A Neural Network for Factoid Question Answering over Paragraphs

code & data: http://cs.umd.edu/~miyyer/qblearn

Mohit Iyyer[1], Jordan Boyd-Graber[2], Leonardo Claudino[1], Richard Socher[3], and Hal Daumé III[1]

University of Maryland, College Park[1] University of Colorado, Boulder[2] Stanford University[3]

THE TASK: QUIZ BOWL

Quiz bowl is a trivia game where players are read paragraph-length questions and can "buzz in" at any point during the question.

Q: *He left unfinished a novel whose title character forges his father's signature to get out of school and avoids the draft by feigning desire to join. A more famous work by this author tells of the rise and fall of the composer* ***Adrian Leverkühn****. Another of his novels features the jesuit* ***Naptha*** *and his opponent* ***Settembrini****, while his most famous work depicts the aging writer* ***Gustav von Aschenbach****. For ten points, name this German author of* ***The Magic Mountain*** *and* ***Death in Venice****.*

A: ***Thomas Mann***

WHY IS THIS CHALLENGING?

- **Question pyramidality:** earlier sentences contain harder clues than later ones
- Early sentences usually contain very few if any named entities indicative of the answer
- Have to decide *when* to answer the question as well as *what* answer to give

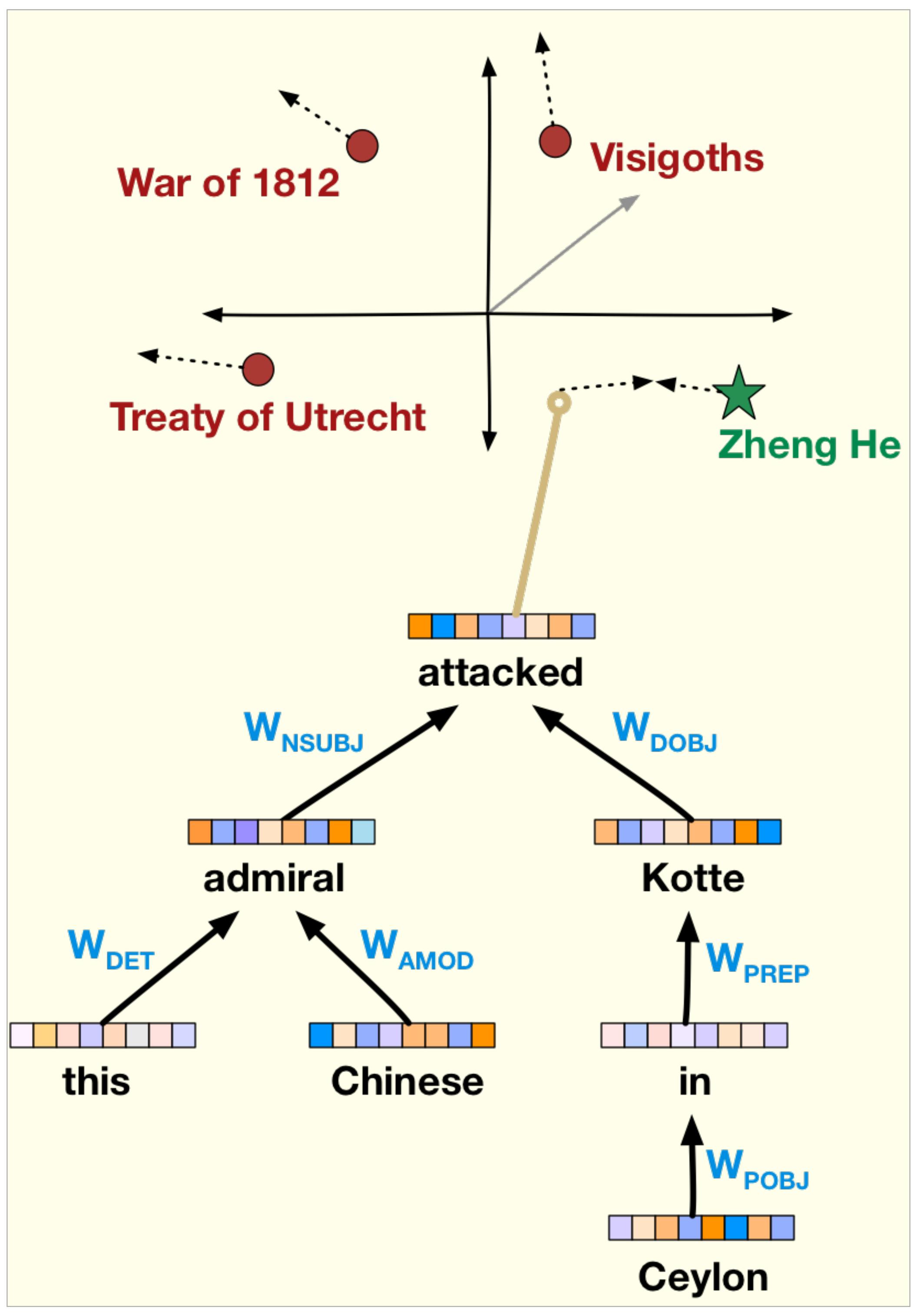
WHY NOT TRADITIONAL QA?

- IR systems work by querying some large knowledge base for terms similar to those in the query. But what if the query lacks informative terms?
- In such cases, we have to model the **compositionality** of the query.

CONTRIBUTIONS OF OUR WORK

- A dependency-tree recursive neural network model, **QANTA**, that computes distributed question representations to predict answers.
- QANTA outperforms multiple strong baselines and defeats human quiz bowl players when combined with IR methods.

HOW DOES IT WORK?



- Each word is associated with a vector $\mathbf{x_w}$
- Each dependency relation **r** is associated with a matrix $\mathbf{W_r}$
- The hidden representation $\mathbf{h_n}$ at node **n** is:

$$f(W_v \cdot x_w + b + \sum_{k \in K(n)} W_{R(n,k)} \cdot h_k)$$

- Paragraph representations are computed by averaging sentence representations.

HOW IS IT TRAINED?

- We want to push a computed question representation $\mathbf{h_q}$ close to its answer and far away from incorrect answers.
- We randomly sample *j* incorrect answers for each question and minimize a contrastive max-margin objective.
- The WARP loss proposed in Weston et al. (IJCAI, 2011) significantly improves accuracy.

EXPERIMENTAL MODELS:

- **BoW, BoW-DT** – unigram bag-of-words logistic regression baseline
- **IR-QB, IR-WIKI** – uses Whoosh, an IR engine, to search a knowledge base of training QA pairs and Wikipedia with BM-25 term weighting, query expansion, and fuzzy queries.
- **QANTA, FIXED QANTA** – our DT-RNN model, trained only on QA pairs, vary answer training
- **QANTA + IR-WIKI** – combines DT-RNN features with IR scores, our best model

DATA:

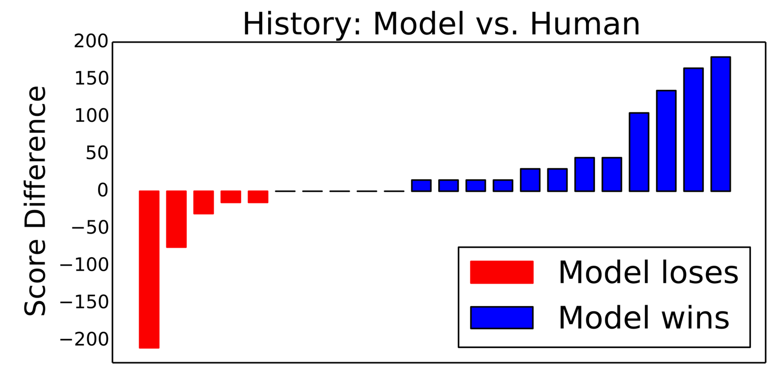
- Data was provided by NAQT (naqt.com).
- History dataset: *4,460* questions (*16,985* sentences), literature dataset: *5,685* questions (*21,549* sentences).

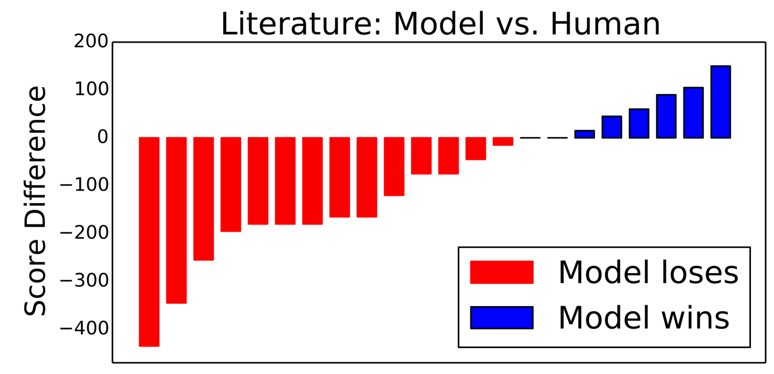
RESULTS:

	History			Literature		
Model	Pos 1	Pos 2	Full	Pos 1	Pos 2	Full
BOW	27.5	51.3	53.1	19.3	43.4	46.7
BOW-DT	35.4	57.7	60.2	24.4	51.8	55.7
IR-QB	37.5	65.9	71.4	27.4	54.0	61.9
FIXED-QANTA	38.3	64.4	66.2	28.9	57.7	62.3
QANTA	**47.1**	**72.1**	**73.7**	**36.4**	**68.2**	**69.1**
IR-WIKI	53.7	76.6	77.5	41.8	74.0	73.3
QANTA+IR-WIKI	**59.8**	**81.8**	**82.3**	**44.7**	**78.7**	**76.6**

HUMAN EVALUATION:

We compare our model to 22 skilled quiz bowl players on both datasets; we beat the average human at history questions.

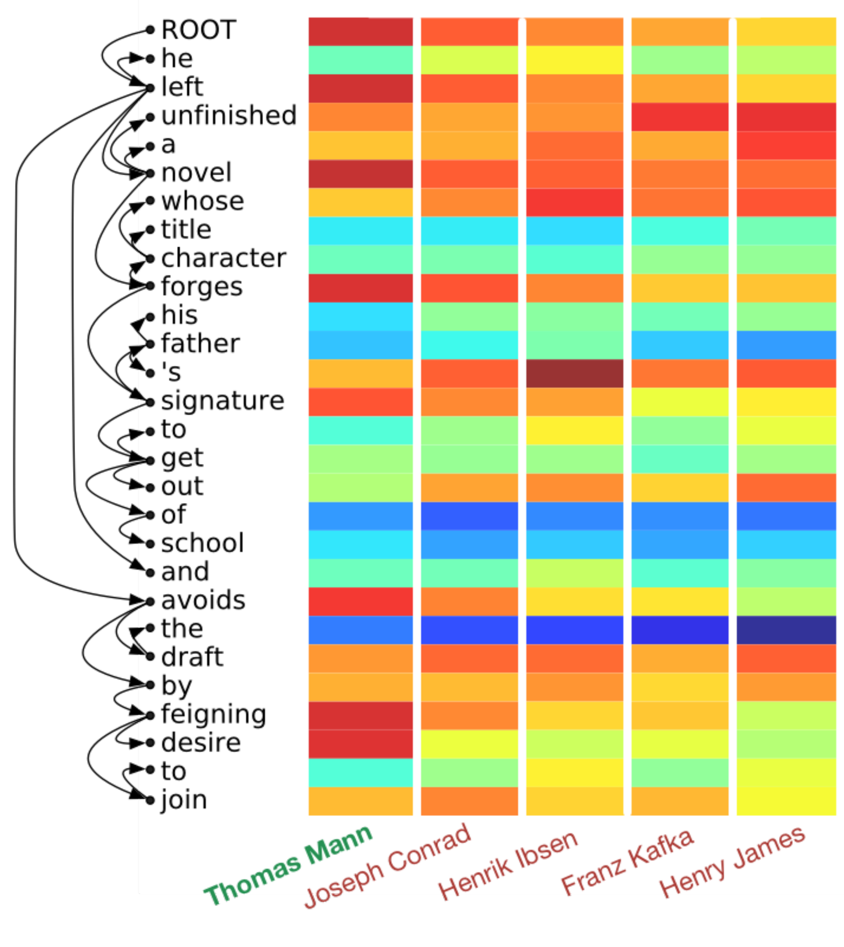




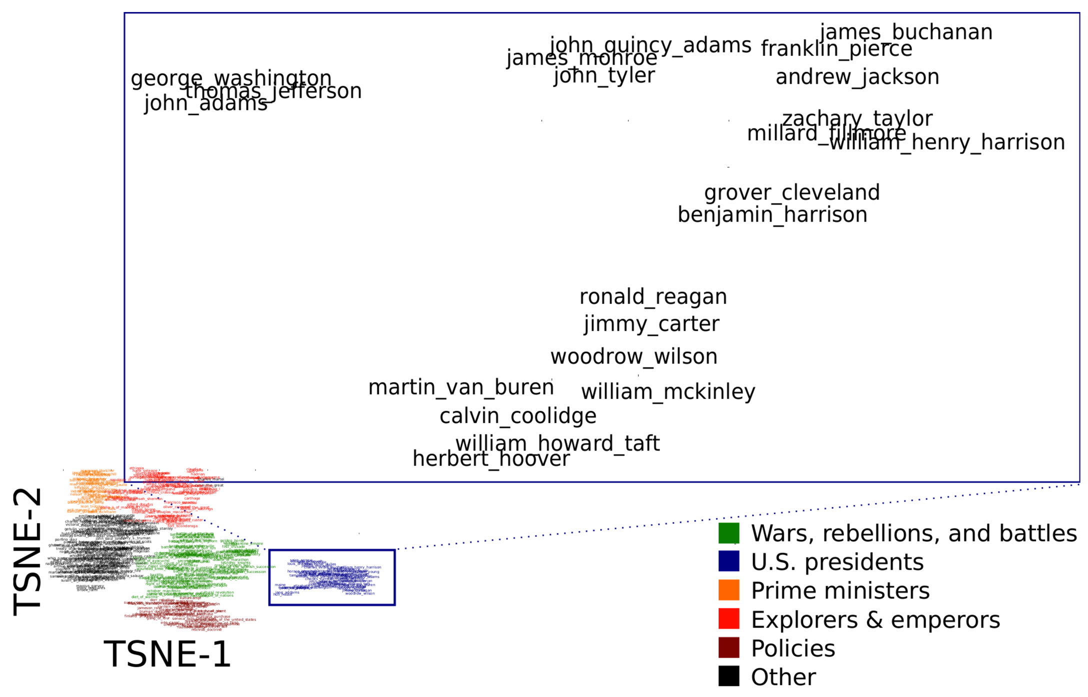
MOTIVATING THE MODEL

- QANTA builds on the DT-RNN model introduced by Socher et al. (TACL, 2014) for caption→image mapping.
- The key difference: **we train both the questions and answers in the same vector space.**
- Why is this useful? We don't want to treat answers as independent of one another.
  - The ***Battle of the Bulge*** may occur in questions about ***World War II***, and vice versa.
- The bag-of-words model of Boyd-Graber et al. (EMNLP, 2012) lost to human players because it was unable to answer quickly.

QANTA IN ACTION:



LEARNING ATTRIBUTES



FUTURE WORK:

- **demo** QANTA at the 2015 NAQT High School National Championships