Ashutosh Modi and Ivan Titov

INCLUDING NEURAL MODELS OF SCRIPT KNOWLEDGE

Selected Topics in Semantic and Discourse
Professor Dr. Pinkal
Margarita Chikobava

Overview

- Motivation
- Model
  - Event representation
  - Learning to Order
- Experiments
  - Learning from Crowdsourced Data
  - Learning from Natural Text
- Results and Discussion

How was it done before?

- Manual
- Scripts, graphs and chains
- Human interpretable
- Not natural text

How is it done now?

- Statistical
- Without knowledge representation as a graph
- With distributed representation of event realizations
- Predicates and their arguments
- Capture relevant features (transitivity, particles…)

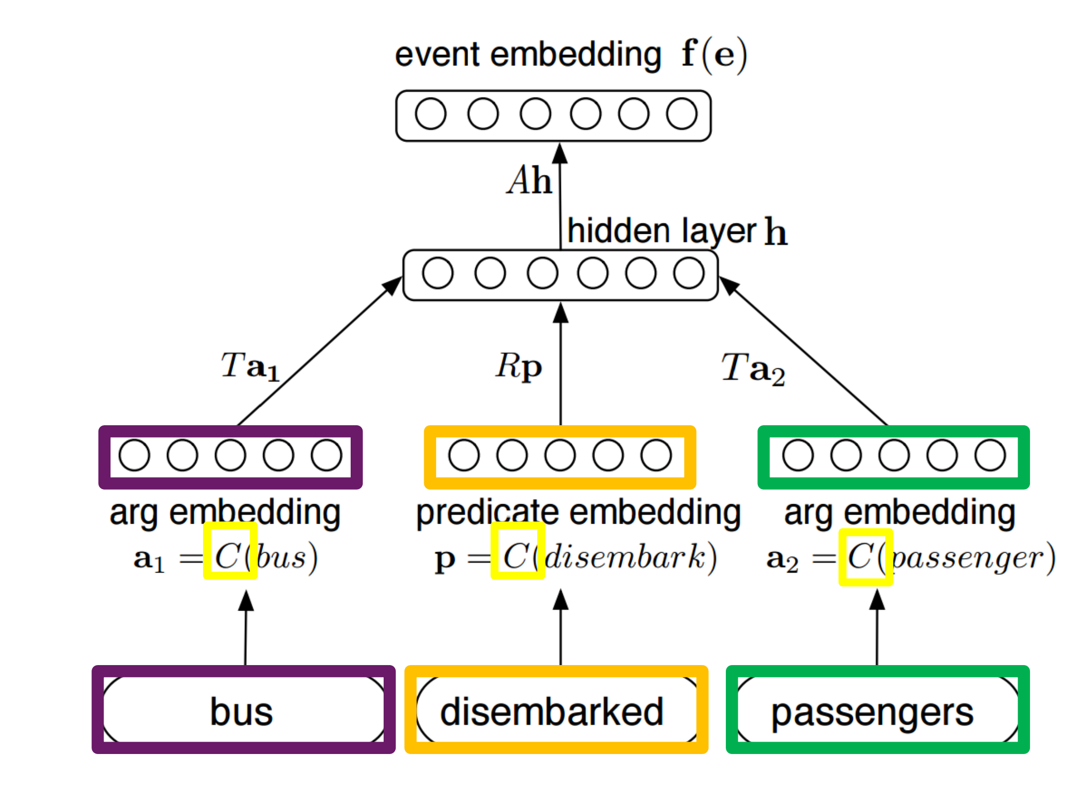
- Task: Ordering? Two event mentions. Paraphrases of the same event?

Model. Event Representation

- Simple compositional model:

event = predicate + argument

- Each word c_i in the vocabulary is mapped to a real vector



Model. Event Representation

- Simple compositional model: event = predicate + argument
- Each word c_i in the vocabulary is mapped to a real vector

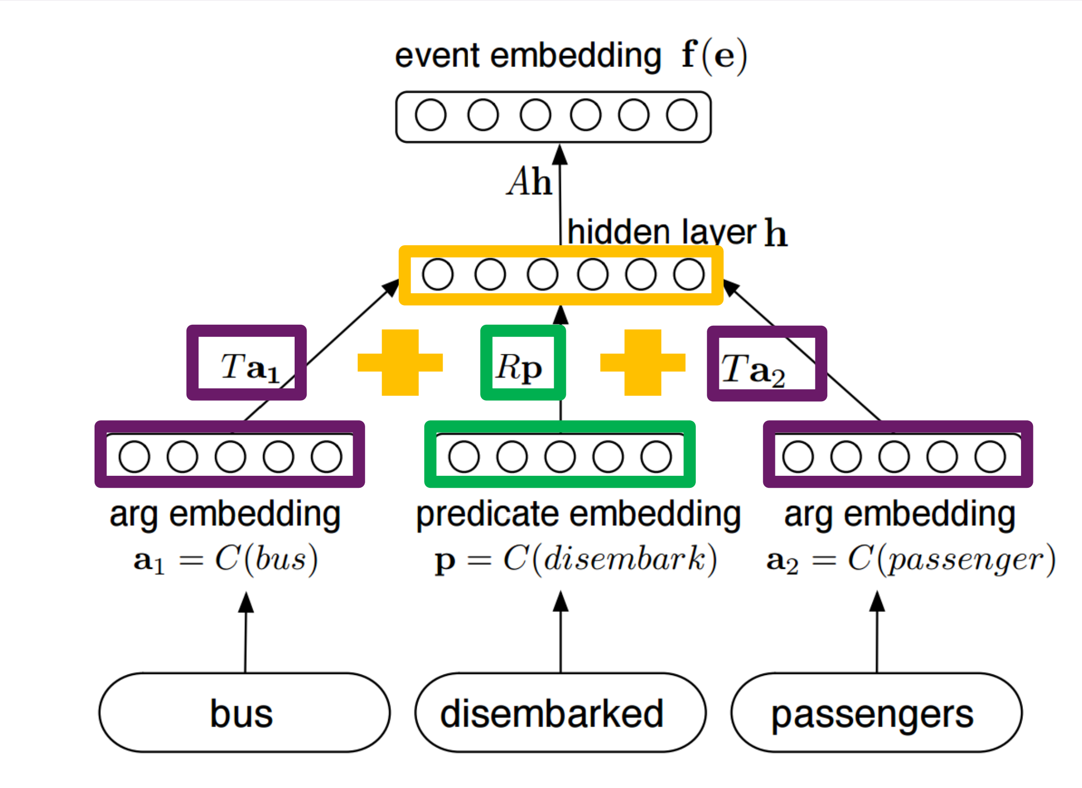
- Hidden layer =

linearly transformed predicate

and argument

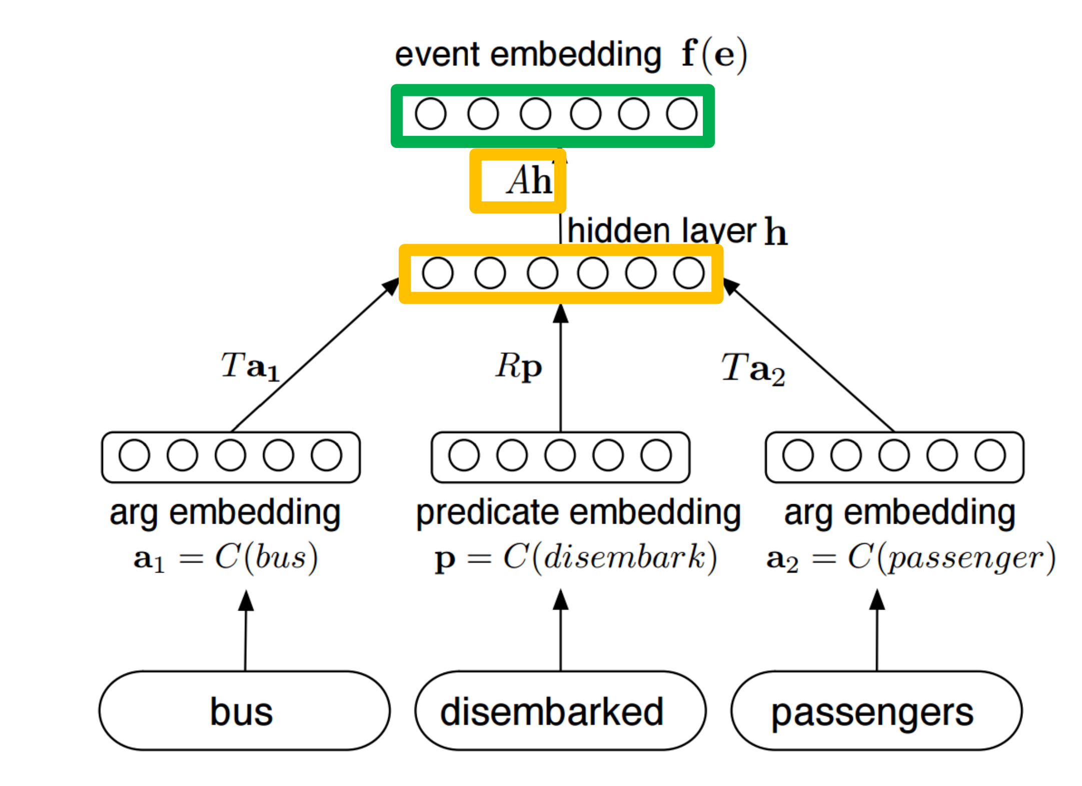
summing

+ logistic sigmoid function



Model. Event Representation

- Simple compositional model: event = predicate + argument
- Each word c_i in the vocabulary is mapped to a real vector
- Hidden layer = summing linearly transformed predicate and argument + logistic sigmoid function



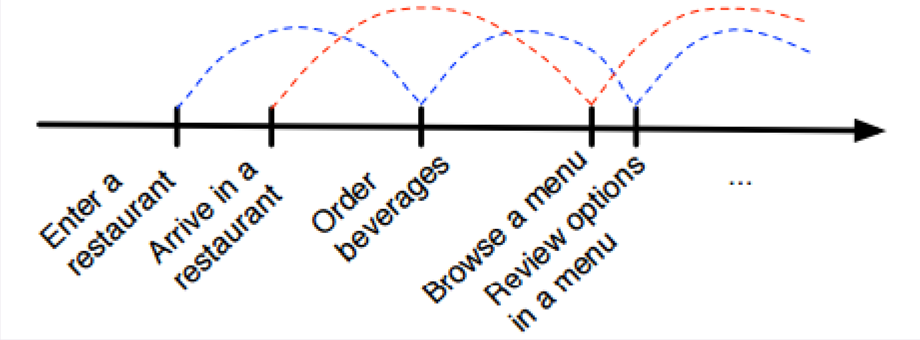
- event representation =

linear transformation

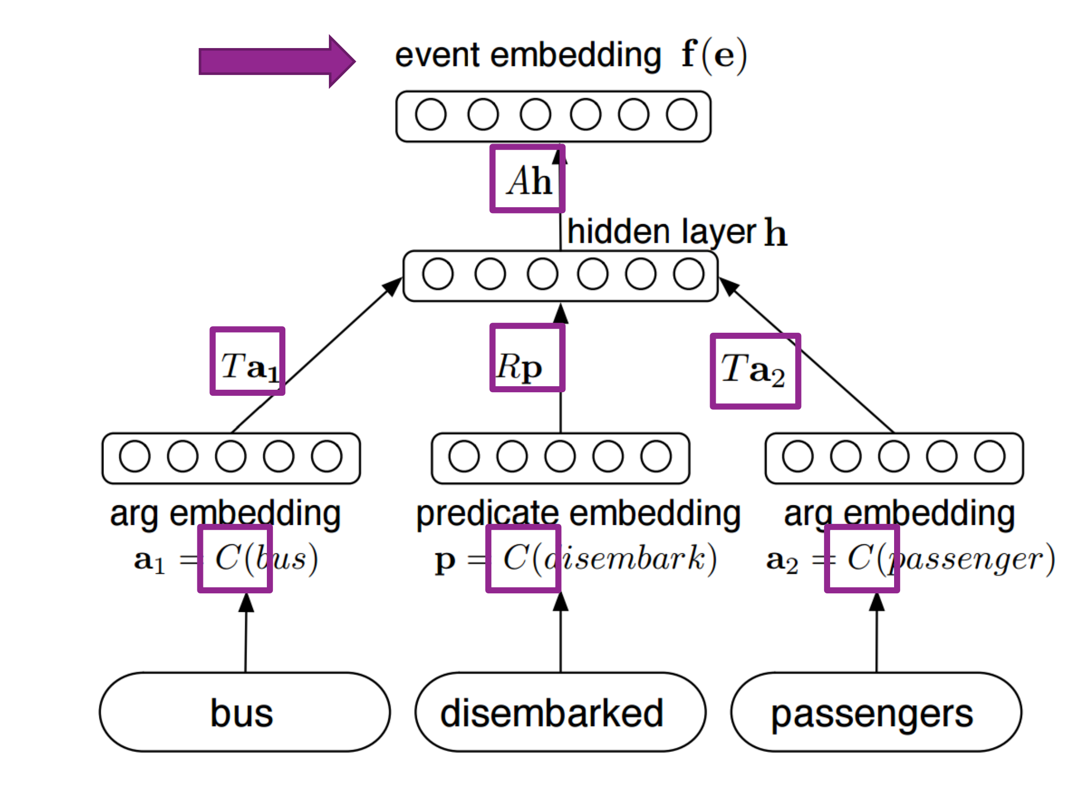
+ sigmoid function

Model. Learning to Order

- Linear ranker w ?
- Ranking score



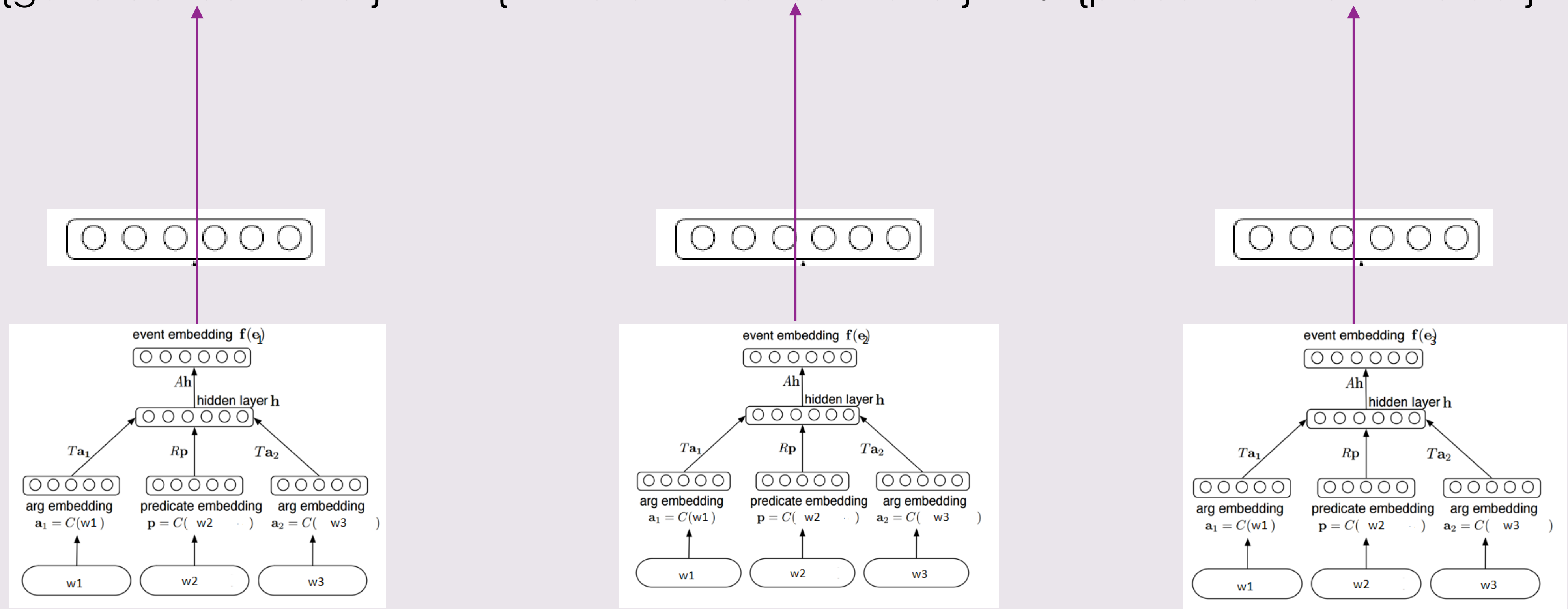
- Representation parameters



Linear ranker

1. {go to coffee maker} → 2. {fill water in coffee maker} → 3. {place the filter in holder}

ranker



Experiments. CD

- Regneri: event sequence descriptions(ESD)
- Scenario ≈ 30 ESDs

- Challenging learning task:
  - Limited amount of training data
  - Variability
  - optionality

- 4 scenarios to choose model parameters
- 2000 epochs

{go to coffee maker} → {fill water in coffee maker} → {place the filter in holder} → {place coffee in filter} → {place holder in coffee maker} → {turn on coffee maker}

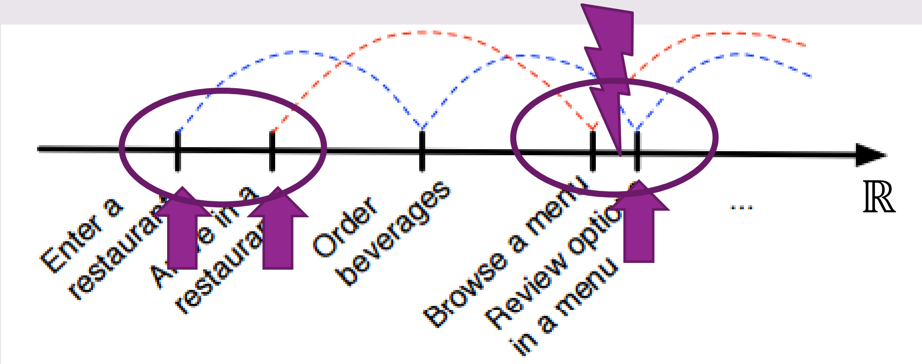
CD: Results and discussion

	Precision (%)					Recall (%)					F1 (%)				
	BL	EE_{verb}	MSA	BS	EE	BL	EE_{verb}	MSA	BS	EE	BL	EE_{verb}	MSA	BS	EE
Bus	70.1	81.9	80.0	76.0	85.1	71.3	75.8	80.0	76.0	91.9	70.7	78.8	80.0	76.0	88.4
Coffee	70.1	73.7	70.0	68.0	69.5	72.6	75.1	78.0	57.0	71.0	71.3	74.4	74.0	62.0	70.2
Fastfood	69.9	81.0	53.0	97.0	90.0	65.1	79.1	81.0	65.0	87.9	67.4	80.0	64.0	78.0	88.9
Return	74.0	94.1	48.0	87.0	92.4	68.6	91.4	75.0	72.0	89.7	71.0	92.8	58.0	79.0	91.0
Iron	73.4	80.1	78.0	87.0	86.9	67.3	69.8	72.0	69.0	80.2	70.2	69.8	75.0	77.0	83.4
Microw.	72.6	79.2	47.0	91.0	82.9	63.4	62.8	83.0	74.0	90.3	67.7	70.0	60.0	82.0	86.4
Eggs	72.7	71.4	67.0	77.0	80.7	68.0	67.7	64.0	59.0	76.9	70.3	69.5	66.0	67.0	78.7
Shower	62.2	76.2	48.0	85.0	80.0	62.5	80.0	82.0	84.0	84.3	62.3	78.1	61.0	85.0	82.1
Phone	67.6	87.8	83.0	92.0	87.5	62.8	87.9	86.0	87.0	89.0	65.1	87.8	84.0	89.0	88.2
Vending	66.4	87.3	84.0	90.0	84.2	60.6	87.6	85.0	74.0	81.9	63.3	84.9	84.0	81.0	88.2
Average	69.9	81.3	65.8	85.0	**83.9**	66.2	77.2	78.6	71.7	**84.3**	68.0	79.1	70.6	77.6	**84.1**

Table 1: Results on the crowdsourced data for the verb-frequency baseline (BL), the verb-only embedding model (EE_{verb}), Regneri et al. (2010) (MSA), Frermann et al. (2014)(BS) and the full model (EE).

Paraphrasing

- Mapping event to the position on the time line
- Simplifying assumption: intervals do not overlap
- Constraint: same ESD → not the same interval
- Simple greedy algorithm
- Modification



Paraphrasing

- Not using external knowledge
- Not very robust to noise

Scenario	F1 (%)			
	APBL	MSA	BS	EE
Take bus	53.7	74.0	47.0	63.5
Make coffee	42.1	65.0	52.0	63.5
Order fastfood	37.0	59.0	80.0	62.6
Return food back	64.8	71.0	67.0	$\boxed{81.1}$
Iron clothes	43.3	67.0	60.0	56.7
Microwave cooking	43.2	75.0	82.0	57.8
Scrambled eggs	57.6	69.0	76.0	53.0
Take shower	42.1	78.0	67.0	55.7
Answer telephone	71.0	89.0	81.0	79.4
Vending machine	56.1	69.0	77.0	69.3
Average	51.1	**71.6**	68.9	64.5

Experiments. NT

- Gigaword corpus - news
- Explicit clues (then, after…)
- Ordered pairs
- Training

- Preprocessing
  - rule-based temporal classifier
  - No clues No semantic differences
  - No information about order in the text

	Accuracy (%)
BL	60.7
CJ08	60.1
EE_{verb}	75.9
EE	**83.5**

Summary

- Interesting? Statistical model for prototypical event ordering
- New? Without representation as a script
- Well done? Outperforms baselines

- Practical? Paraphrasing. Ordering

References

◦ Modi, A., & Titov, I. (2014). Inducing neural models of script knowledge. CoNLL-2014

Thank you for your attention!

- Questions?