Zero-Shot Learning of Intent Embeddings for Expansion by Convolutional Deep Structured Semantic Models

Carnegie Mellon

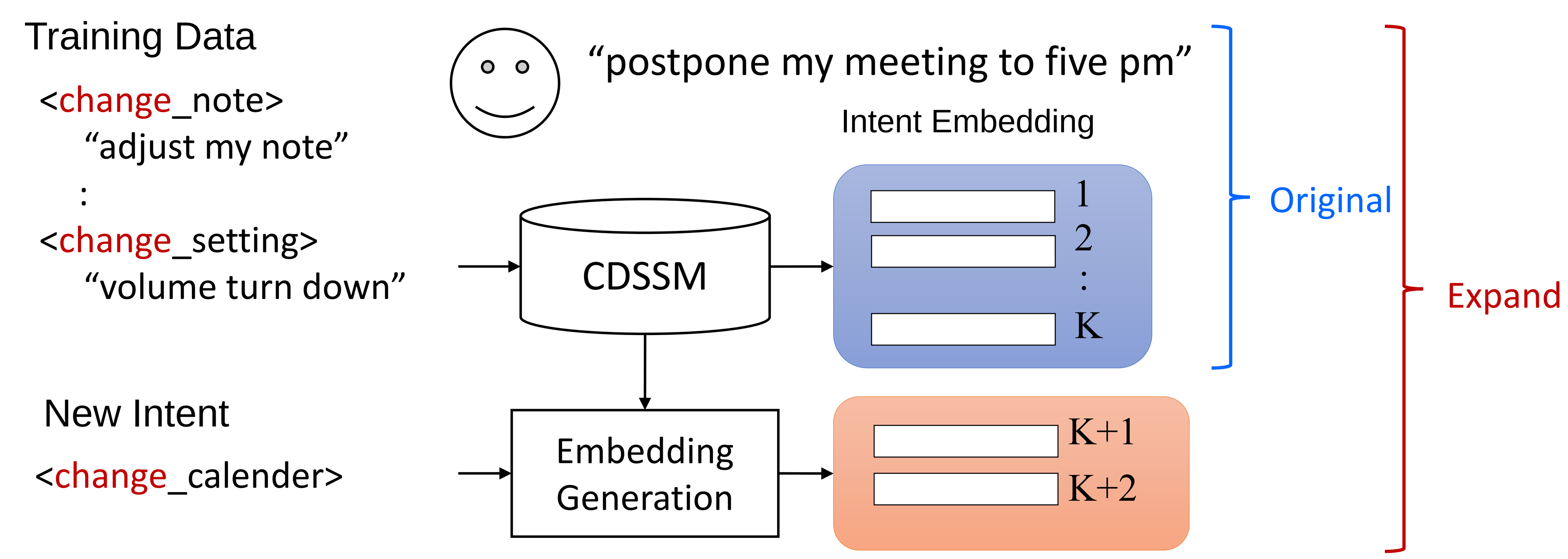
Yun-Nung (Vivian) Chen, Dilek Hakkani-Tür, and Xiaodong He

Microsoft Research

Summary

- Motivation: ***Domain Constraint & Inflexible Intent Schema***
  - Intents are usually predefined and inflexible to expand and transfer across domains, where re-designing intent semantic schemes requires human annotation and model re-training.
- Approach: ***Learning Intent Embedding***
  - Applying CDSSM to learn high-level semantic representations to bridge the semantic relation across domains for intent expansion (e.g. "find movie" and "find weather" belong to different domains, but they share the semantics about "find".)
- Result
  - CDSSM is capable of performing zero-shot learning effectively, e.g. generating embeddings of previously unseen intents, and therefore expand to new intents without re-training, and outperforms other semantic embeddings.

1. Framework



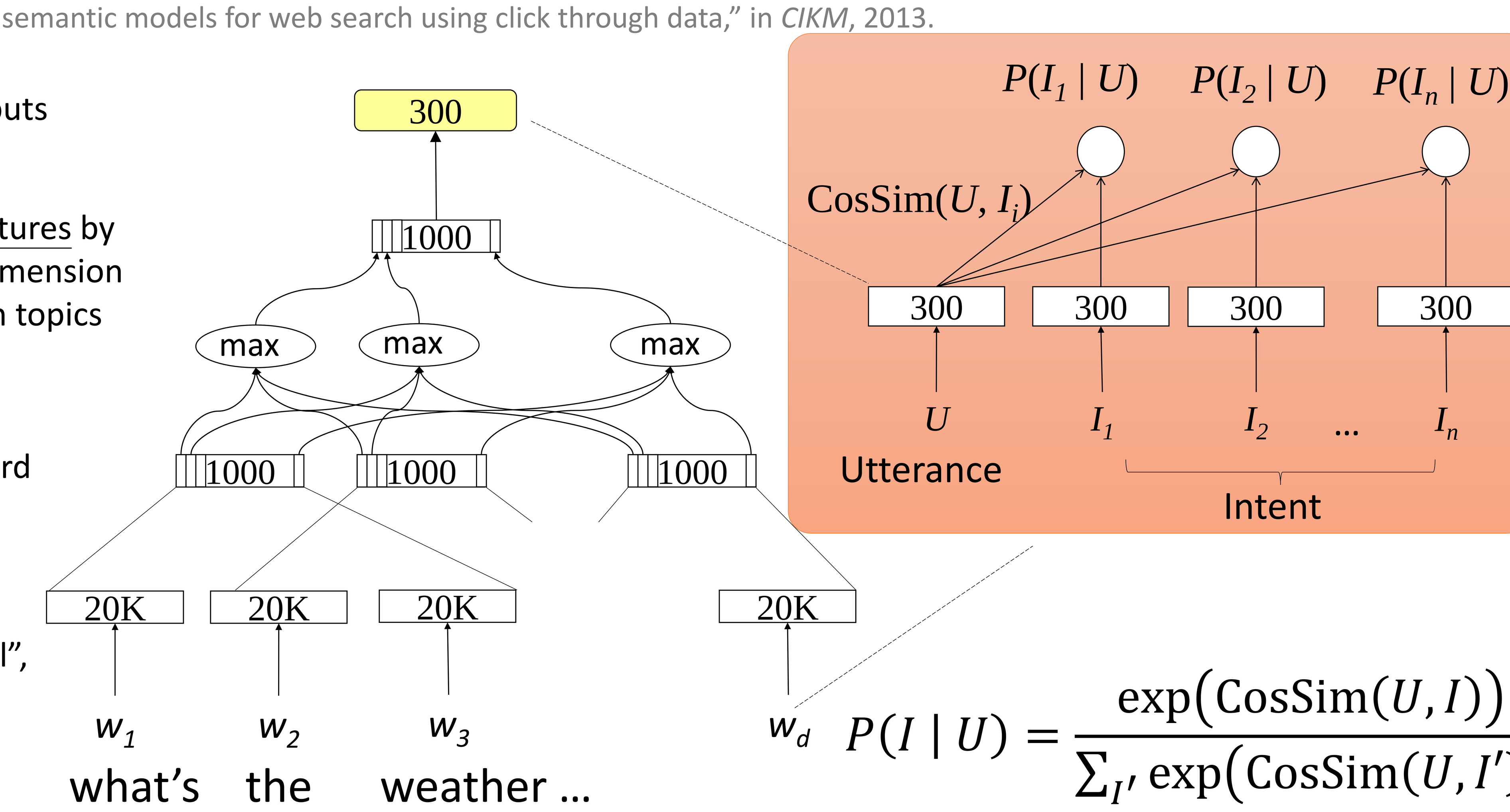
2. Convolutional Deep Structured Semantic Models (CDSSM)

Model Architecture

Shen et al., "A latent semantic model with convolutional-pooling structure for information retrieval," in *CIKM*, 2014.

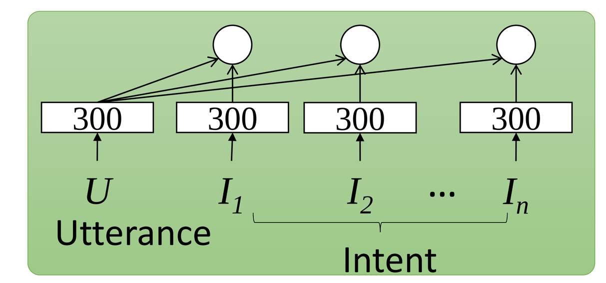
Huang et al., "Learning deep structured semantic models for web search using click through data," in *CIKM*, 2013.

- Semantic Layer: y — feed-forward neural network layers outputs the final non-linear semantic features
  Projection Matrix: W_s
- Max Pooling Layer: l_m — only retain the <u>most prominent local features</u> by applying the max operation over each dimension of l_c to keep the max activation of hidden topics across the whole word sequence
  Max Pooling Operation
- Convolutional Layer: l_c — contextual features c_i for each target word
  Convolution Matrix: W_c — $l_{ci} = \tanh(W_c^T c_i)$
- Word Hashing Layer: l_h — one-hot word vector → tri-letter vector (e.g. "email" → "#em", "ema", "mai", "ail", "il#")
  Word Hashing Matrix: W_h
- Word Sequence: x — user utterance / intent



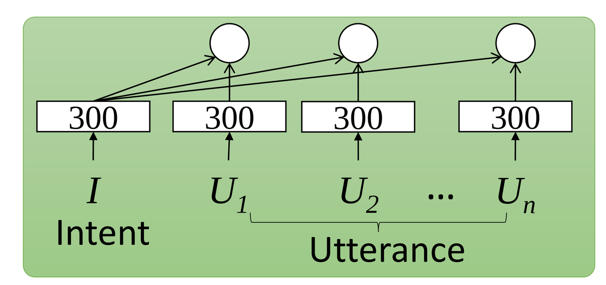
$$P(I \mid U) = \frac{\exp(\mathrm{CosSim}(U, I))}{\sum_{I'} \exp(\mathrm{CosSim}(U, I'))}$$

Training Procedure



→ Predictive Model

$$\Lambda(\theta_1) = \log \prod_{(U, I^+)} P(I^+ \mid U)$$



→ Generative Model

$$\Lambda(\theta_2) = \log \prod_{(I, U^+)} P(U^+ \mid I)$$

Intent Prediction by Bidirectional Estimation

- incorporate the effectiveness of predictive and generative models

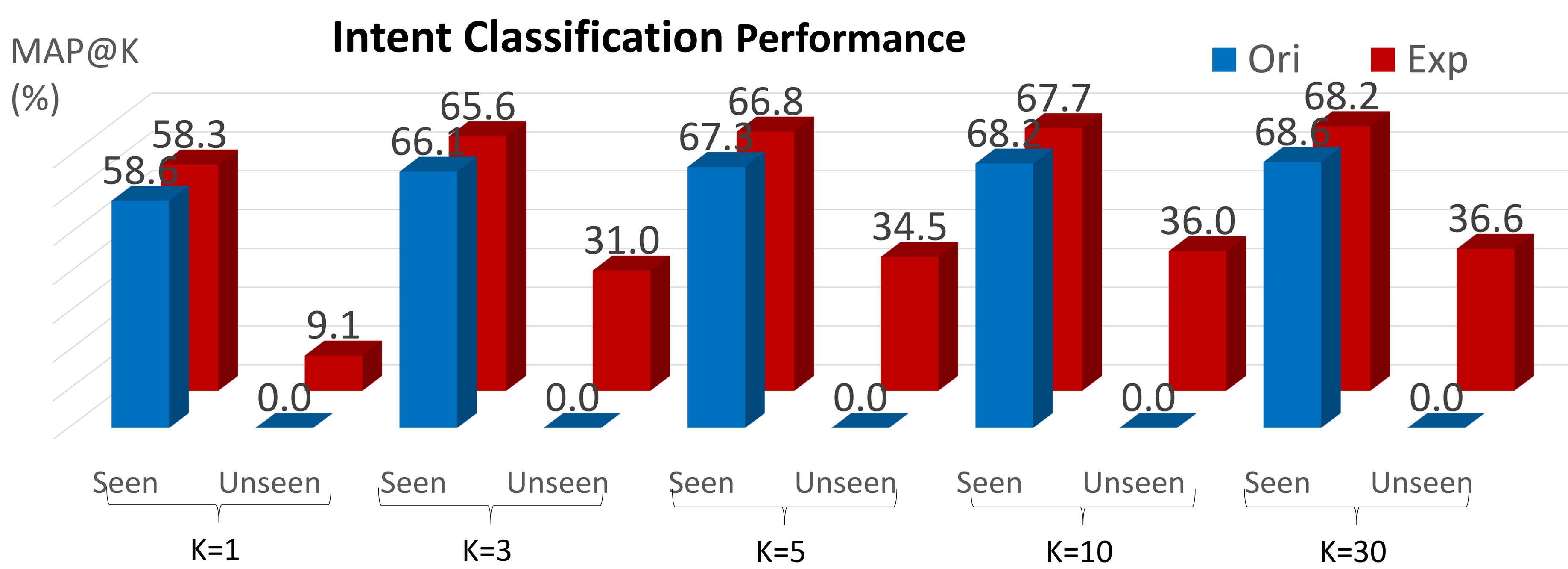
$$S_{\mathrm{Bi}}(U, I) = \gamma \underbrace{S_{\mathrm{P}}(U, I)}_{\text{Predictive Model}} + (1 - \gamma) \underbrace{S_{\mathrm{G}}(U, I)}_{\text{Generative Model}}$$

> The trained model is able to generate representations given word sequences without model re-training, so the intent embeddings can generalize to different domains.

3. Experiments

- Dataset: collected via the Microsoft Cortana (> 100 intents)
  - Segmented into seen and unseen intents
    - Unseen: randomly chose 7 intents with different verbs; ~100K utterances
    - Seen: ~1M annotated utterances (2/3 for training CDSSM, 1/3 for testing)
- Intent Prediction
  - For each utterance vector, the semantic similarity can be estimated using vectors for both seen and unseen intents.
  - The unseen intent vectors can be generated from CDSSM by feeding the tri-letter vectors of the new intent as input <u>without model re-training</u>.
- Evaluation Metrics: Mean average precision at K (MAP@K)

Intent Classification Performance



> The expanded models <u>consider new intents without training samples</u>, and produces similar but slightly worse than original models for seen intents due to higher uncertainty from more intent candidates.
> For unseen intents, expanded models are able to capture the correct intents and achieve higher than 30% of MAP when K ≥ 3, which indicates the encouraging performance when considering more than 100 intents.

- Effectiveness of Bidirectional Estimation

CDSSM-Expand	Seen Intents				Unseen Intents			
	K=1	K=3	K=5	K=10	K=1	K=3	K=5	K=10
Predictive	**58.9**	**65.9**	**67.1**	**67.9**	5.2	18.7	23.4	26.1
Generative	44.7	52.0	53.5	54.6	6.7	23.2	26.5	28.7
Bidirectional	58.3	65.6	66.8	67.7	**9.1**	**31.0**	**34.5**	**36.0**

> Although the predictive model performs better for seen intents, the bidirectional estimation is more robust to unseen intents, which is crucial to intent expansion.

- Effectiveness of Seen Intent Classification

Approach		ACC
Baseline	Paragraph Vector (doc2vec)	45.3
Proposed	CDSSM-Expand: Predictive	**58.9**
	CDSSM-Expand: Generative	44.7
	CDSSM-Expand: Bidirectional	**58.3**

> Outperform other embeddings for intent classification, including the state-of-the-art doc2vec (Le and Mikolov, 2014).

Conclusion

- A convolutional deep structured semantic model (CDSSM) is applied to perform zero-shot learning of intent embeddings to bridge the semantic relation across domains.
- The experiments of intent expansion show that CDSSM can
  - capture the semantics borrowed from other domains and can be used to flexibly expand the intents through high-level representations
  - carrying the crucial high-level semantics and can be applied to different domains for easy adaptation and extension
  - generate more flexible intent embeddings without training samples and model re-training, removing the domain constraint in dialogue systems for practical usage.