

# Question Answering over Freebase with Multi-Column Convolutional Neural Networks



Li Dong<sup>1</sup>, Furu Wei<sup>2</sup>, Ming Zhou<sup>2</sup>, Ke Xu<sup>1</sup>

<sup>1</sup>SKLSDE, Beihang University, Beijing, China

<sup>2</sup>Microsoft Research, Beijing, China

# Question Answering over Freebase

- Freebase

- Large-scale knowledge base
- A rich resource to answer open-domain questions


**Question:**

*when did Avatar release in UK* →  → **Answer:** *2009-12-17*

- Challenge

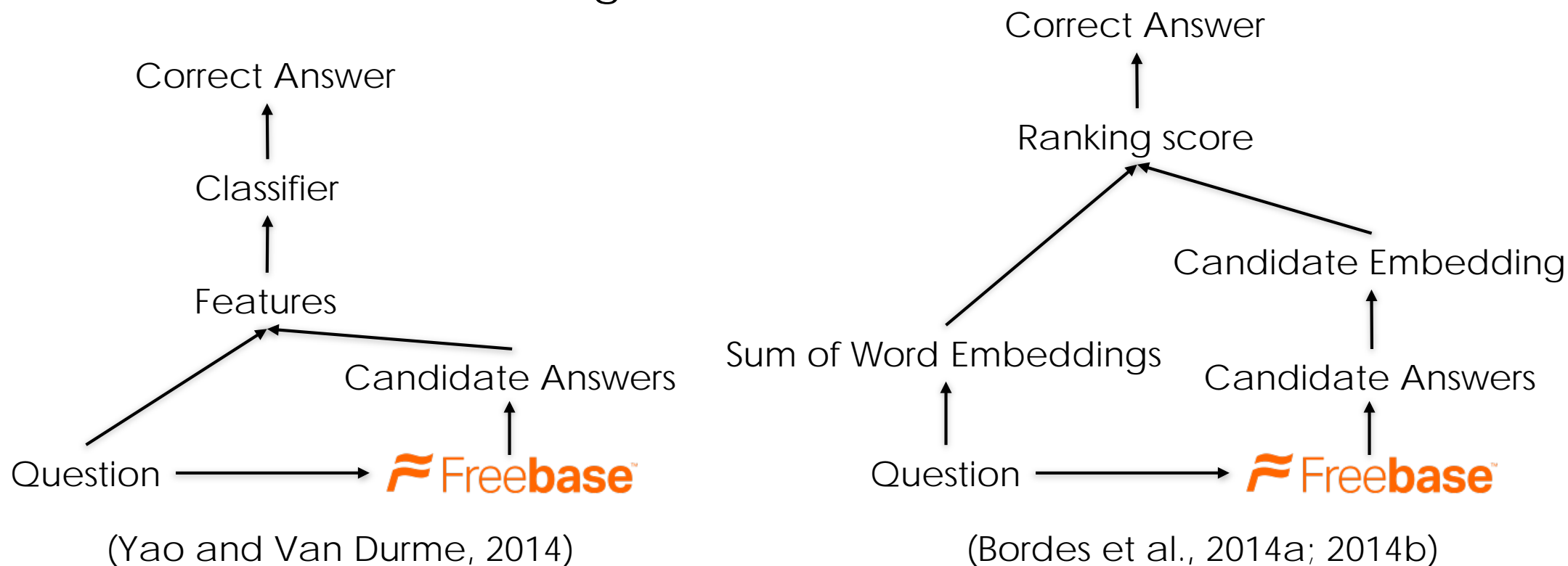
- natural language questions ~ structured semantics of Freebase
- How to bridge the gap?

# Mainstream Methods (1/2)

- Semantic parsing (Berant et al., 2013; Bao et al., 2014; etc.)
  - Question → Formal Meaning Representation → Structured Queries
  - →  Freebase → Answer
  - Example
    - Utterance: Which college did Obama go to
    - Logical form: (and (Type University) (Education BarackObama))
    - Denotation: Occidental College, Columbia University
  - Challenges
    - Huge search space
    - Lexical triggers

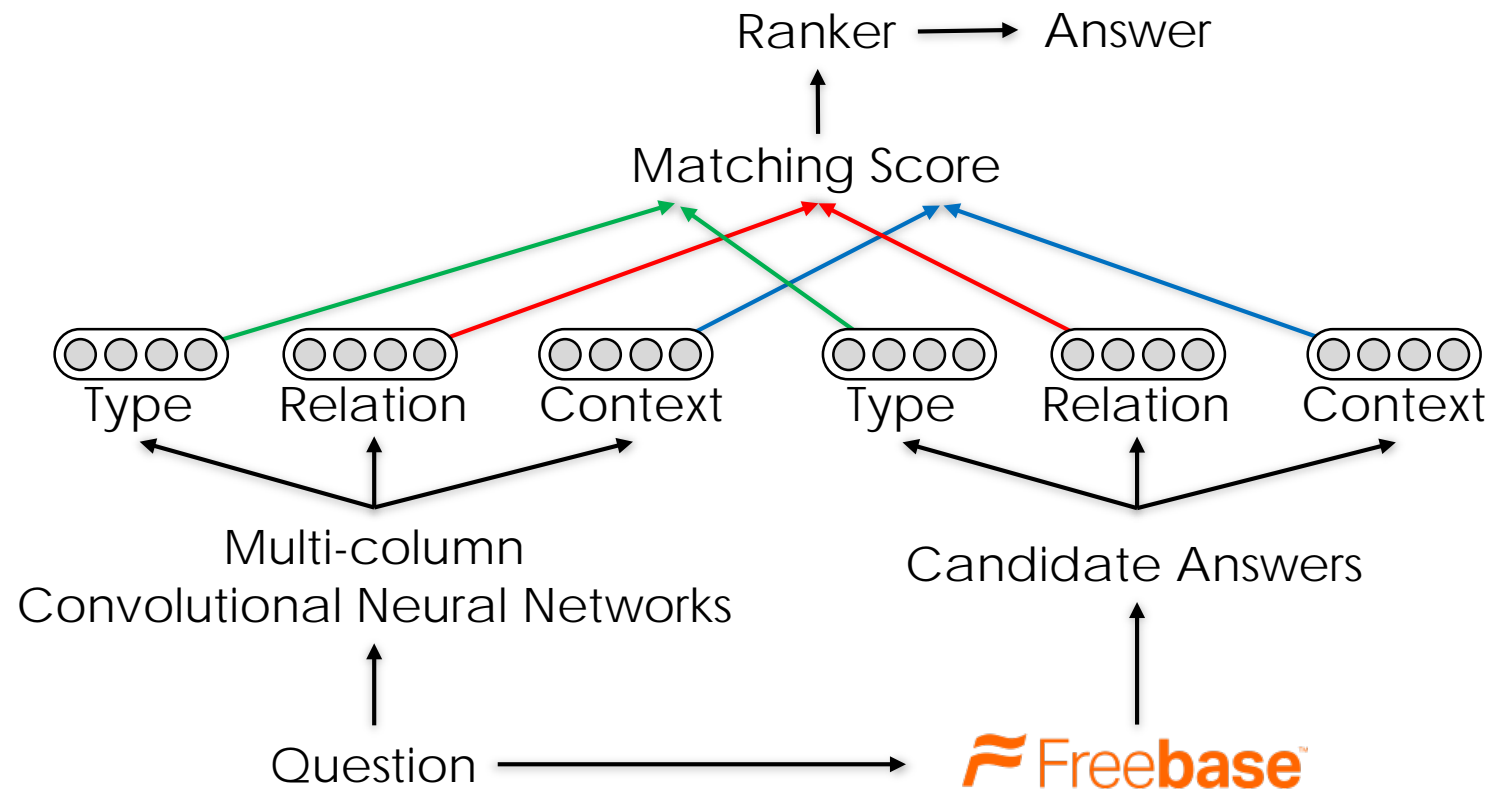
# Mainstream Methods (2/2)

- Information extraction over knowledge base
  - 1. Retrieve candidate answers from Freebase
  - 2. Extract features
  - 3. Classification / Ranking

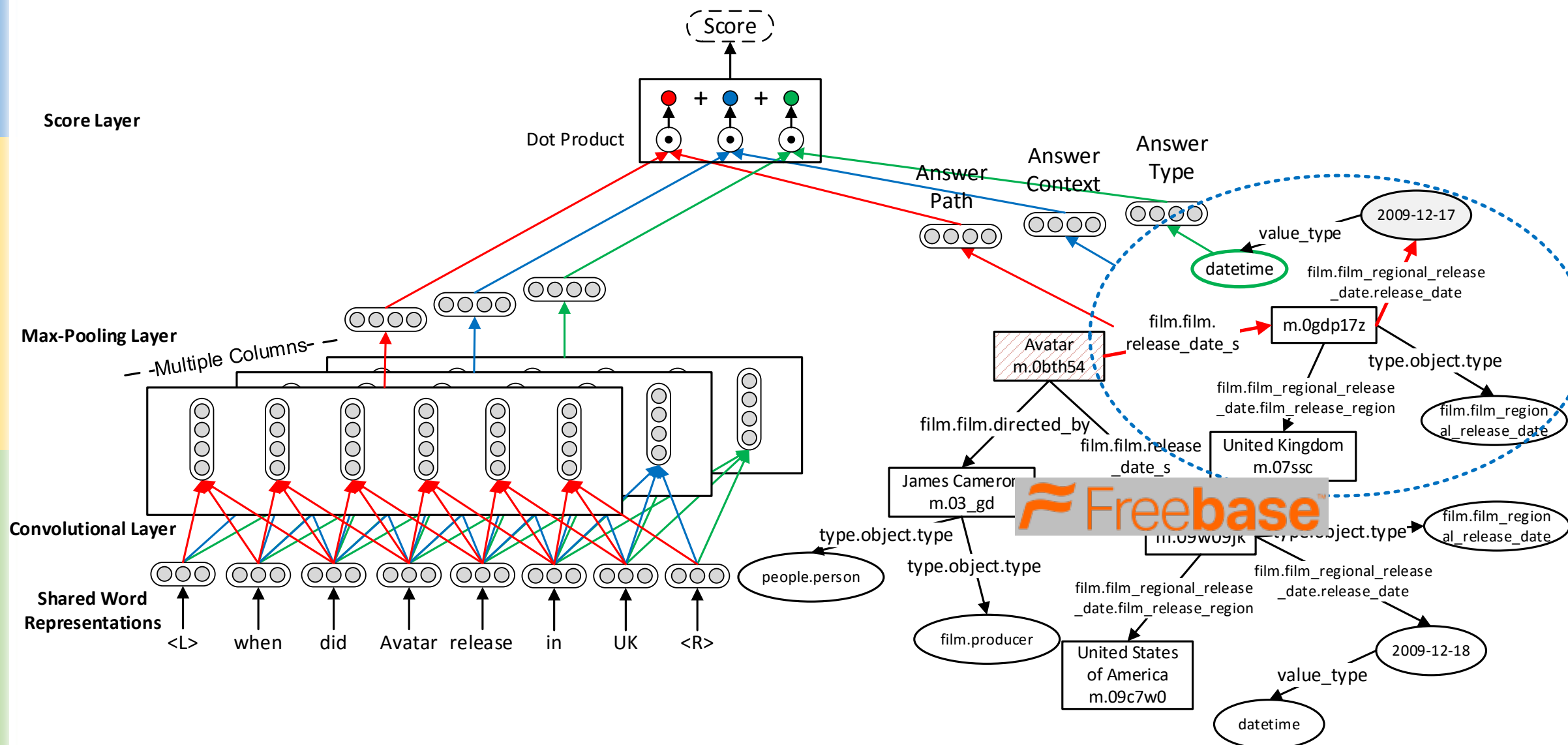


# Proposed Method

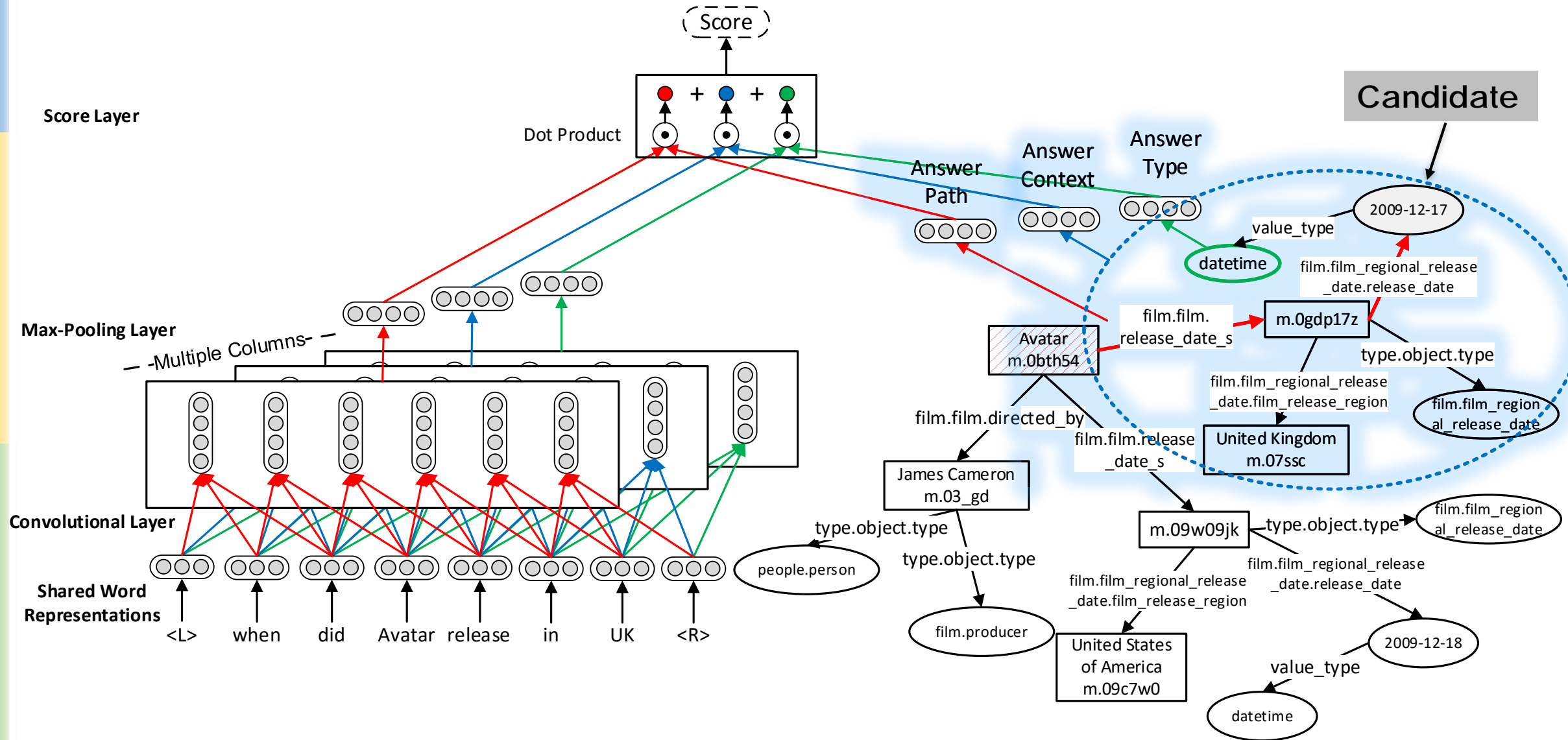
- Question answering -> Constraint matching
  - Answer type, answer path (relation), answer context
- Question understanding with convolutional neural networks



# Model Overview

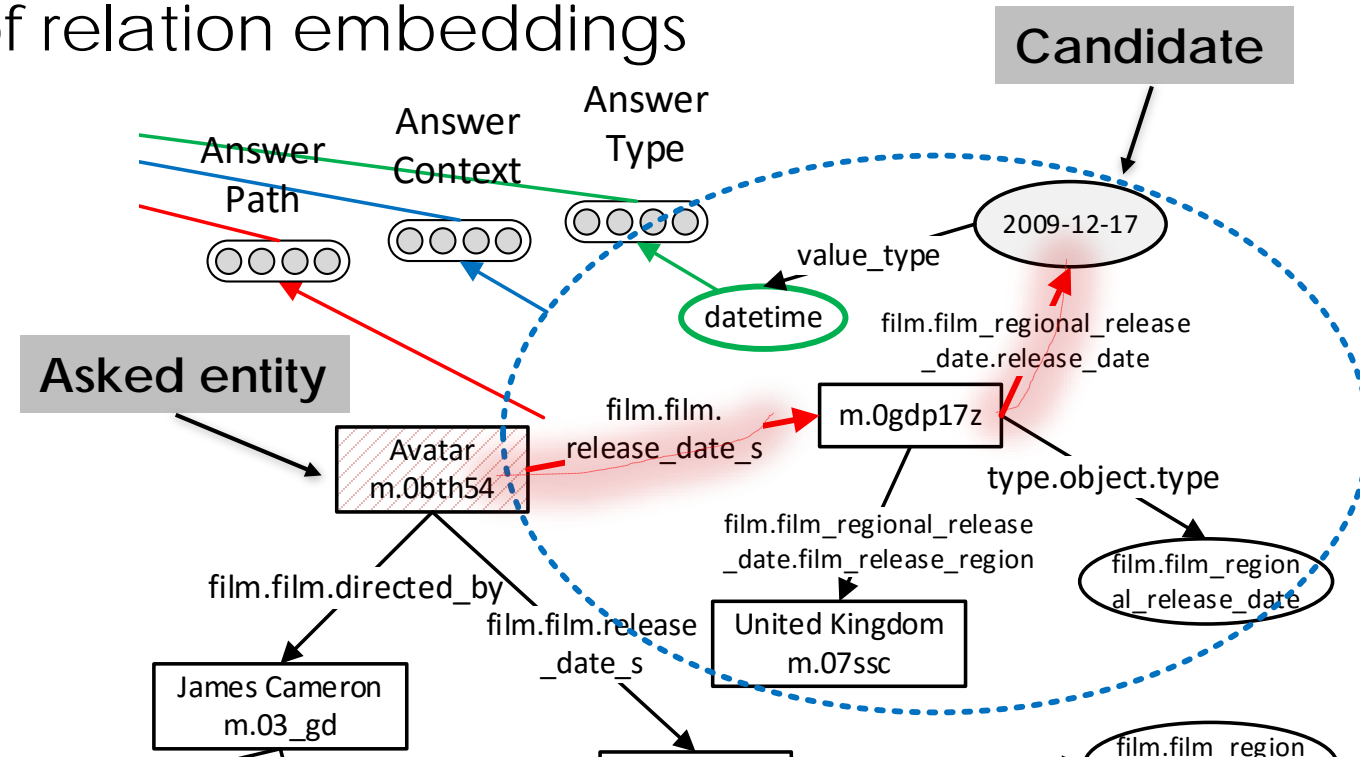


# Model Overview



# Embedding Candidate Answers

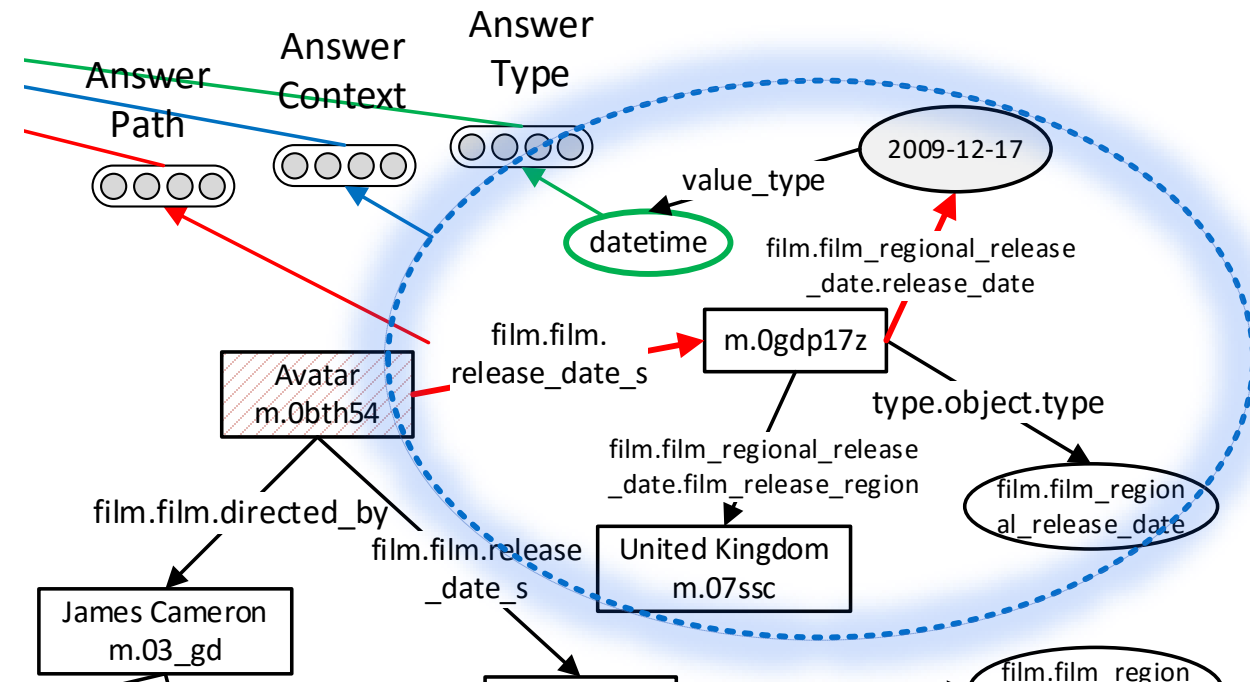
- Learn vector representations for candidate answers
  - (Bordes et al., 2014a; Bordes et al., 2014b)
- Answer path
  - relations between the candidate node and the entity asked in question
  - $avg(r_1, r_2, \dots, r_n)$ : average of relation embeddings
- Answer context
- Answer type





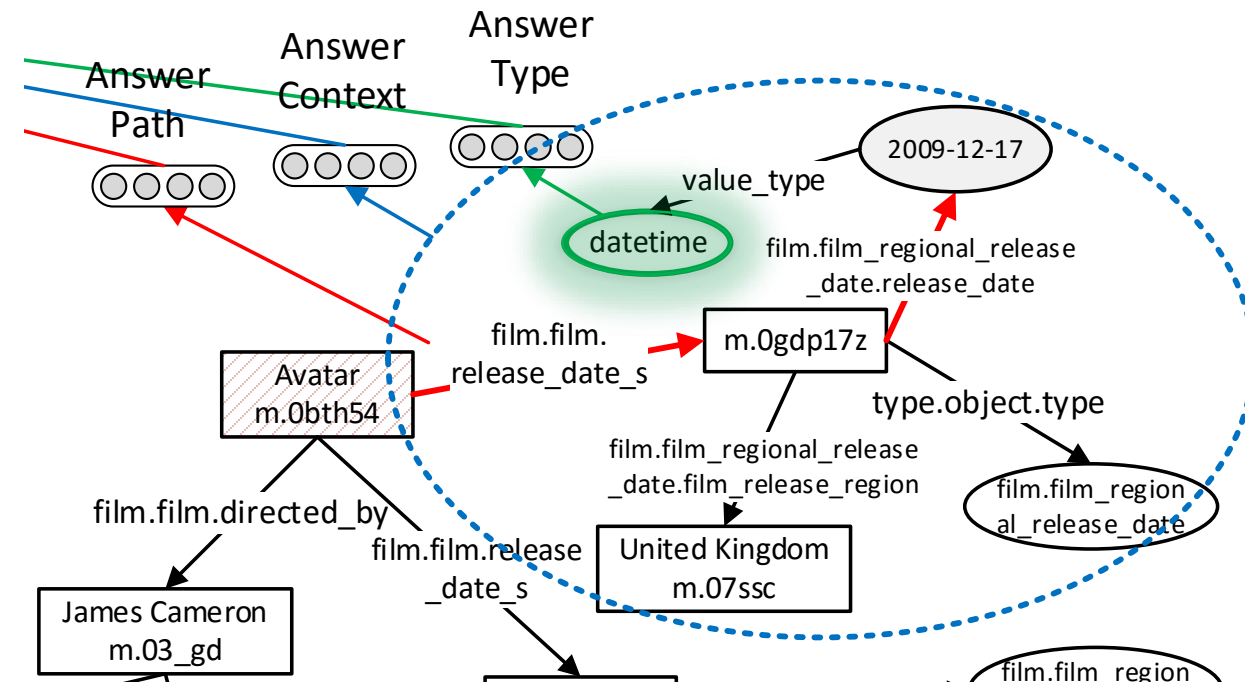
# Embedding Candidate Answers

- Learn vector representations for candidate answers
  - (Bordes et al., 2014a; Bordes et al., 2014b)
- Answer context
  - 1-hop entities and relations connected to the answer path
  - $avg(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n)$ : average of context entity and relation embeddings
- Answer path
- Answer type

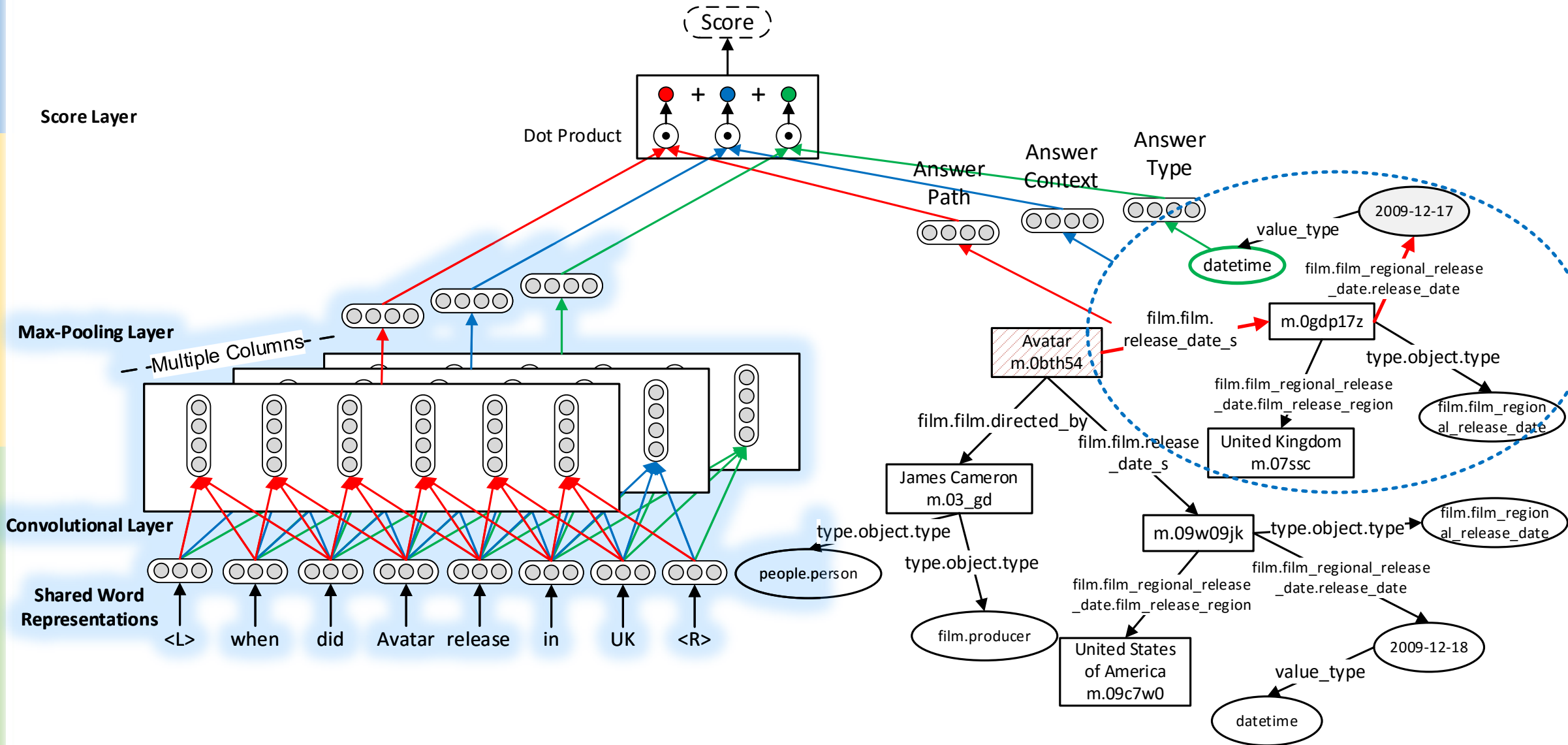


# Embedding Candidate Answers

- Learn vector representations for candidate answers
  - (Bordes et al., 2014a; Bordes et al., 2014b)
- Answer type
  - common.topic.notable\_types, value type (e.g., float, string, datetime)
  - $avg(t_1, t_2, \dots, t_n)$ : average of type embeddings
- Answer path
- Answer context

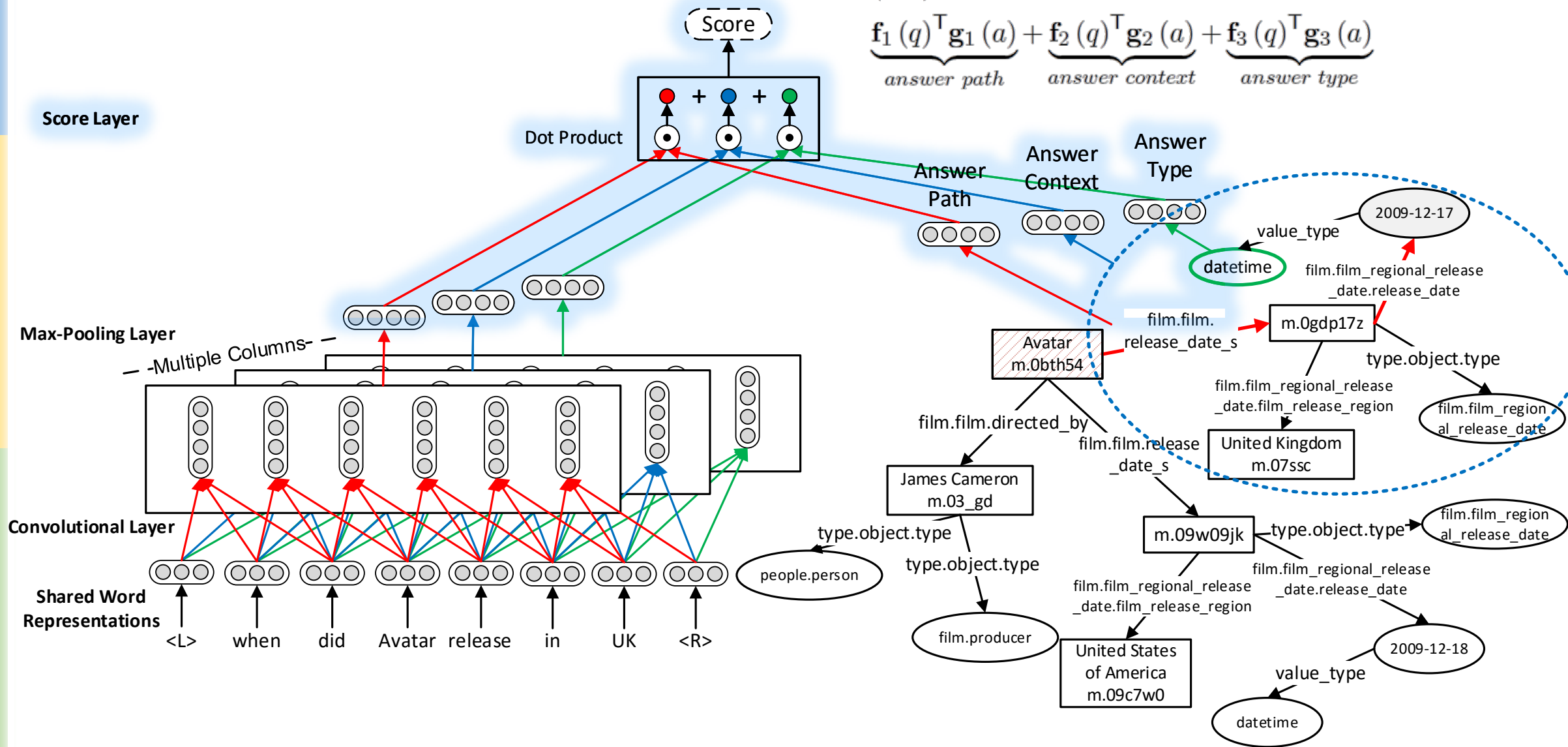


# Model Overview



# Model Overview

$$S(q, a) = \underbrace{\mathbf{f}_1(q)^\top \mathbf{g}_1(a)}_{\text{answer path}} + \underbrace{\mathbf{f}_2(q)^\top \mathbf{g}_2(a)}_{\text{answer context}} + \underbrace{\mathbf{f}_3(q)^\top \mathbf{g}_3(a)}_{\text{answer type}}$$



# Model Training

- Negative instance  $a'$  is randomly sampled from the set of candidate answers
- Hinge loss for  $(q, a)$  and  $(q, a')$

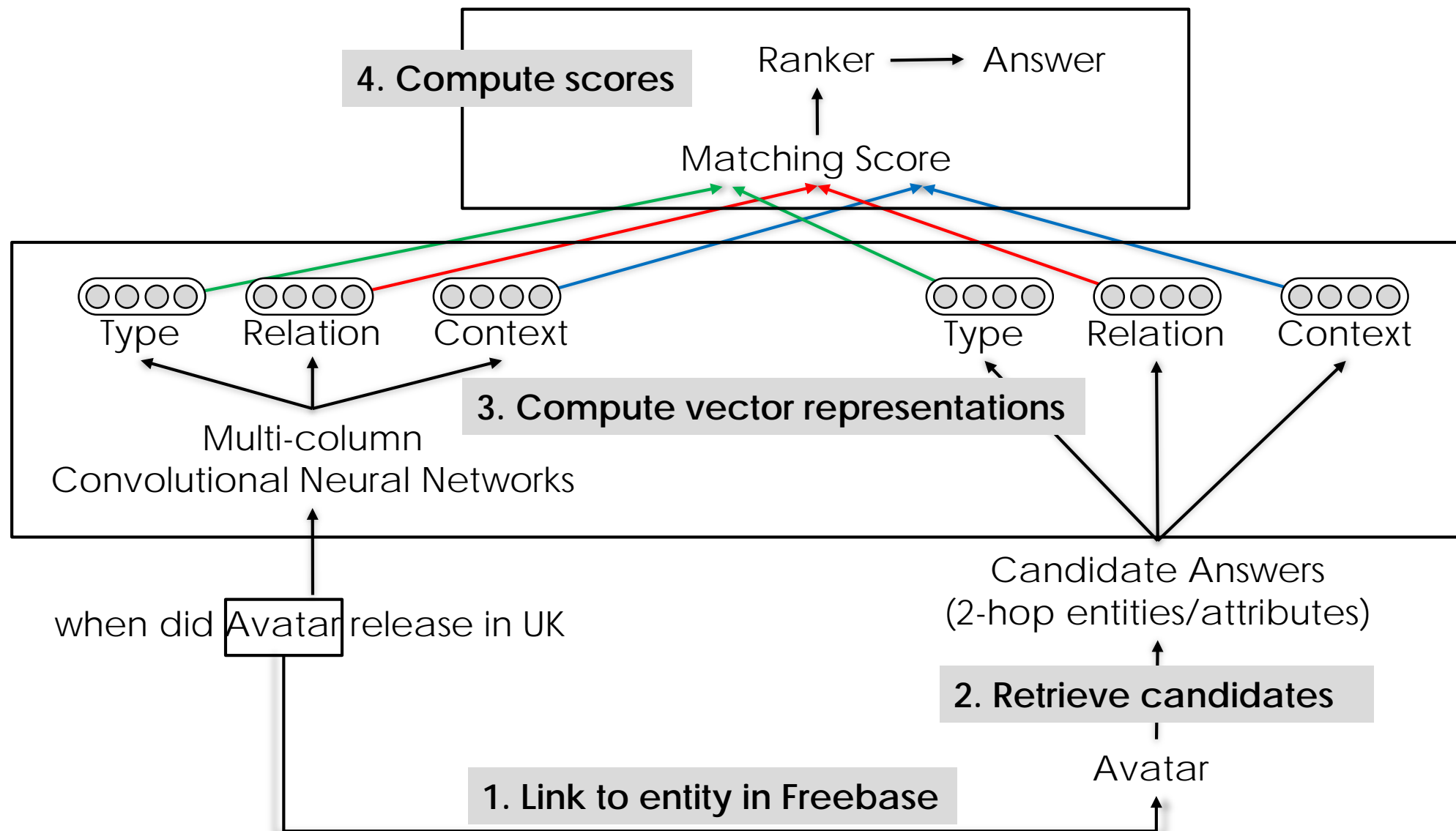
$$l(q, a, a') = (m - S(q, a) + S(q, a'))_+, \text{ where } (z)_+ = \max\{0, z\}$$

- Objective function
  - $A_q$ : set of correct answers
  - $R_q \subseteq C_q \setminus A_q$ : set of wrong answers

$$\min \sum_q \frac{1}{|A_q|} \sum_{a \in A_q} \sum_{a' \in R_q} l(q, a, a')$$

- Back-propagation, AdaGrad, max-norm regularization

# Inference (During Test)



# Inference (During Test)

- If there are more than one correct answers
  - Use the margin  $m$  in objective function as threshold
  - Candidates whose scores are not far from the best answer are regarded as predicted results

$$\hat{A}_q = \{\hat{a} \mid \hat{a} \in C_q \text{ and} \\ \max_{a' \in C_q} \{S(q, a')\} - S(q, \hat{a}) < m\}$$

# Question Paraphrases for Multi-Task Learning

- Question understanding results of paraphrases should be same
  - *who is the father of A*
  - *who is A's father*
- So, the vectors of paraphrases computed by neural networks should be similar
  - Hinge loss
  - Negative instance is randomly sampled

$$l_p(q_1, q_2, q_3) = \sum_{i=1}^3 \left( m_p - \mathbf{f}_i(q_1)^T \mathbf{f}_i(q_2) + \mathbf{f}_i(q_1)^T \mathbf{f}_i(q_3) \right)_+$$



# Experiments

- WebQuestions (Berant et al., 2013)
  - wh- questions collected by querying Google Suggest API
  - Annotated in Amazon Mechanical Turk
  - Train: 3023, Dev: 755, Test: 2032
- Example
  - Question: what is the name of justin beiber brother?
  - Url: [http://www.freebase.com/view/en/justin\\_bieber](http://www.freebase.com/view/en/justin_bieber)
  - Answers: {Jazmyn Bieber, Jaxon Bieber}
- Paraphrases (Fader et al., 2013)
  - Collected from the WikiAnswers website
  - ~2.4M questions, grouped into ~355k paraphrase clusters

# Experimental Results

- Better or comparable results than baseline methods

		Method	F1	P@1
Semantic Parsing	{	(Berant et al., 2013)	31.4	-
		(Berant and Liang, 2014)	39.9	-
		(Bao et al., 2014)	37.5	-
Information Extraction	{	(Yao and Van Durme, 2014)	33.0	-
		(Bordes et al., 2014a)	39.2	40.4
		(Bordes et al., 2014b)	29.7	31.3
		MCCNN (our)	<b>40.8</b>	<b>45.1</b>

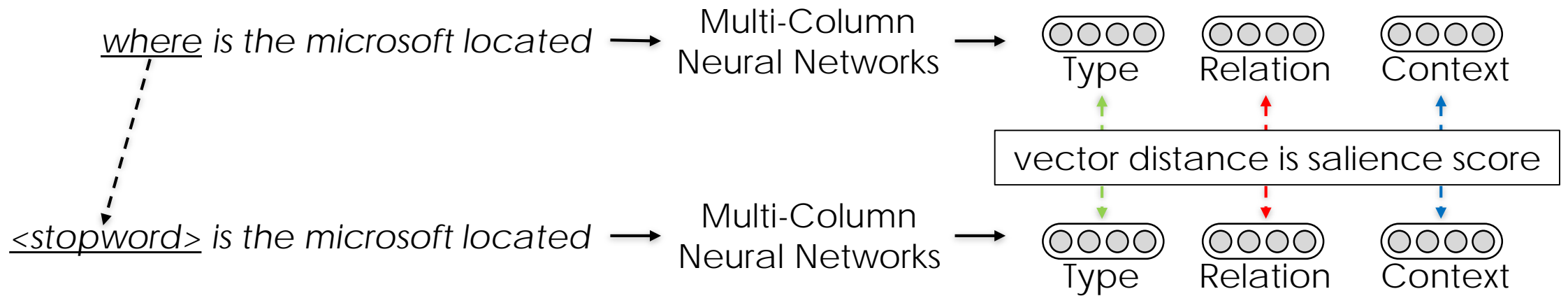
# Model Analysis

- Ablation experiments
  - w/o path/type/context
    - without the specific column
  - w/o multi-column
    - tying parameters of multiple columns
  - w/o paraphrase
    - without question paraphrases
  - 1-hop
    - 1-hop paths to generate candidates

Setting	F1	P@1
all	<b>40.8</b>	<b>45.1</b>
w/o path	32.5	37.1
w/o type	37.7	40.9
w/o context	39.1	41.0
w/o multi-column	38.4	41.8
w/o paraphrase	40.0	43.9
1-hop	29.3	32.2

# Salient Question Words Detection

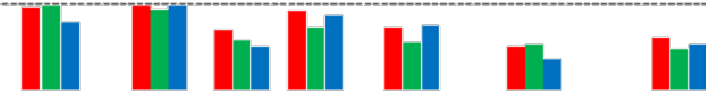
- Salience score
  - How much a word affects question understanding
  - Replace a word with stop words, how much the vectors are affected



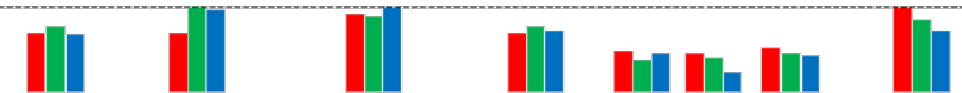
# Salient Question Words Detection

- Observations
  - wh- words
  - nouns dependent of the wh- words
    - *type/country/leader*
  - verbs
    - *speak/located*

what type of car does weston drive



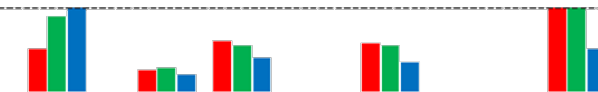
what countries speak german as a first language



who is the current leader of cuba today



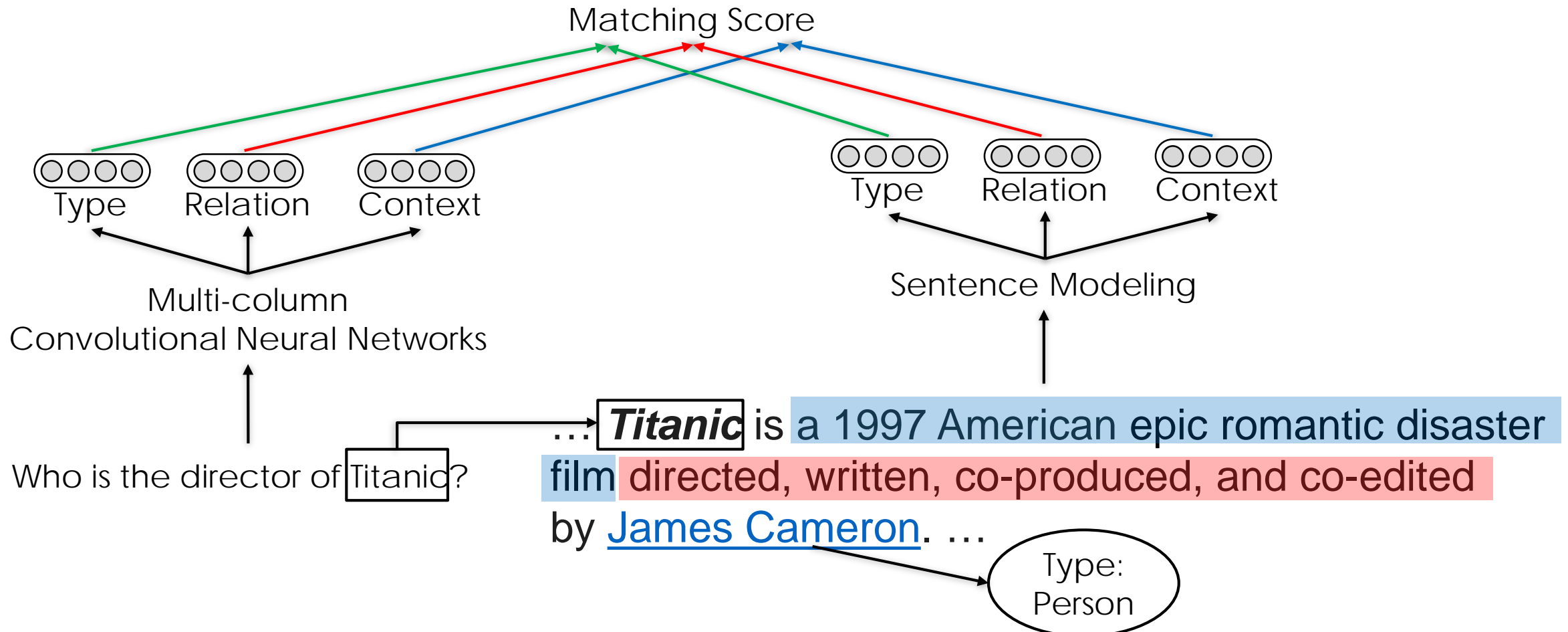
where is the microsoft located



■ Answer Path  
■ Answer Type  
■ Answer Context

# Future Work

- Question answering over unstructured text



# THANKS!



Column 1 (Answer Path)	Column 2 (Answer Type)	Column 3 (Answer Context)
<b>what to do in hollywood can this weekend</b> what to do in midland tx this weekend what to do in cancun with family what to do at fairfield can what to see in downtown asheville nc what to see in toronto top 10	<b>where be george washington originally from</b> where be george washington carver from where be george bush from where be the thame river source where be the main headquarters of google in what town do ned kelly and he family grow up	<b>where do charle draw go to college</b> where do kevin love go to college where do pauley perrette go to college where do kevin jame go to college where do charle draw go to high school where do draw bree go to college wikianswer
<b>who found collegehumor</b> who found the roanoke settlement who own skywest who start mary kay who be the owner of kfc who own wikimedia foundation	<b>who be the leader of north korea today</b> who be the leader of syrium now who be the leader of cuba 2012 who be the leader of france 2012 who be the current leader of cuba today who be the minority leader of the house of representative now	<b>who be judy garland father</b> who be clint eastwood date who be emma stone father who be robin robert father who miley cyrus engage to who be chri cooley marry to
<b>what type of money do japanese use</b> what kind of money do japanese use what type of money do jamaica use what type of currency do brazil use what type of money do you use in cuba what money do japanese use	<b>what be the two official language of paraguay</b> what be the local language of israel what be the four official language of nigerium what be the official language of jamaica what be the dominant language of jamaica what be the official language of brazil now	<b>what be the timezone in vancouver</b> what be my timezone in californium what be los angeles california time zone what be my timezone in oklahoma what be my timezone in louisiana what be the time zone in france