Machine Comprehension with Discourse Relations

Karthik Narasimhan
Regina Barzilay

CSAIL, Massachusetts Institute of Technology

Sally liked going outside. She put on her shoes. She went outside to walk. [...] Missy the cat meowed to Sally. Sally waved to Missy the cat. [...] Sally hears her name. ”Sally, Sally, come home”, Sally’s mom calls out. Sally runs home to her Mom. Sally liked going outside.

Why did Sally put on her shoes?

A) To wave to Missy the cat

B) To hear her name

C) Because she wanted to go outside

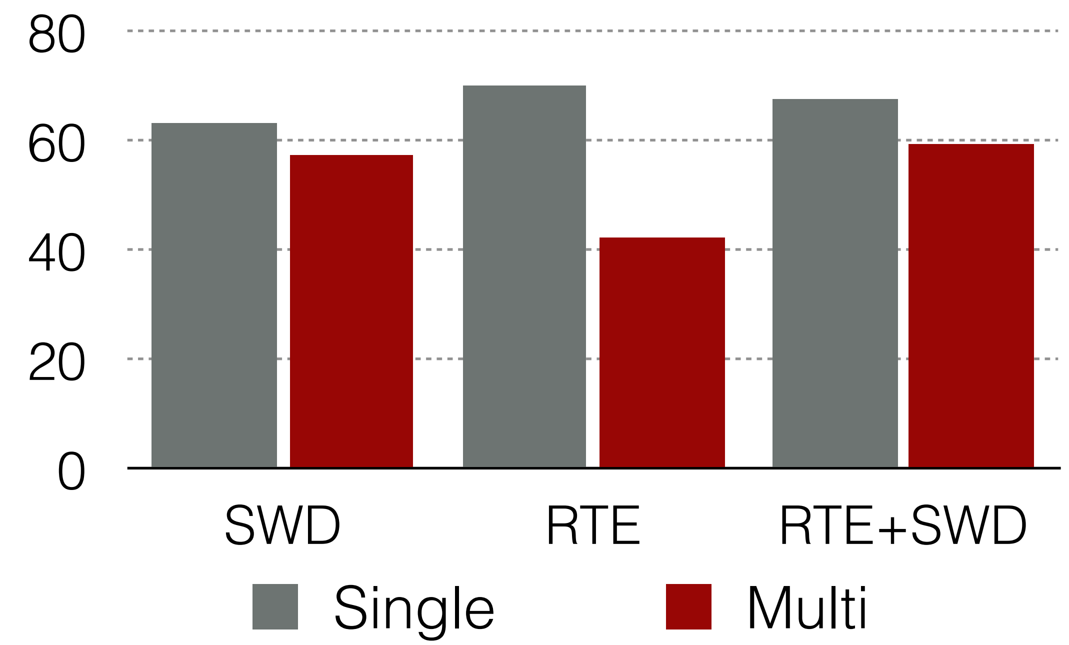
D) To come home

Sample passage excerpt and question in a Machine Comprehension task

Reasoning over multiple sentences

Multi-sentential questions are significantly harder than single sentence ones

Accuracy of baseline systems on MC500-test



We focus on modeling multi-sentence relations to improve Q&A performance.

Is there only a single relation?

Causality *Temporality*

She put on her shoes She went outside to walk

Why did Sally put on her shoes?

When did Sally put on her shoes?

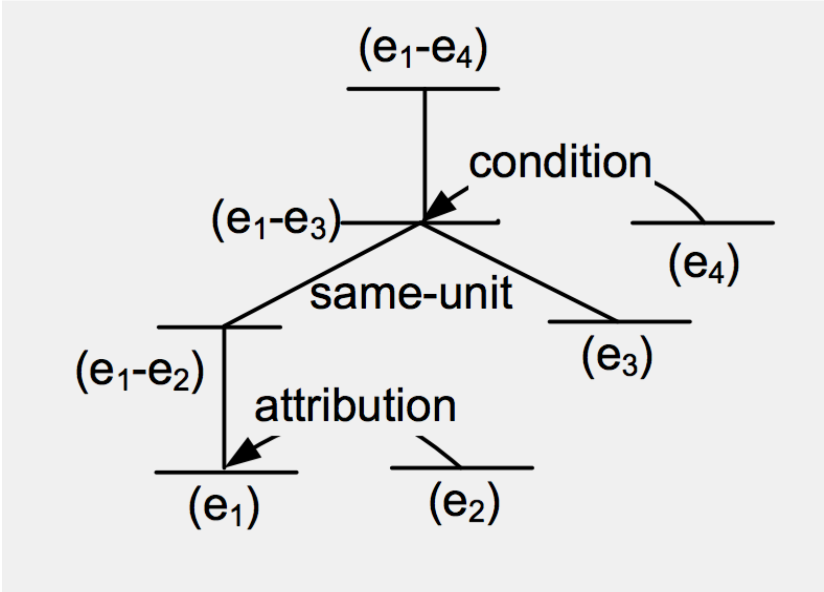
Relation between two clauses is question-dependent.

Key idea: Learn relations optimized for MC

Sally liked going outside. [...]

Why did Sally put on her shoes?
C) Because she wanted to go outside ✓

Training data: Q&A pairs



Traditional approach: Use off-the-shelf discourse analyzers
(Source: Feng and Hirst, 2012)

Hypothesis: Task-based discourse relations can facilitate better Comprehension Q&A

Fully supervised case

Causality

She put on her shoes

She went outside to walk

Why did Sally put on her shoes?
A) To wave to Missy the cat
B) To hear her name
C) Because she wanted to go outside ✓
D) To come home

She put on her shoes

She went outside to walk

Why did Sally put on her shoes?

A) To wave to Missy the cat

B) To hear her name

C) Because she wanted to go outside ✓

D) To come home

Sally liked going outside. She put on her shoes. She went outside to walk. [...] Missy the cat meowed to Sally. Sally waved to Missy the cat. [...] Sally hears her name. ”Sally, Sally, come home”, Sally’s mom calls out. Sally runs home to her Mom. Sally liked going outside.

Why did Sally put on her shoes?
A) To wave to Missy the cat
B) To hear her name
C) Because she wanted to go outside ✓
D) To come home

Key Steps

Infer correct relation

Causality

Identify relevant sentences

She put on her shoes

She went outside to walk

Why did Sally put on her shoes?

C) Because she wanted to go outside ✓

Select correct answer

Three models

Identify most relevant sentence from passage

Expand to a set of sentences

Infer inter-sentential relations

Sally liked going outside. She put on her shoes. She went outside to walk. [...] Missy the cat meowed to Sally. Sally waved to Missy the cat. [...] Sally hears her name. ”Sally, Sally, come home”, Sally’s mom calls out. Sally runs home to her Mom. Sally liked going outside.

Why did Sally put on her shoes?

A) To wave to Missy the cat

B) To hear her name

C) Because she wanted to go outside

D) To come home

Sentence - z

Question - q

Answer - a

Identifying a relevant sentence (Model 1)

- Retrieve a single relevant sentence from passage.
- Joint model over sentence z and answer choice a, given question q.

$$P(a, z \mid q) = P(z \mid q) \cdot P(a \mid z, q)$$

Identifying a relevant sentence set (Model 2)

Extends model 1 to select a *pair* of relevant sentences from passage.

- Retrieve a second sentence z_2 conditioned on both question q and the first retrieved sentence z_1.

$$P(a, z_1, z_2 \mid q) = P(z_1 \mid q) \cdot P(z_2 \mid z_1, q) \cdot P(a \mid z_1, z_2, q)$$

Incorporating Relations (Model 3)

Capture inter-sentential relations, modeled as hidden variables.

$$
\begin{aligned}
P(a, r, z_1, z_2 \mid q) = P(z_1 \mid q) \cdot P(r \mid q) \\
\cdot P(z_2 \mid z_1, r, q) \cdot P(a \mid z_1, z_2, r, q)
\end{aligned}
$$

Flexibility to induce relations between sentences *conditioned on the question*.

Learning

- *Supervision:* question-answer pairs.
- Marginalize over hidden variables *z* and *r* to get P(*a* | *q*).
- Maximize the following objective (model 3):

$$L_3(\theta; \mathcal{P}_{train}) = \log \sum_{i,j,m,r \in \mathcal{R}} \sum_{n \in [m-k, m+k]} P(a^*_{ij}, z_{im}, z_{in}, r \mid q_{ij}) - \lambda ||\theta||^2$$

Prediction

For a given question *q*, simply choose answer with highest P(*a* | *q*).

- Marginalize over all hidden variables *z* and *r*.

$$\hat{a}_j = \underset{k}{argmax}\ P(a_{jk}|q_j)$$

Lexical Features

Type 1 (q, z):

- Unigram and bigram matches + entity and action matches

Type 2 (q, a, z1, [z2]):

- Capture interactions between a, q and sentence(s) (z1, z2).

Relational Features

Type 3 (q, r, z1, z3) and Type 4 (q, r):

- Inter-sentence distance, presence of relation-specific markers (small seed list) in sentences.
- Second-order: cross of above features with entity and action match counts.
- Connect question word with relation type (Ex. *why* and *Causality*)

Discourse in Q&A

Prior work has shown value of domain-independent discourse relations in Q&A.

- *Chai and Jin (2004)* incorporate discourse processing into context Q&A.
- *Verberne et al. (2007)* use Rhetorical Structure Theory (RST) to relate question topics and answers.
- *Jansen et al. (2014)* use discourse information to improve answer re-ranking for non-factoid Q&A.

Experiments

- **Data:** MCTest (Richardson et al., 2013)

Split	MC160		MC500	
	Passages	Questions	Passages	Questions
Train	70	280	300	1200
Dev	30	120	50	200
Test	60	240	150	600

- > 50% of questions require information from multiple sentences.

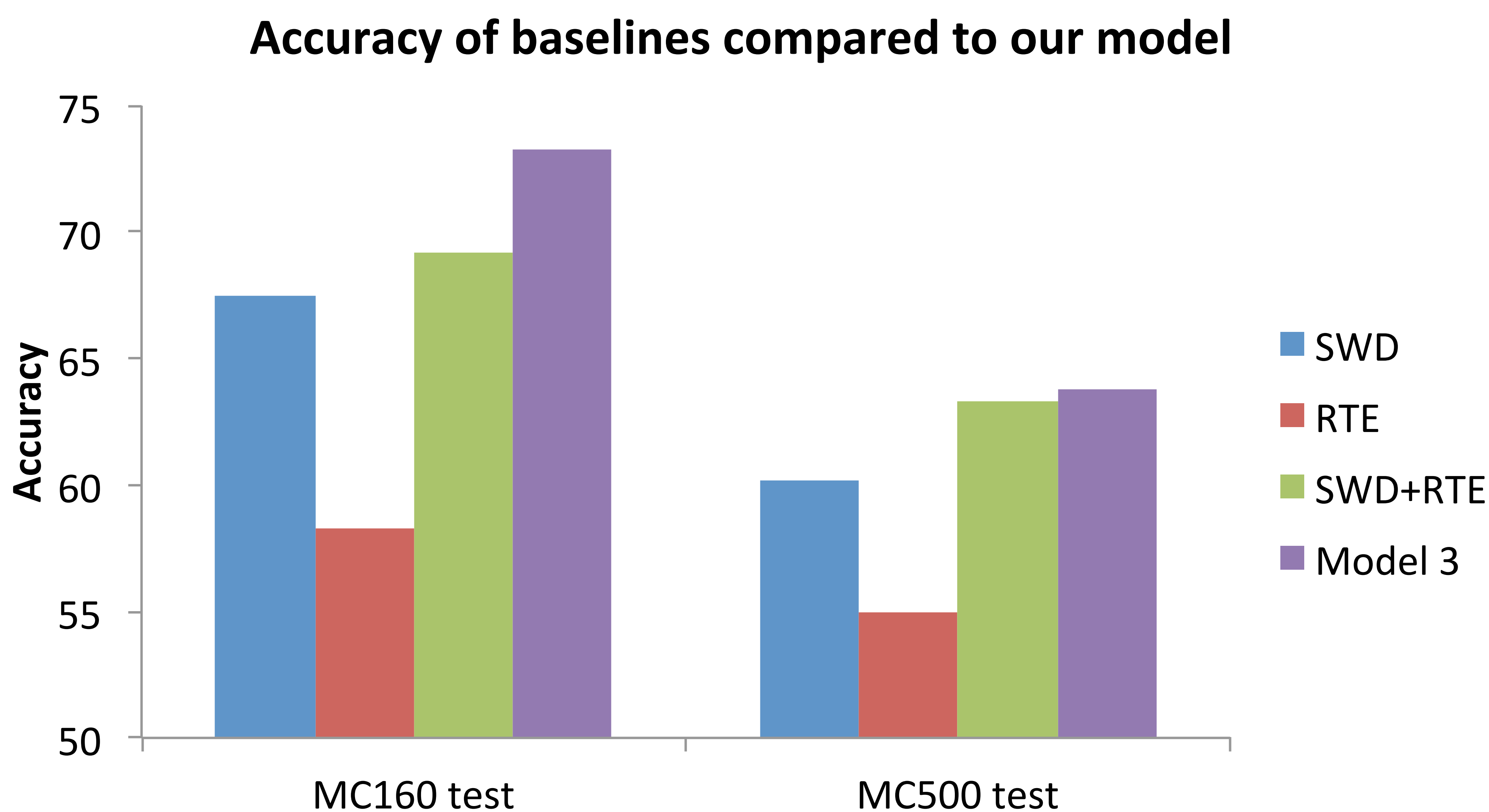
- **Evaluation:** Answering accuracy with partial credit for ties (as previously used).

Baselines

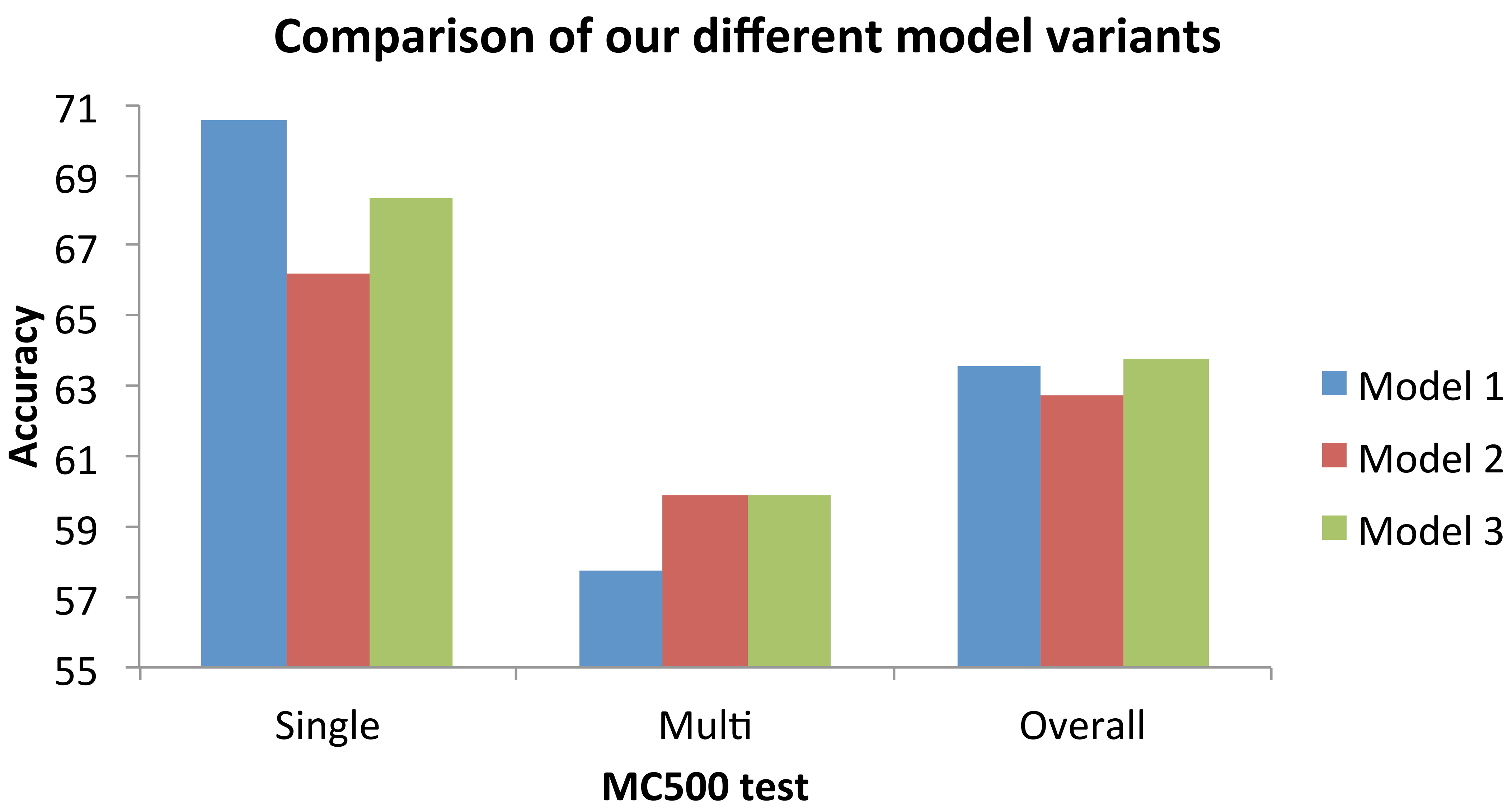
Systems from Richardson et al. (2013)

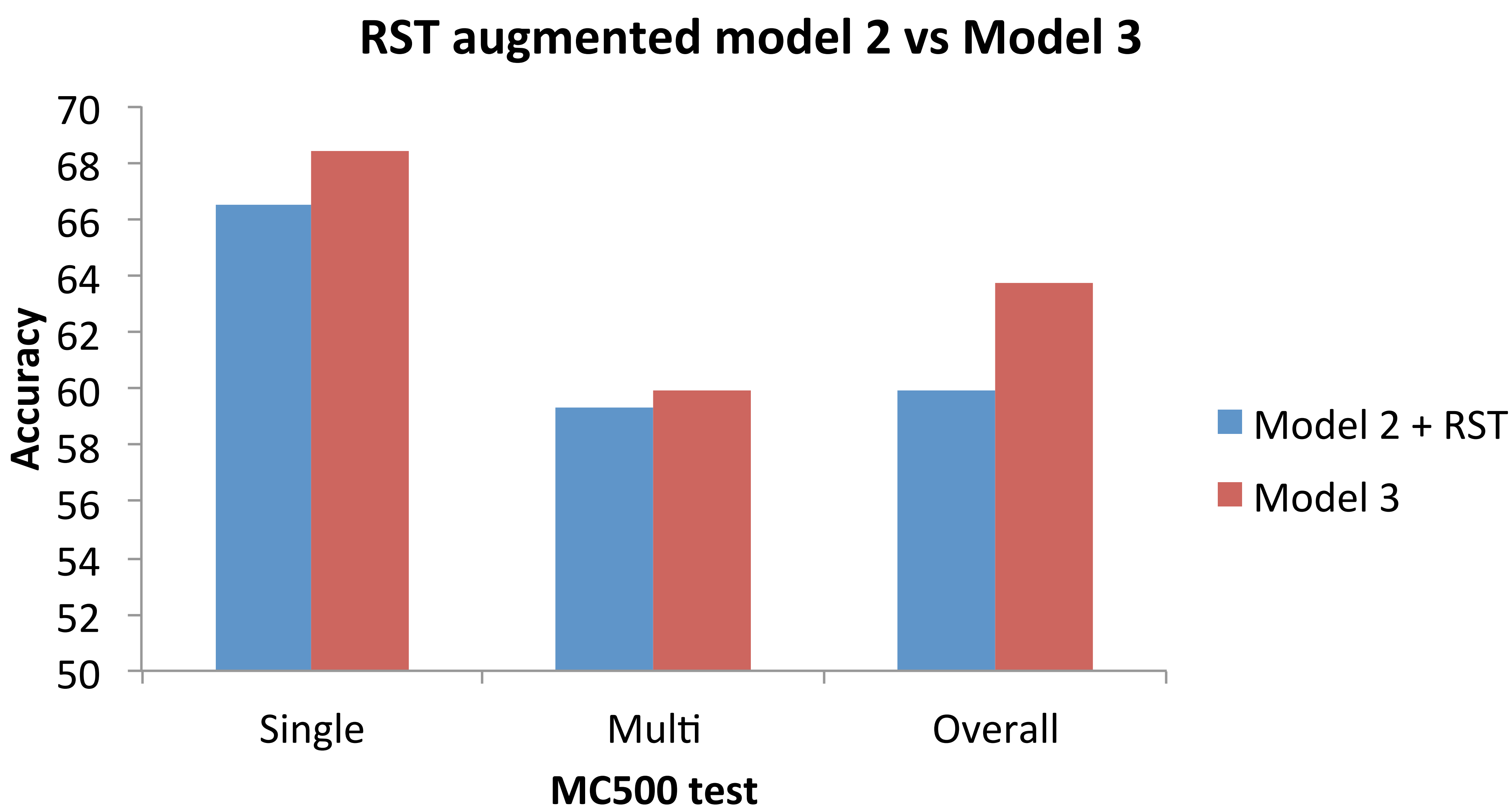
- SWD: uses sliding window to count matches between passage words and words in answer.
- RTE: utilizes a textual entailment system to determine if answer is entailed by passage.
- RTE+SWD: weighted combination of systems above

Comprehension Accuracy



Accuracy by Question Type





Task-based discourse relations can facilitate better Comprehension Q&A

77% of the predicted RST relations are *Elaboration!*

Evaluation using Human judgements

We annotated 240 questions from MC160 test set with most relevant sentence(s) in passage, and relations between sentence pairs.

- 103 sentence pairs with annotated relations
    - 34% of these have relevant discourse markers occurring anywhere in sentences.
    - Only 9% of sentences have a marker at an end.

Sentence Retrieval

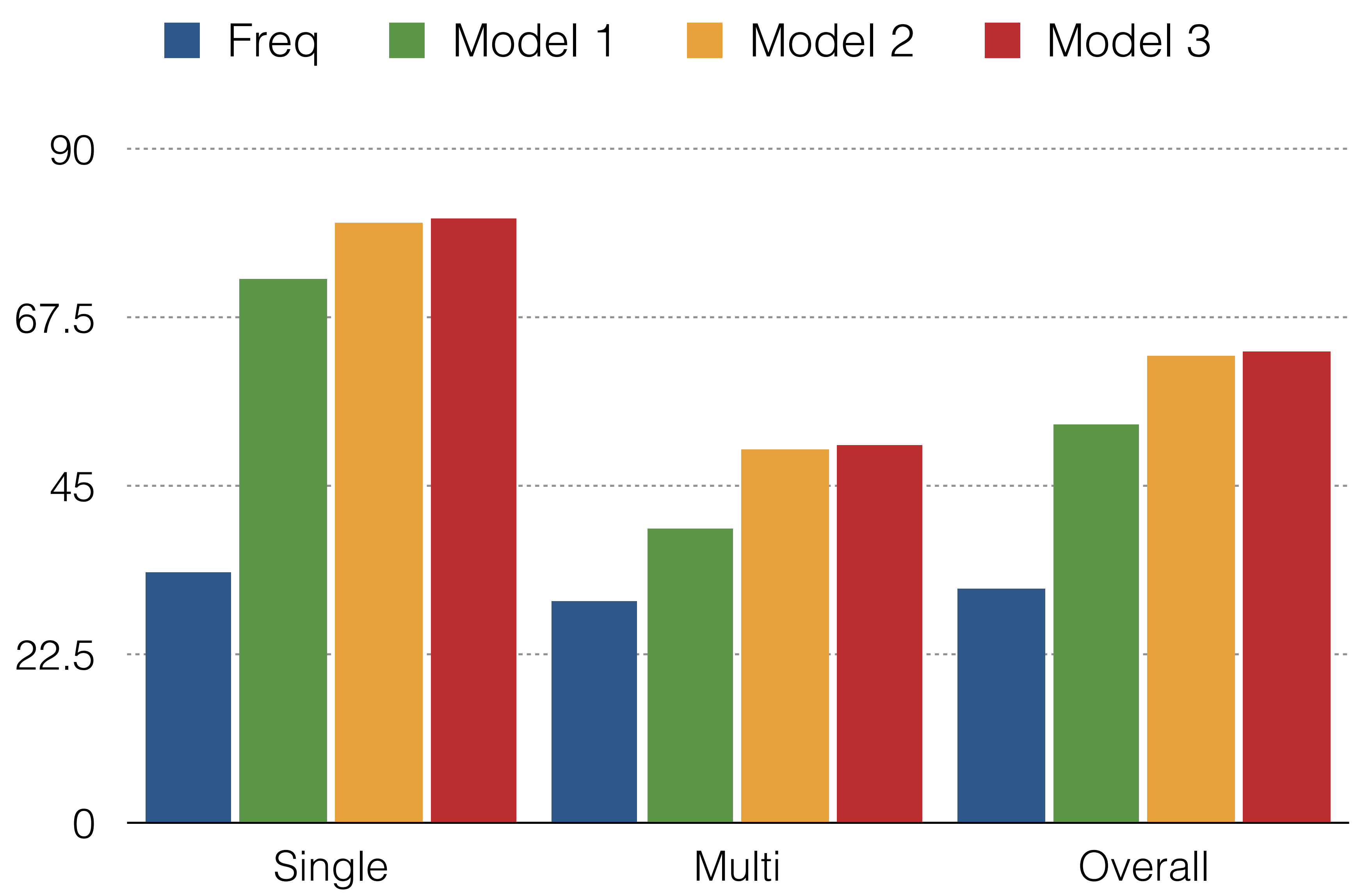


Table: Recall (@5) of relevant sentences retrieved by different models compared to human judgements.

Relation Prediction

Relation	**R @ 1**	**R @ 2**
Causal	56.25	75.00
Temporal	27.27	54.54
Explanation	16.66	33.33
Other	57.40	64.81
Overall	51.45	65.04

Table: Recall of annotated relations at various thresholds in ranking produced by Model 3

Conclusions

- Discourse relations help in the task of machine comprehension Q&A involving multiple sentences.
- A task-specific approach of incorporating discourse information does better than using off-the-shelf analyzers.

Code and data will be available at:
http://people.csail.mit.edu/karthikn/mcdr/