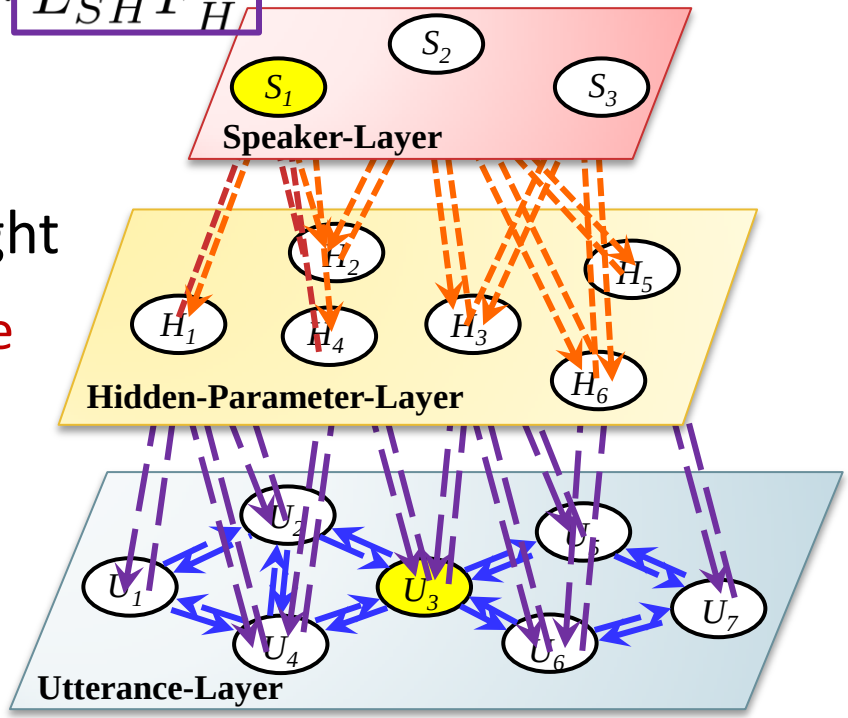
Multi-Layer Mutual Reinforced Random Walk (1/2)

- Mathematical Formulation

scores propagated from parameters weighted by between-layer relation

$$\begin{cases} F_H^{(t+1)} = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^{(t)} + \alpha \cdot L_{SH}^T F_S^{(t)} \\ F_U^{(t+1)} = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^{(t)} \\ F_S^{(t+1)} = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^{(t)} \end{cases}$$

- Original importance
  - Hidden parameter: equal weight
  - Utterance: $I(U, d)$ → baseline
  - Speaker: equal weight



Multi-Layer Mutual Reinforced Random Walk (2/2)

- Mathematical Formulation

$$\begin{cases} F_H^* = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^* + \alpha \cdot L_{SH}^T F_S^* \\ F_U^* = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^* \\ F_S^* = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^* \end{cases}$$

Parameter node H can get higher score when

① More important utterances have higher weights corresponding to the hidden parameter H

② More important speakers have higher weights corresponding to the hidden parameter H

→ We can unsupervised learn important parameters (terms/topics) by the approach

Multi-Layer Mutual Reinforced Random Walk (2/2)

- Mathematical Formulation

$$\begin{cases} F_H^* = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^* + \alpha \cdot L_{SH}^T F_S^* \\ F_U^* = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^* \\ F_S^* = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^* \end{cases}$$

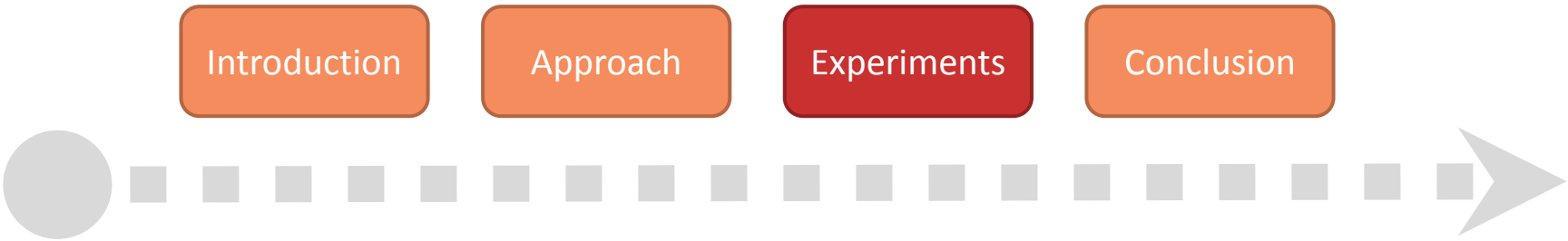
Utterance node U can get higher score when

① Higher original importance I(U, d)

② More important utterances similar to utterance U

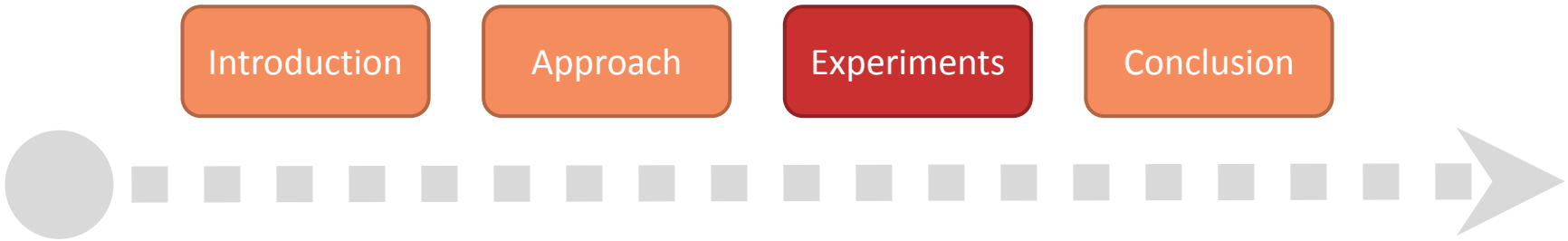
③ More important hidden parameters with higher weights corresponding to utterance U

Outline



- Experimental Setup
- Evaluation Metrics
- Results

Outline



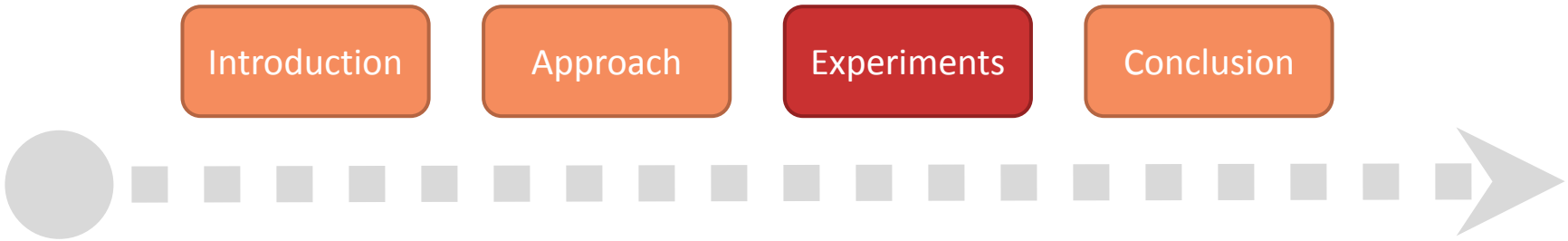
- Experimental Setup
- Evaluation Metrics
- Results

Experimental Setup

- CMU Speech Meeting Corpus
  - 10 meetings from 2006/04 – 2006/06
  - #Speaker: 6 (total), 2-4 (each meeting)
  - WER = 44%
- Reference Summaries
  - Manually labeled by two annotators
- Parameter Setting
  - α = 0.45
  - Extractive summary ratio = 10%, 20%

$$\begin{cases} F_H^* = (1-2\alpha)F_H^{(0)} + \alpha \cdot L_{UH}^T F_U^* + \alpha \cdot L_{SH}^T F_S^* \\ F_U^* = (1-2\alpha)F_U^{(0)} + 2\alpha \cdot L_{UU}^T L_{UH} F_H^* \\ F_S^* = (1-2\alpha)F_S^{(0)} + 2\alpha \cdot L_{SH} F_H^* \end{cases}$$

Outline

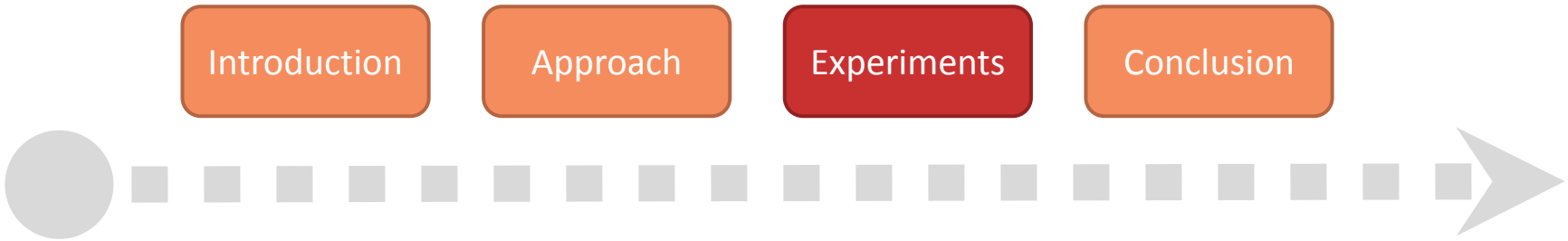


- Experimental Setup
- Evaluation Metrics
- Results

Evaluation Metrics

- ROUGE
  - ROUGE-1
    - F-measure of matched unigram between extracted summary and reference summary
  - ROUGE-L (Longest Common Subsequence)
    - F-measure of matched LCS between extracted summary and reference summary

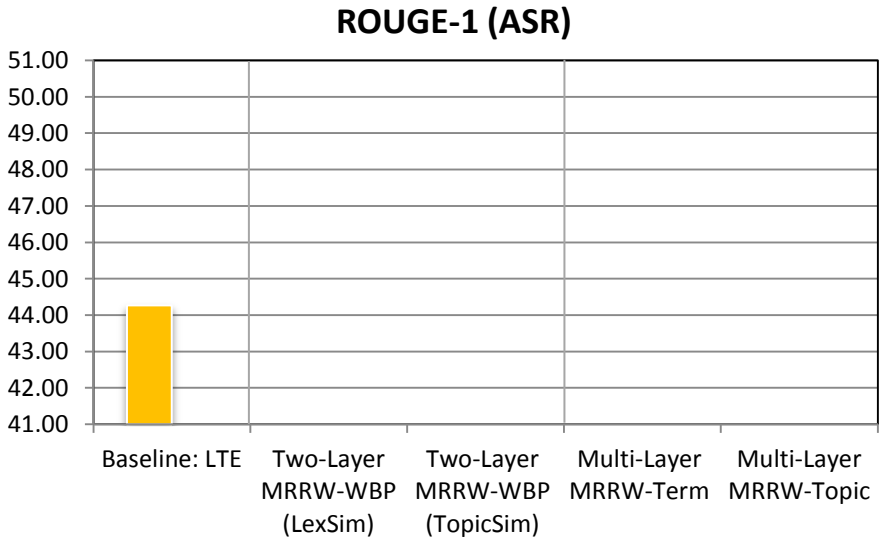
Outline

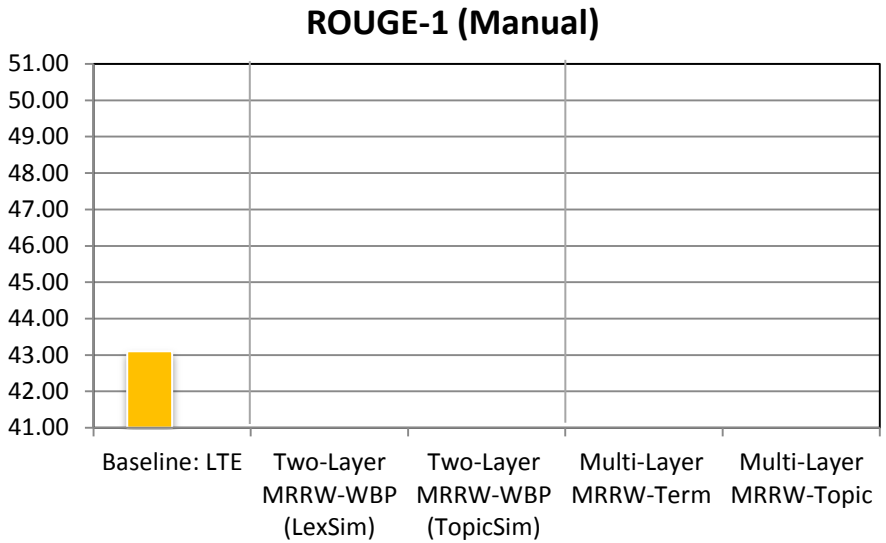


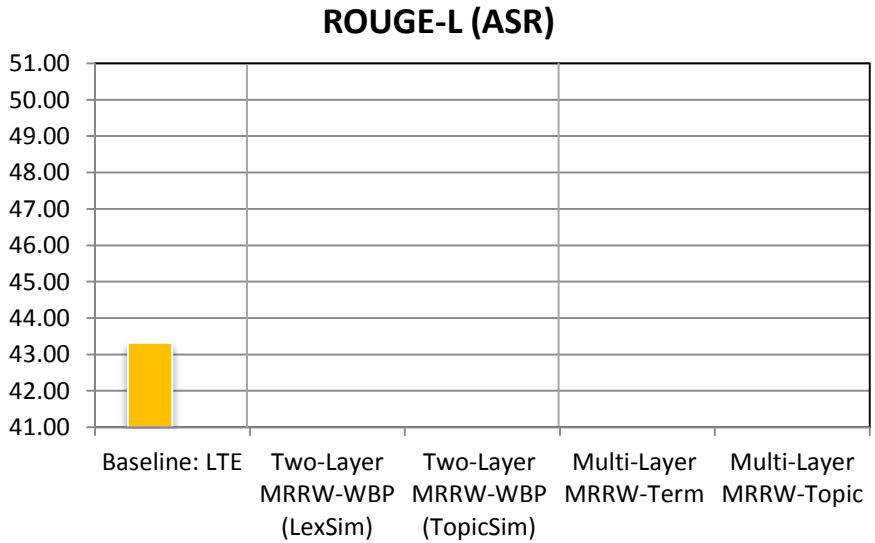
- Experimental Setup
- Evaluation Metrics
- Results

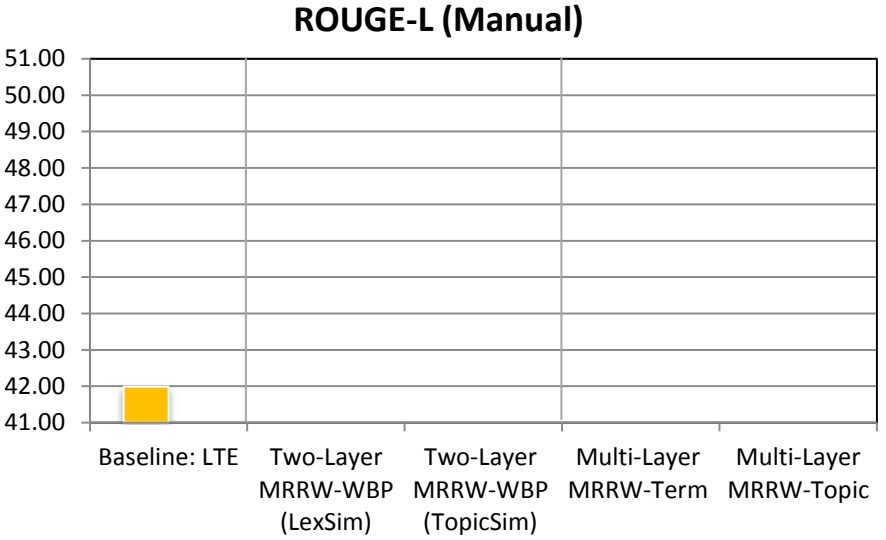
10%

Results


ROUGE-1 (ASR)
51.00
50.00
49.00
48.00
47.00
46.00
45.00
44.00
43.00
42.00
41.00
Baseline: LTE
Two-Layer MRRW-WBP (LexSim)
Two-Layer MRRW-WBP (TopicSim)
Multi-Layer MRRW-Term
Multi-Layer MRRW-Topic

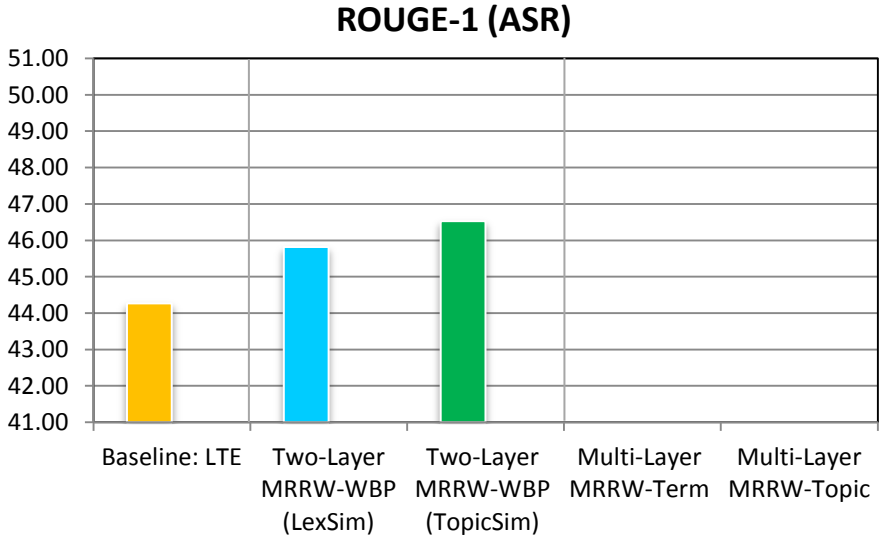

ROUGE-1 (Manual)
51.00
50.00
49.00
48.00
47.00
46.00
45.00
44.00
43.00
42.00
41.00
Baseline: LTE
Two-Layer MRRW-WBP (LexSim)
Two-Layer MRRW-WBP (TopicSim)
Multi-Layer MRRW-Term
Multi-Layer MRRW-Topic

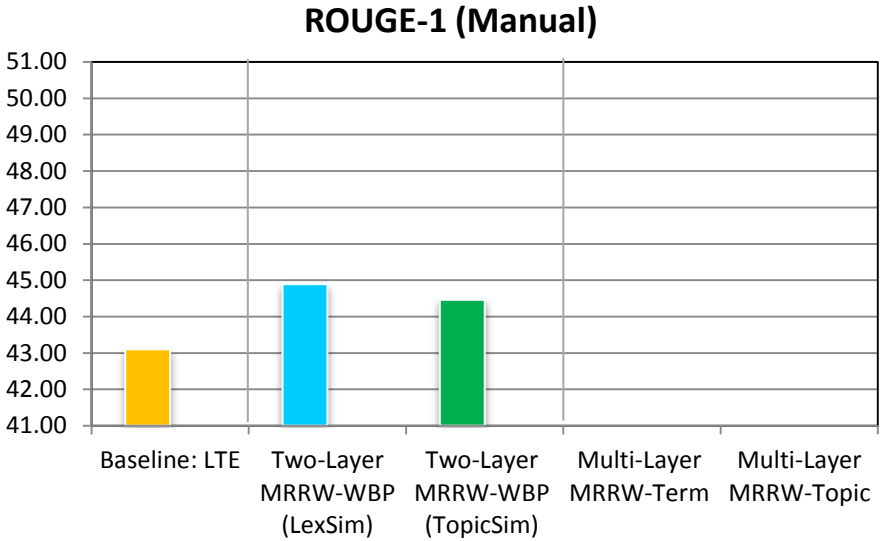

ROUGE-L (ASR)
51.00
50.00
49.00
48.00
47.00
46.00
45.00
44.00
43.00
42.00
41.00
Baseline: LTE
Two-Layer MRRW-WBP (LexSim)
Two-Layer MRRW-WBP (TopicSim)
Multi-Layer MRRW-Term
Multi-Layer MRRW-Topic

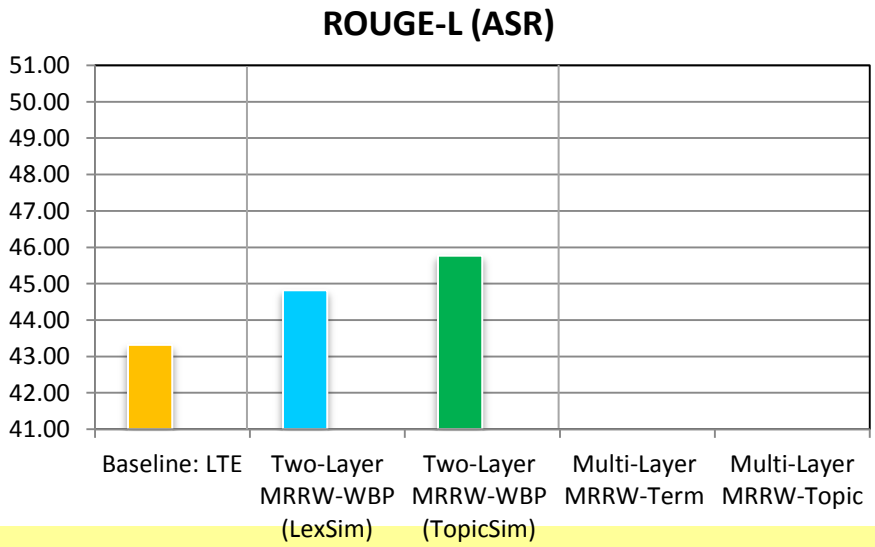

ROUGE-L (Manual)
51.00
50.00
49.00
48.00
47.00
46.00
45.00
44.00
43.00
42.00
41.00
Baseline: LTE
Two-Layer MRRW-WBP (LexSim)
Two-Layer MRRW-WBP (TopicSim)
Multi-Layer MRRW-Term
Multi-Layer MRRW-Topic

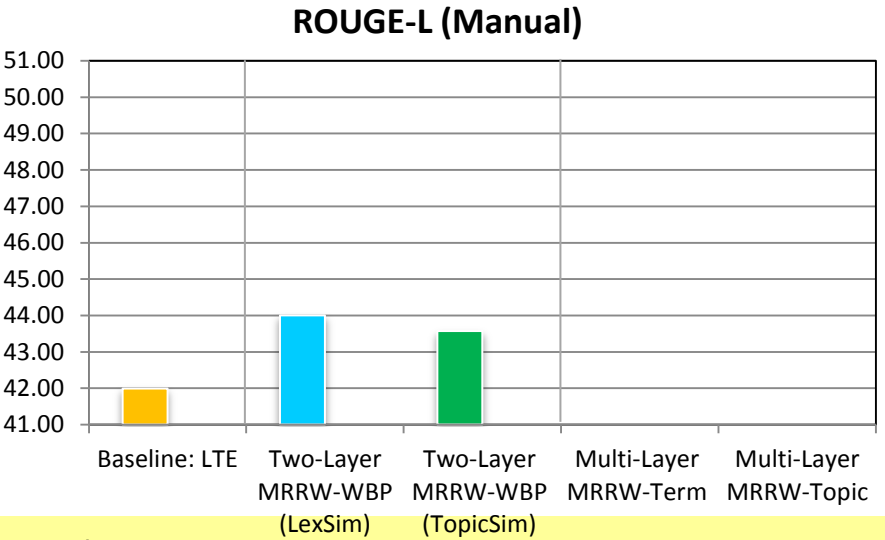
10%

Results – Baseline & Two-Layer Graph Approaches





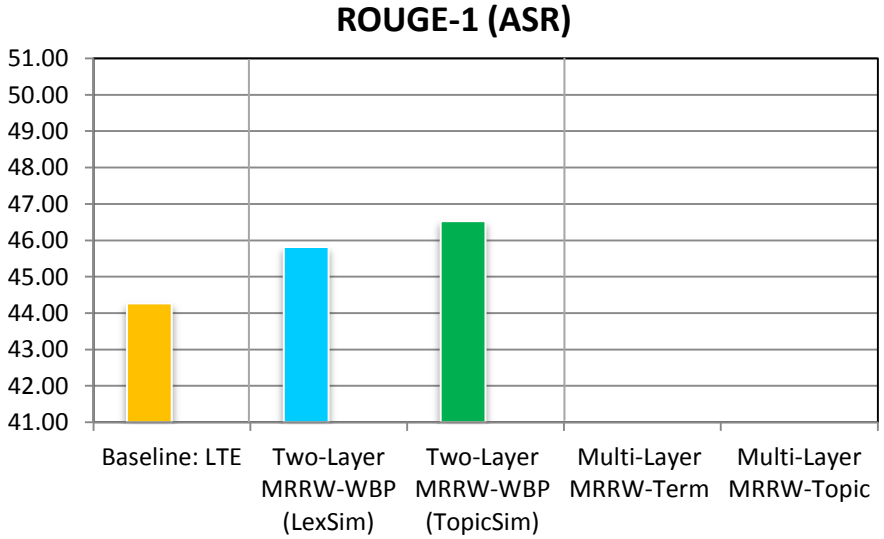


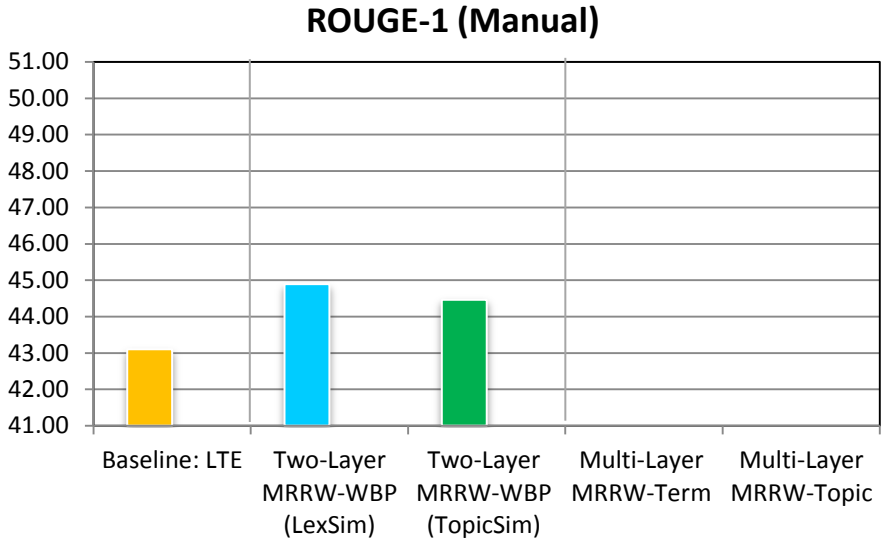


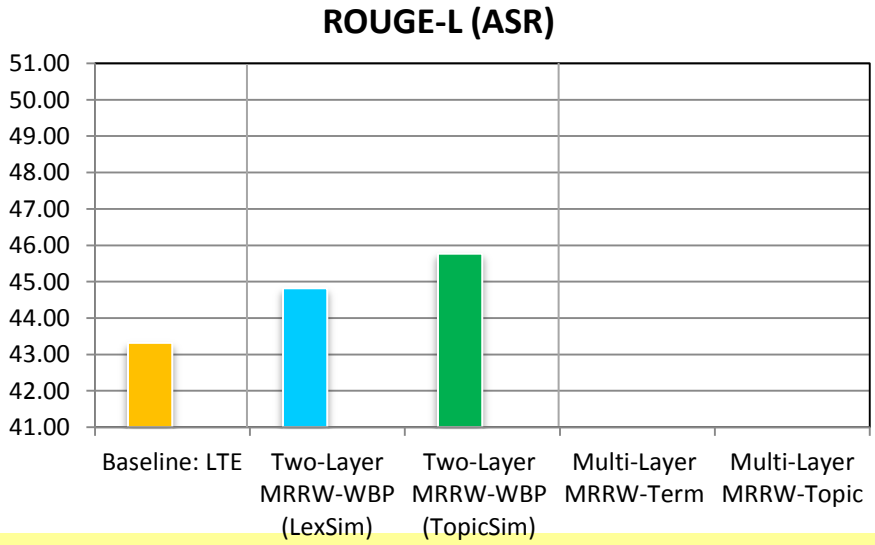
Two-layer approaches (w/o hidden parameters) are significantly better than baseline

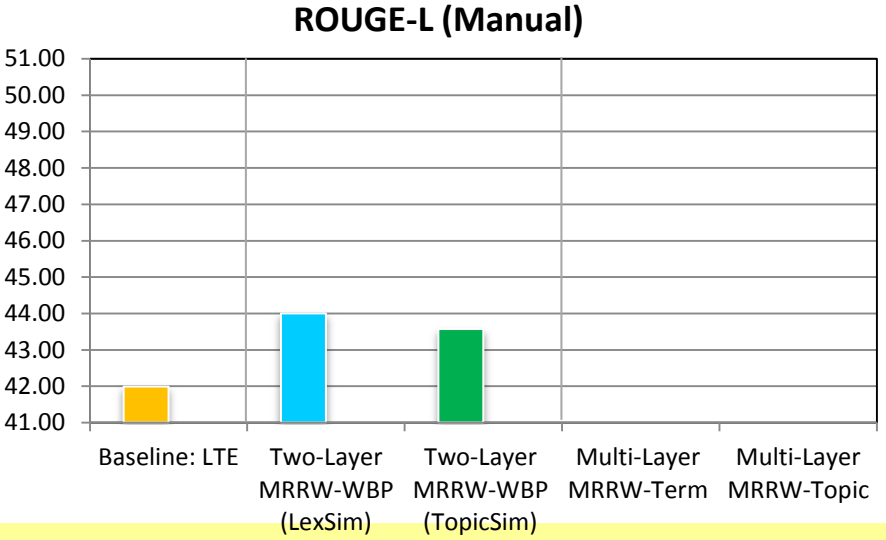
10%

Results – Baseline & Two-Layer Graph Approaches





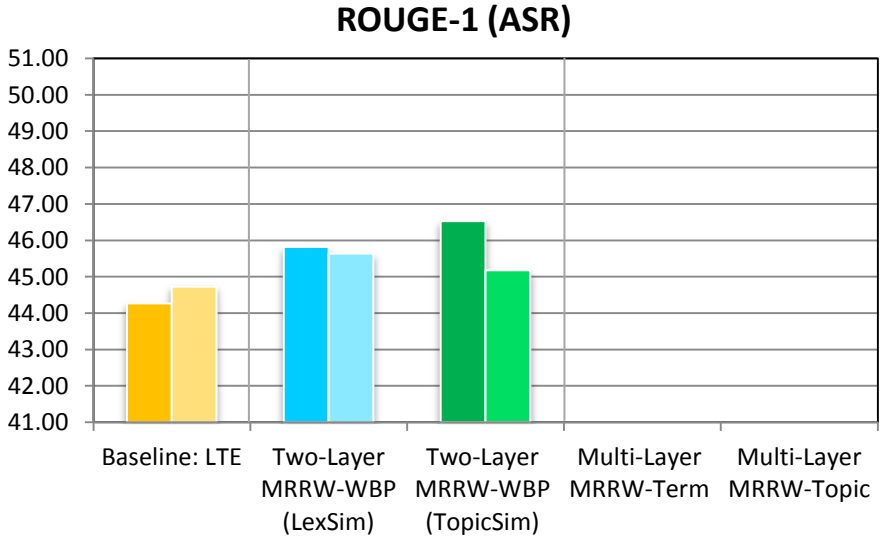


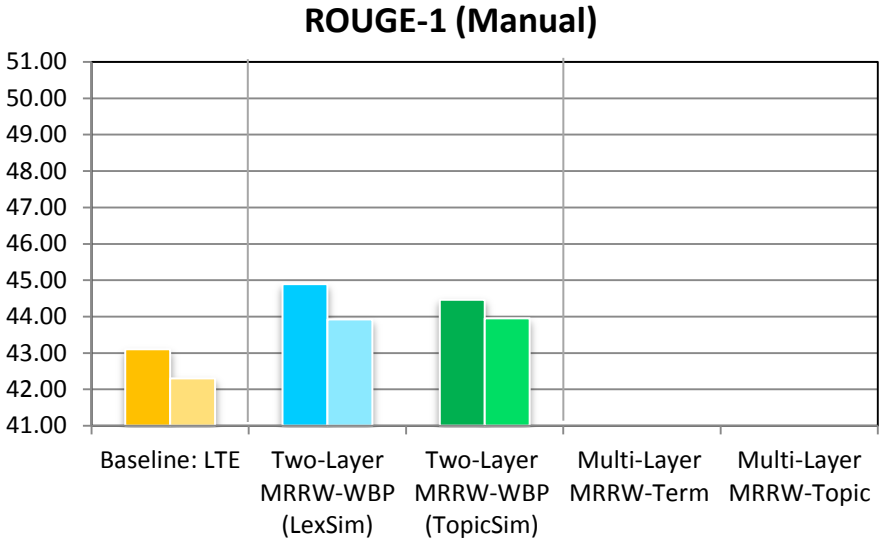


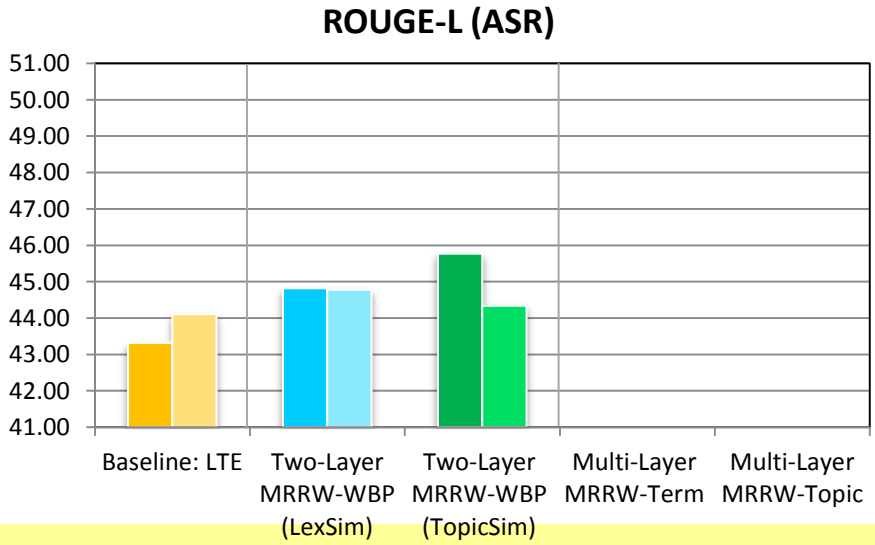
The utterances from the speakers who speak more important utterances tend to be more important

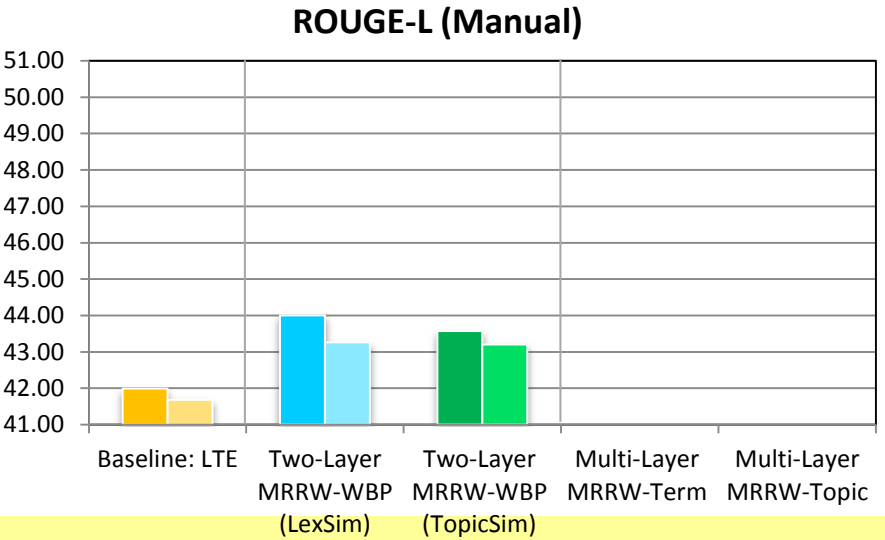
10% & 20%

Results – Baseline & Two-Layer Graph Approaches





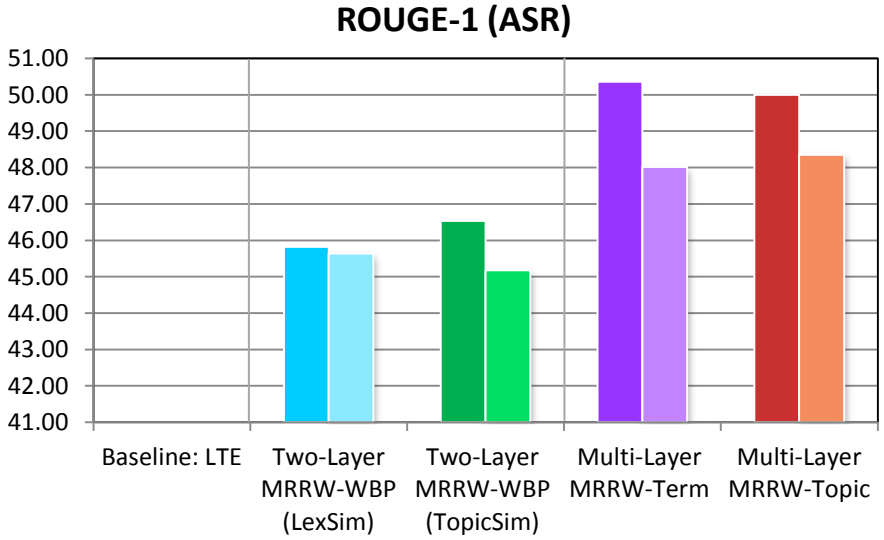


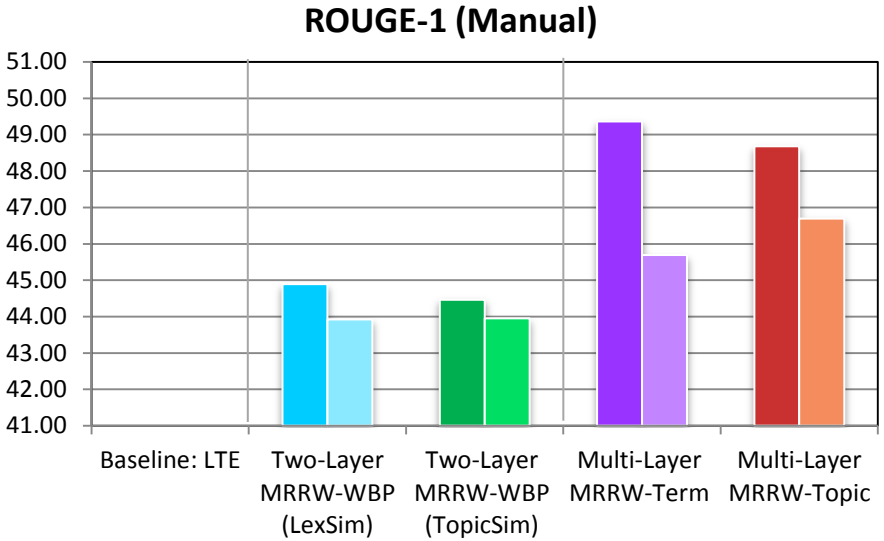


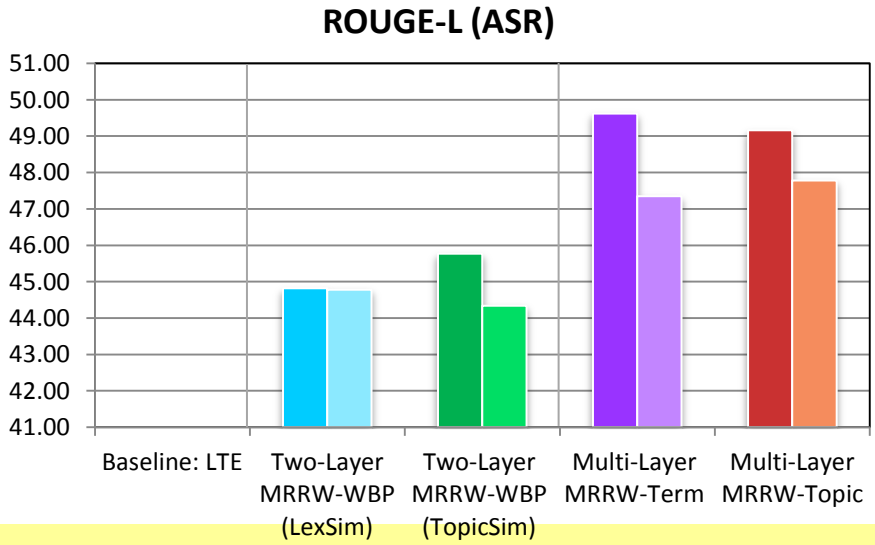
20% summaries perform similarly

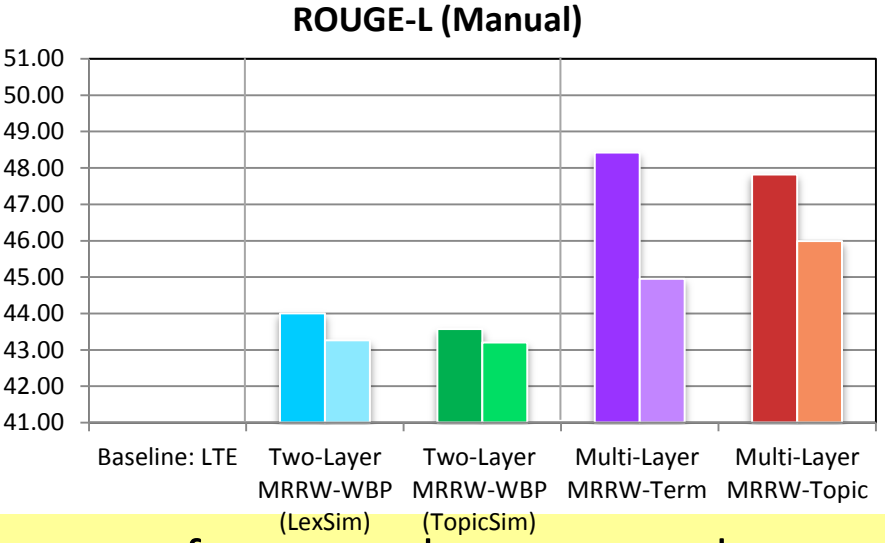
10% & 20%

Results – Effectiveness of Hidden Parameters





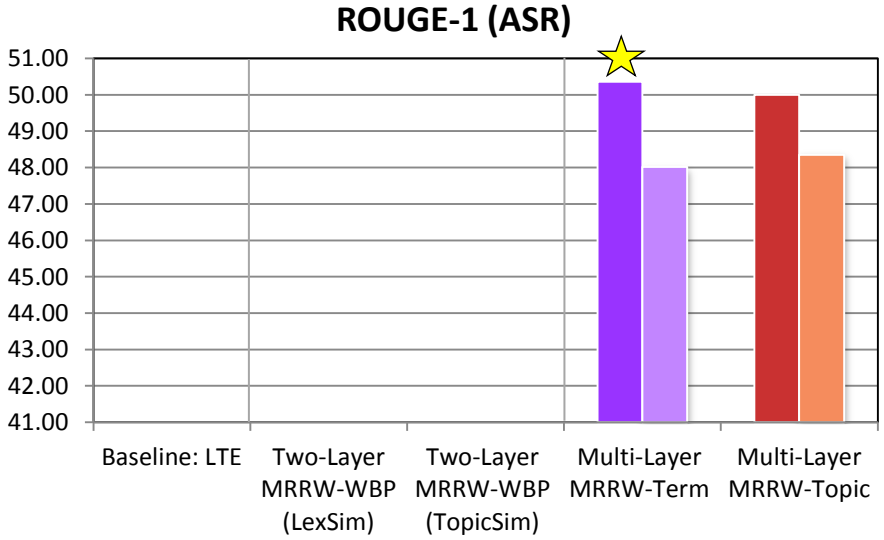


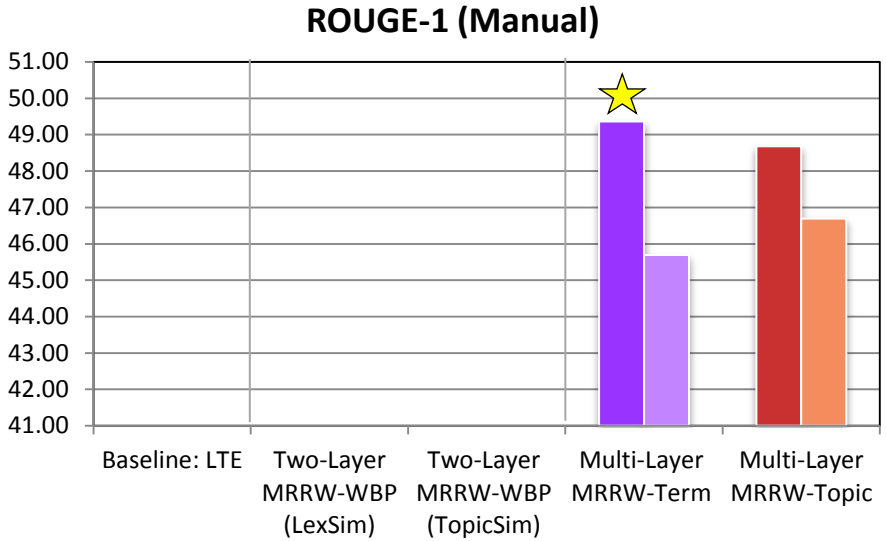


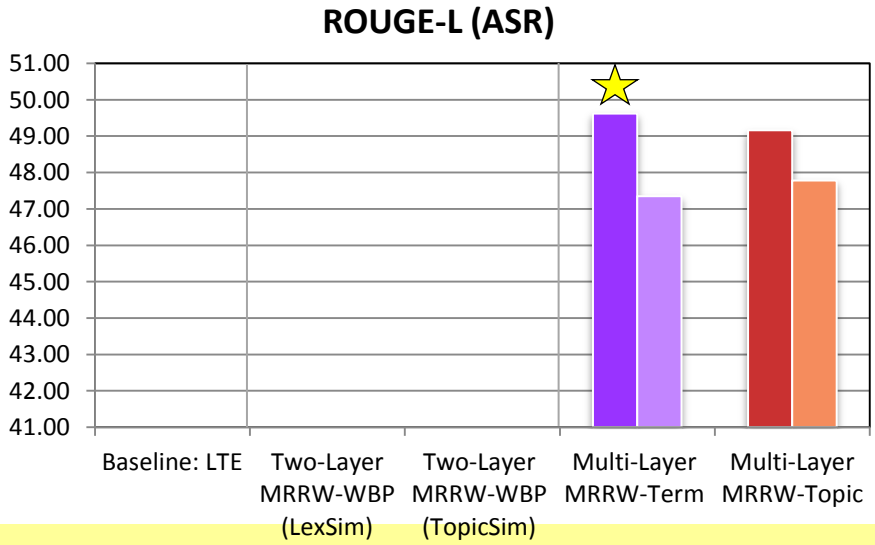
Adding hidden parameters significantly outperforms two-layer approaches

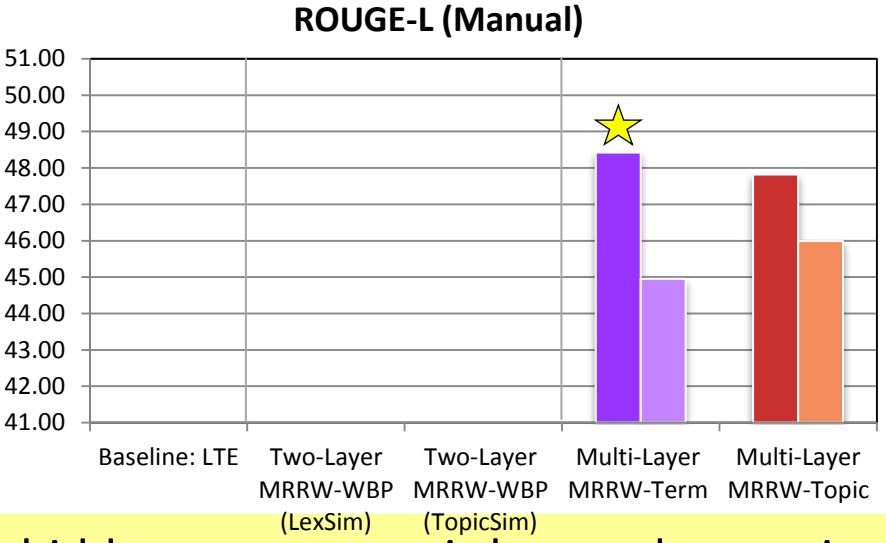
10% & 20%

Results – Different Types of Hidden Parameters





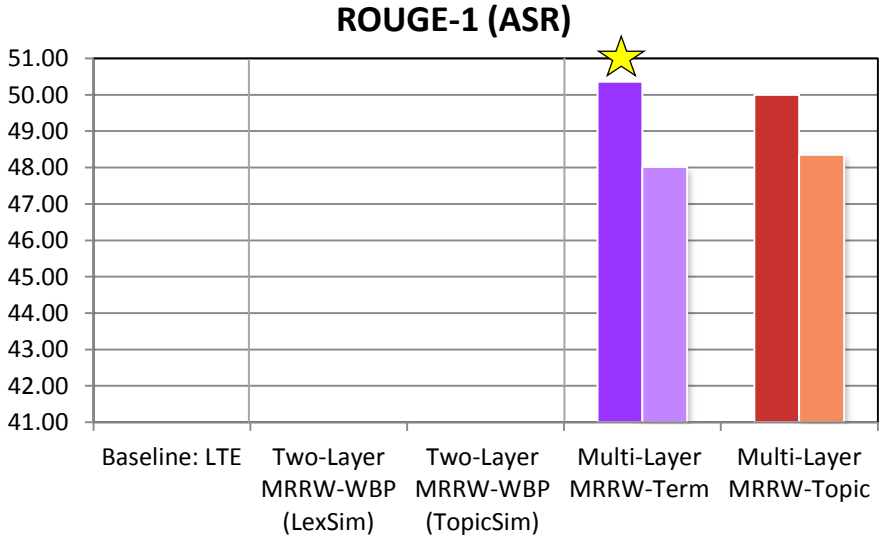


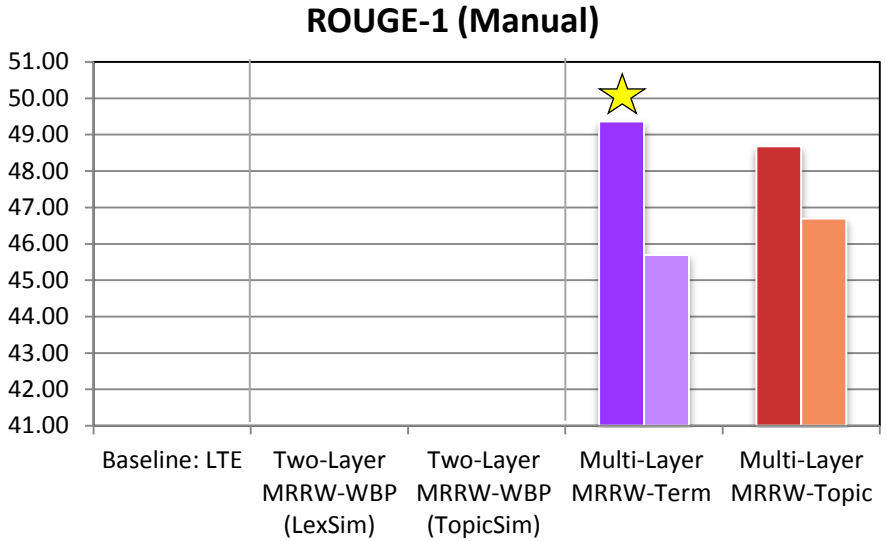


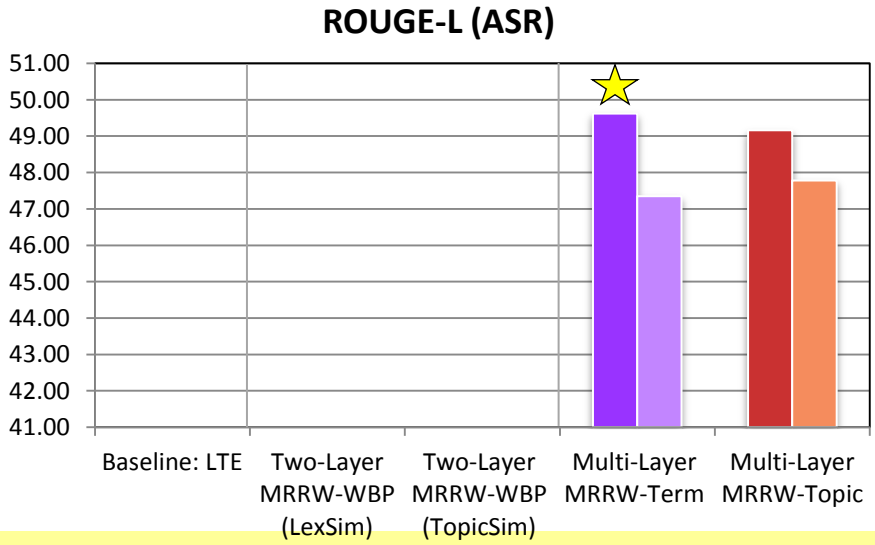
For shorter summaries, using terms as the hidden parameters is better than topics

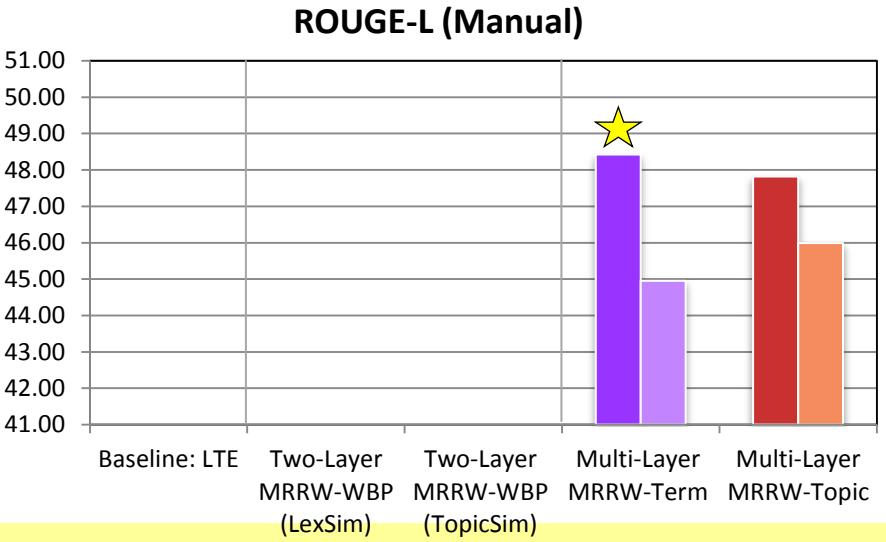
10% & 20%

Results – Different Types of Hidden Parameters





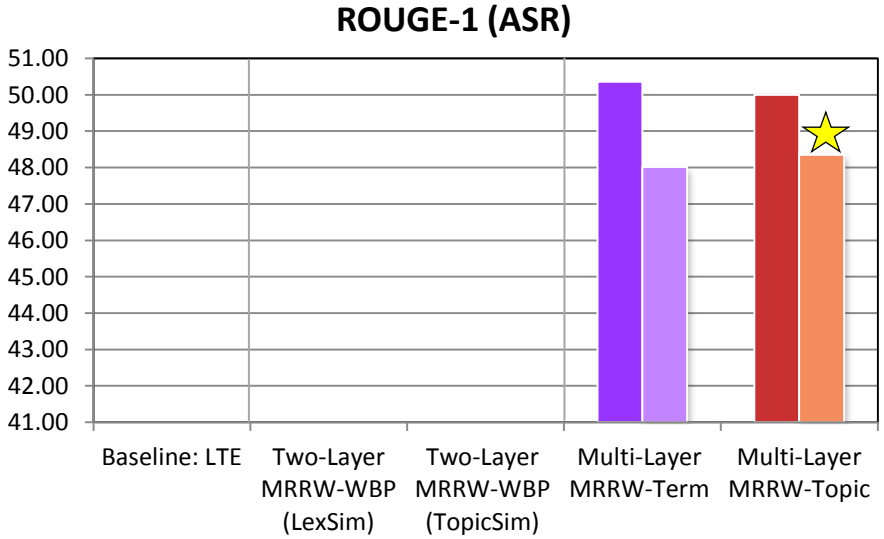


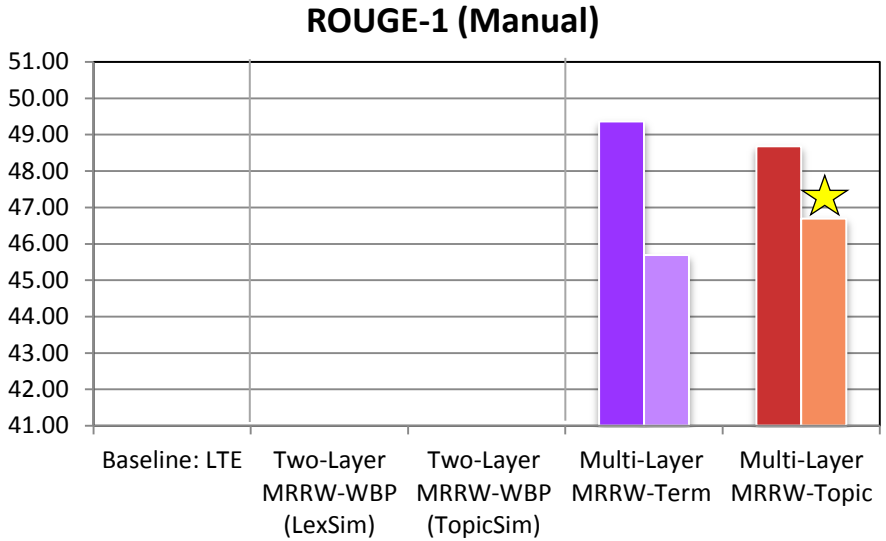


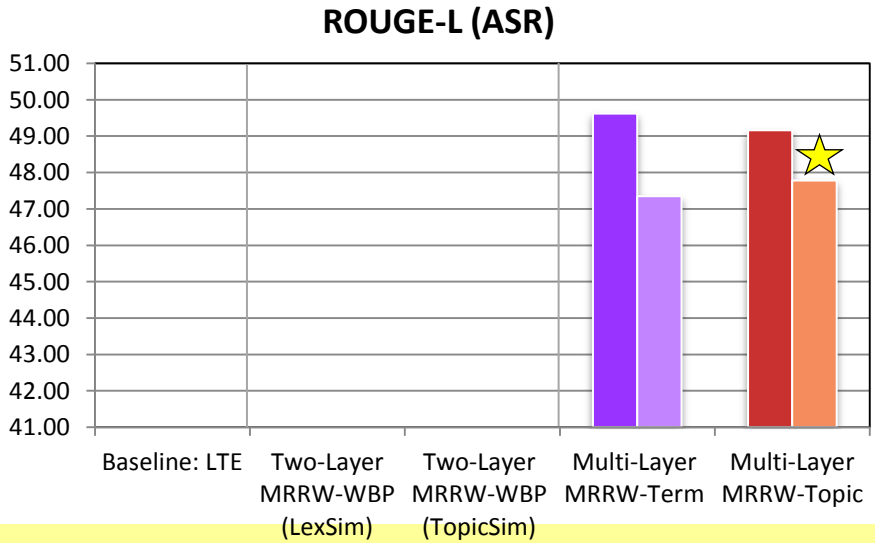
Lexical similarity can model relations accurately since correctly recognized important words

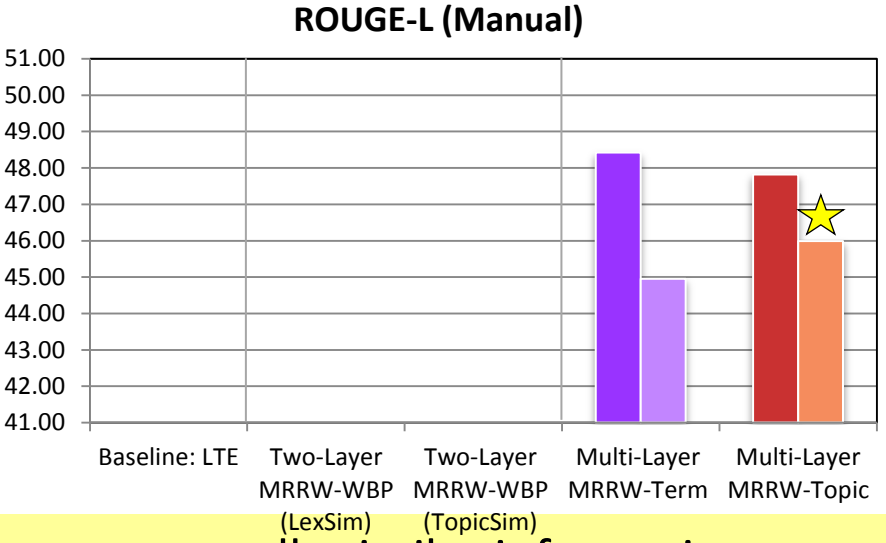
10% & 20%

Results – Different Types of Hidden Parameters





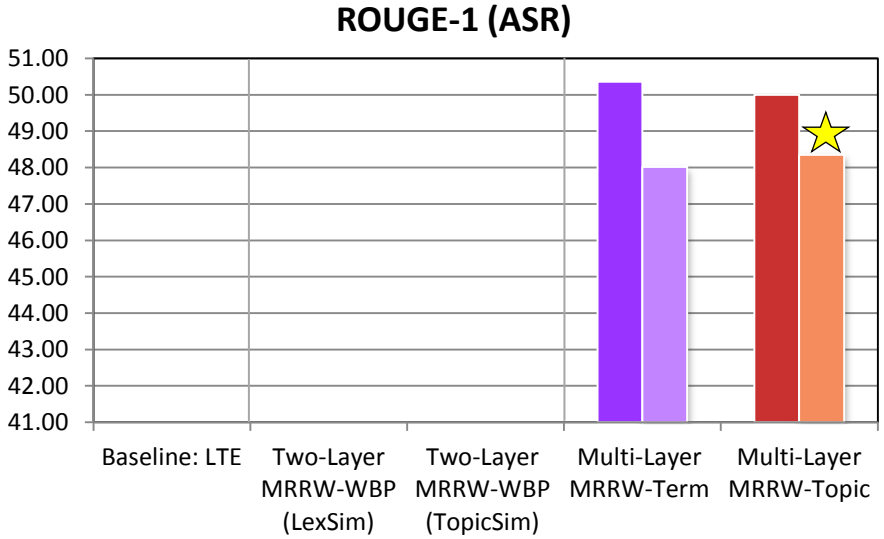


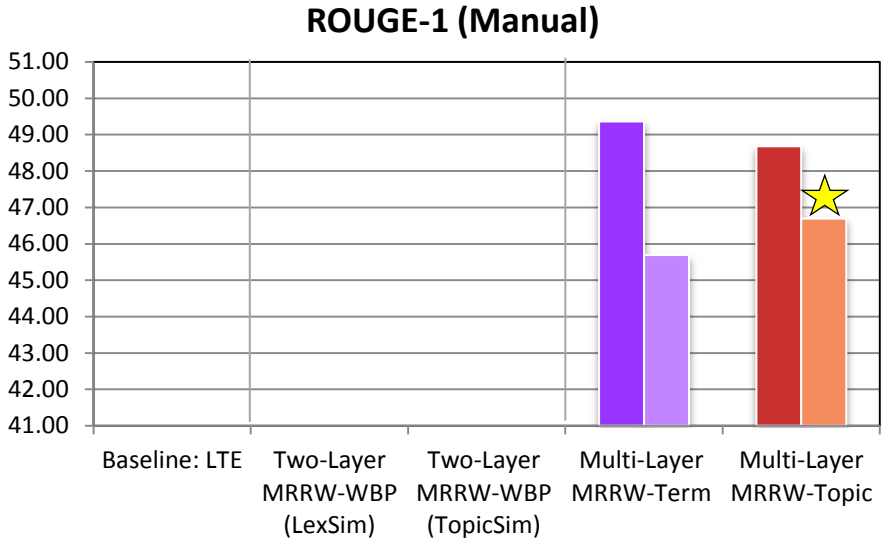


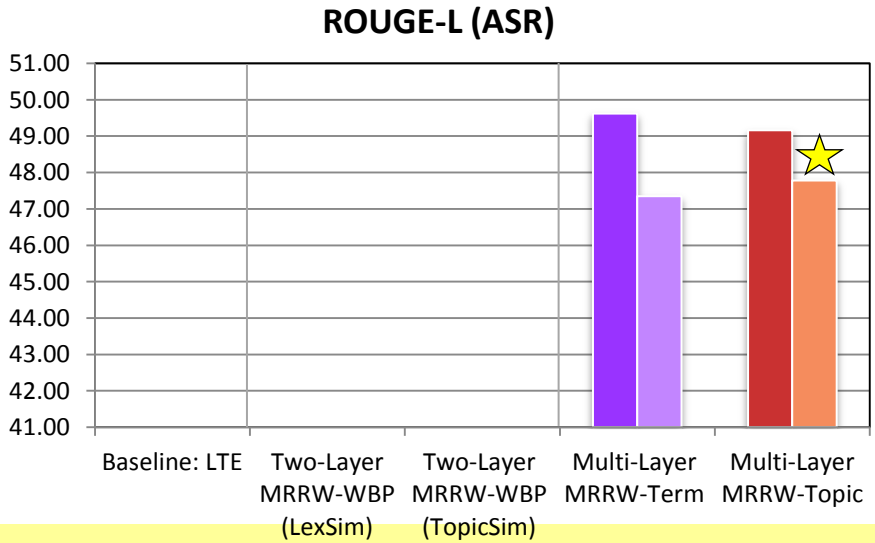
Topic models capture not lexically but conceptually similar information

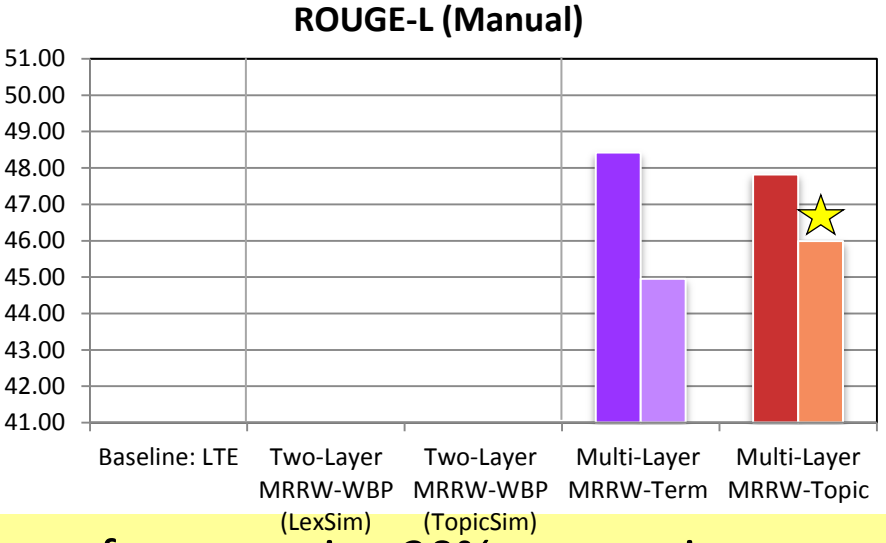
10% & 20%

Results – Different Types of Hidden Parameters





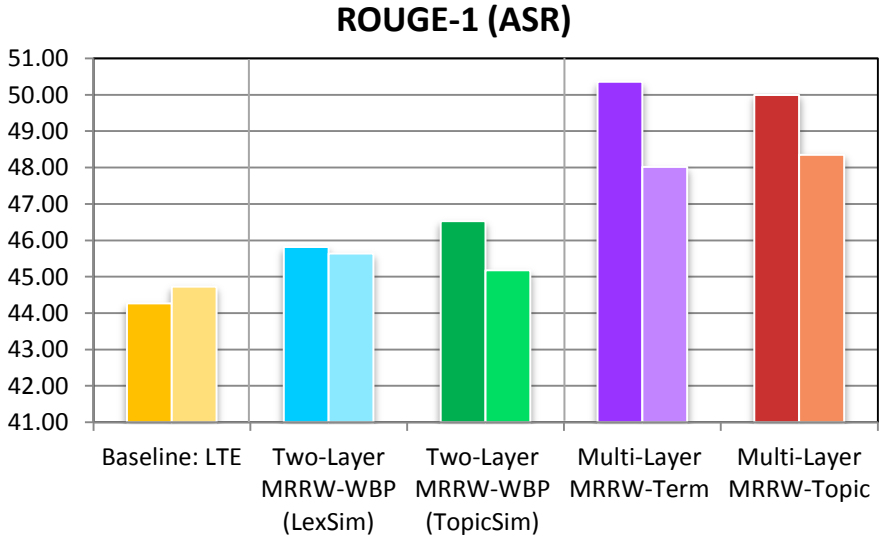


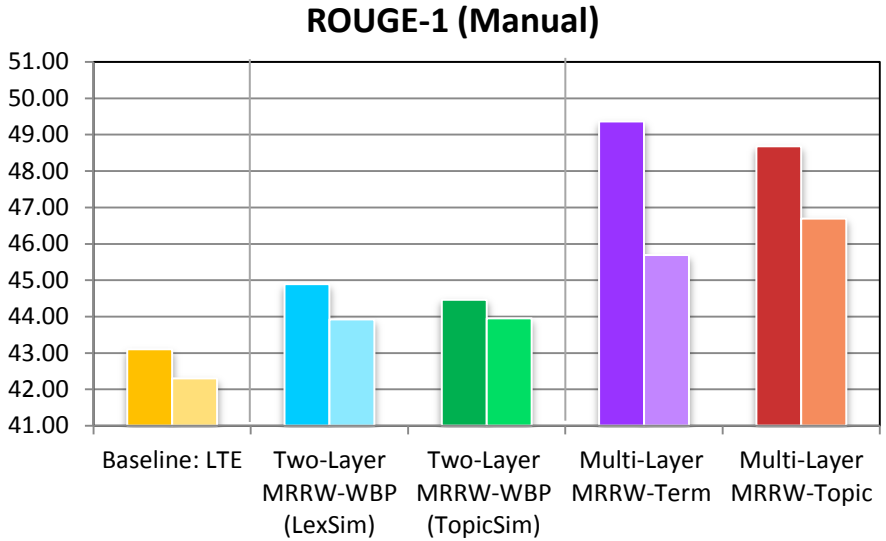


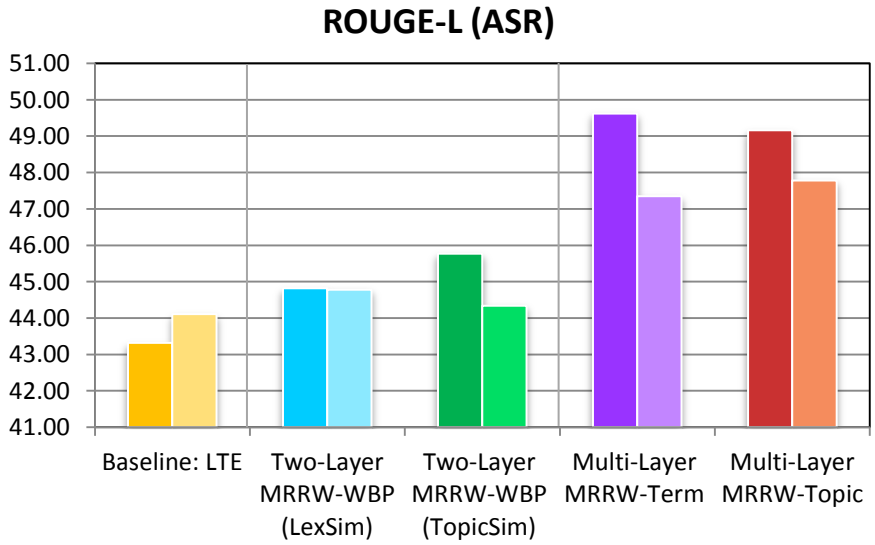
Using topics as the parameters is better for extracting 20% summaries

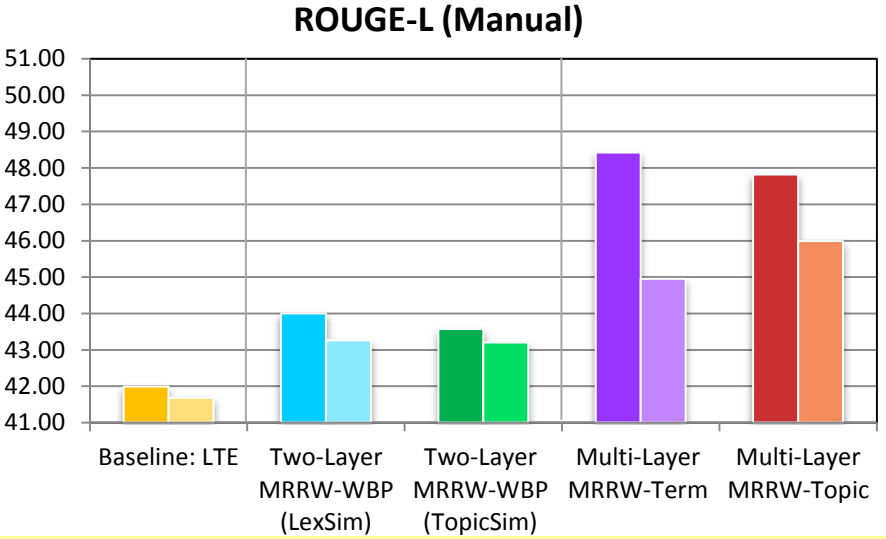
10% & 20%

Results – Complexity Reduction





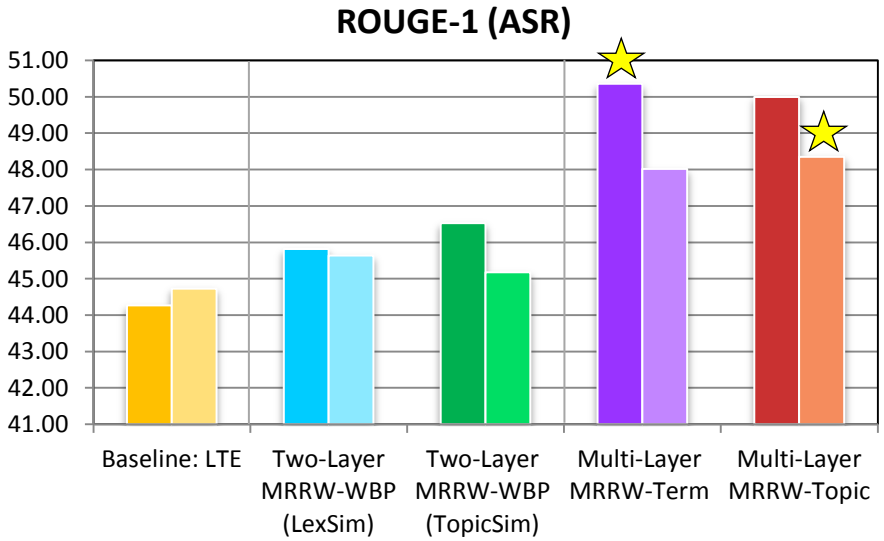


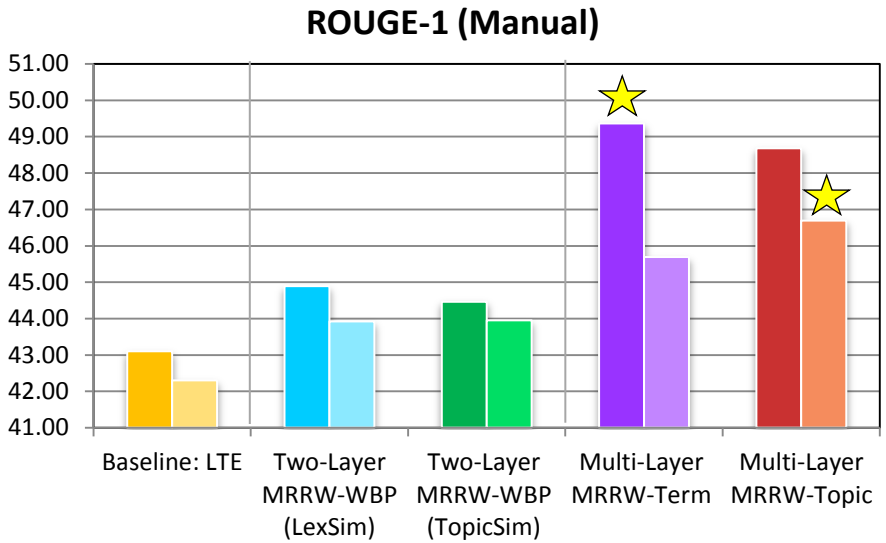


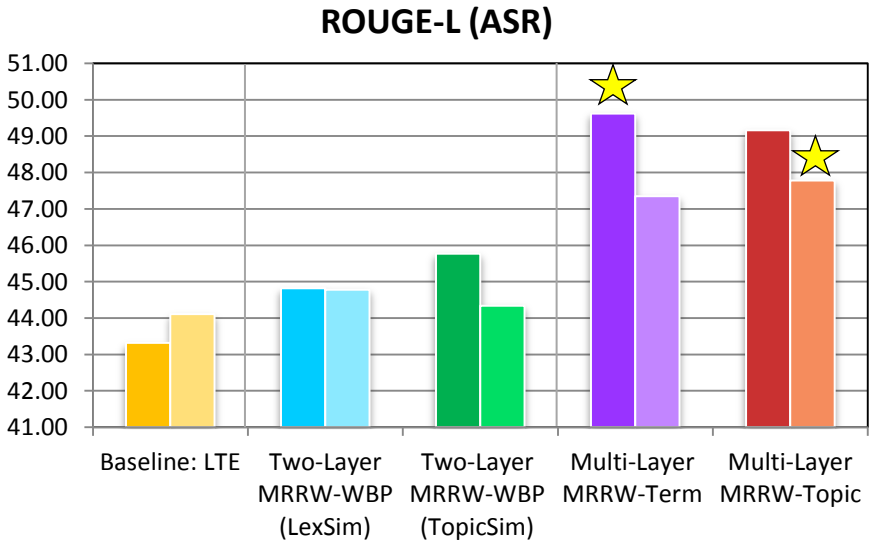
MRRW-Topic uses a smaller hidden-parameter-layer (constant) to reduce the complexity without hurting the performance a lot

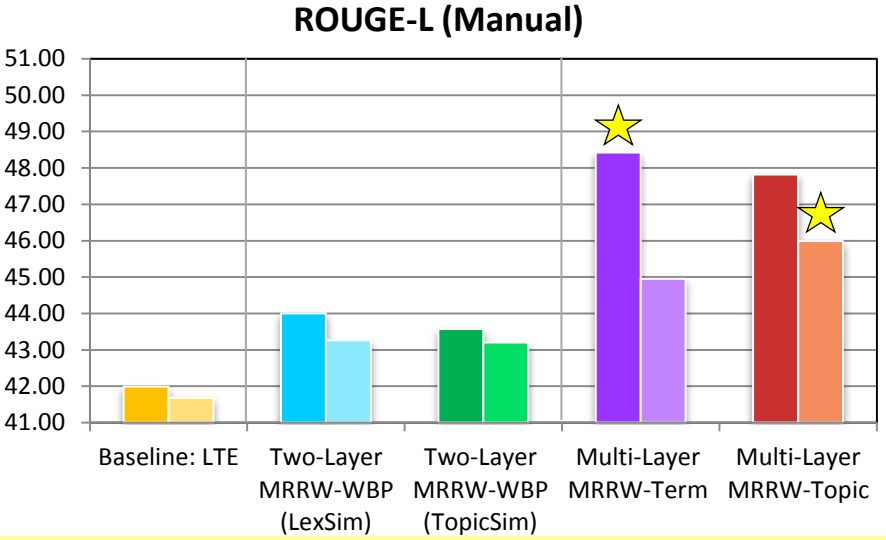
10% & 20%

Results – All



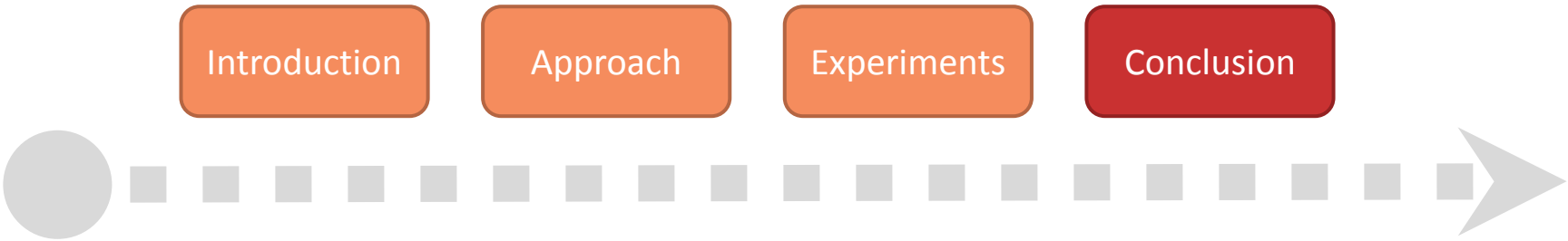






Proposed approaches achieve **13%** and **8%** relative improvement compared to baseline for 10% and 20% results respectively

Outline



- Multi-layer mutually reinforced random walk with hidden parameters can unsupervised model the utterances'/speakers' importance through shared hidden parameters (terms/topics) in the multi-layer graph
- The speaker information can be automatically included in the utterance importance by between-layer propagation and get further improvement
- Using latent topics as hidden parameters can reduce the complexity without hurting the performance

  → showing the practicality and effectiveness
- Our proposed approaches achieve more than 8% relative improvement compared to the baseline

Thanks for your attention! ☺

Q & A