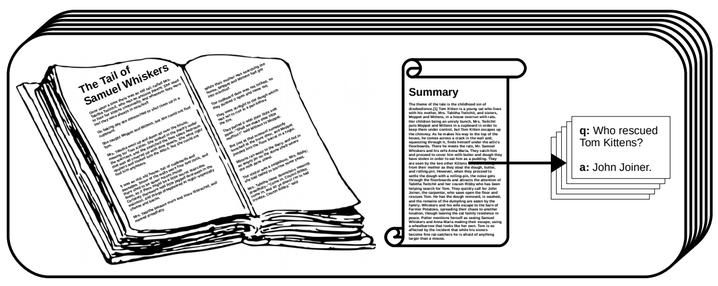
Extractive NarrativeQA with Heuristic Pre-Training

Lea Frermann, Melbourne University, lea.frermann@unimelb.edu.au

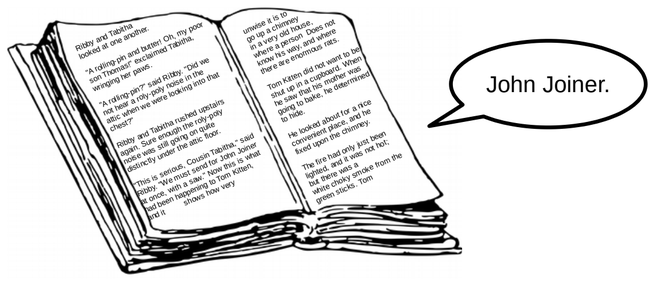
Task and Dataset

The NarrativeQA Dataset



The Original Task

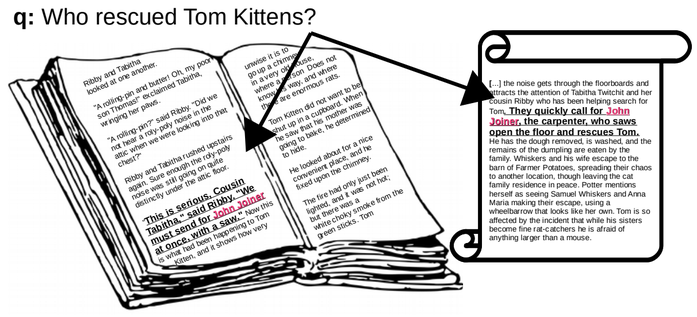
q: Who rescued Tom Kittens?



Here:

Extractive NarrativeQA

q: Who rescued Tom Kittens?



Challenges

- Abstractive answers in original data set
- Answer location in book and summary unknown
- Long-range reasoning required for book-level QA

Take-aways

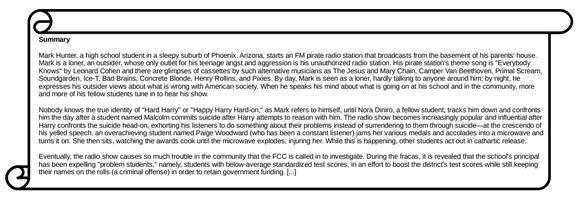
- A simpler, **heuristic data set** leads to effective models
- Heuristics on one domain (summaries) **generalize** to a second (books)
- **Automatic evaluation** metrics are **problematic** for evaluating abstractive QA

Synthetic Training Data

q= Why does Nora track Mark down?

a= {Malcom's suicide, to confront him after Malcolm commits suicide}

c=



rank $s \in \mathbf{c}$: $cos(\phi(s), \phi([\mathbf{q};\mathbf{a}]))$

$\mathbf{r_0}$: Nobody knows the true identity of Hard Harry until Nora Diniro, a fellow student, tracks him down and confronts him the day after a student named Malcolm commits suicide after Harry attempts [...]

$\mathbf{r_1}$: The radio show becomes increasingly [...]

extract: $\mathbf{a}^{\mathbf{ex}} = \arg\max_w heur(w|\mathbf{r}, \mathbf{q}, \mathbf{a})$

$\mathbf{r_0}$: Nobody knows the true identity of Hard Harry until Nora Diniro, a fellow student, tracks him down and confronts him the day after a student named **Malcolm commits suicide** after Harry attempts to [...]

$\mathbf{a}^{\mathbf{ex}}$= **Malcolm commits suicide**

w=word; s=sentence; $\phi()$ = Universal Sentence Encoder; $heur$ = heuristics based on word overlap and sentence ranking.

Approach

Setup

- train extractive models on {**c**, **q**, $\mathbf{a}^{\mathbf{ex}}$} tuples
- test extractive models on summary- and book-level QA

Model

- off-the-shelf BERT for predicting the best span $\hat{a} = w_i...w_j$:

$$z = [CLS]\ BERT(\mathbf{q})\ [SEP]\ BERT(\mathbf{s} \in \mathbf{c})$$

$$b(w_i) = \frac{e^{B \times w_i^s}}{\sum_l e^{B \times w_l^s}} \qquad e(w_{j>i}) = \frac{e^{E \times w_j^s}}{\sum_l e^{E \times w_l^s}}$$

$$\hat{a} = B \times b(w_i) + E \times e(w_j)$$

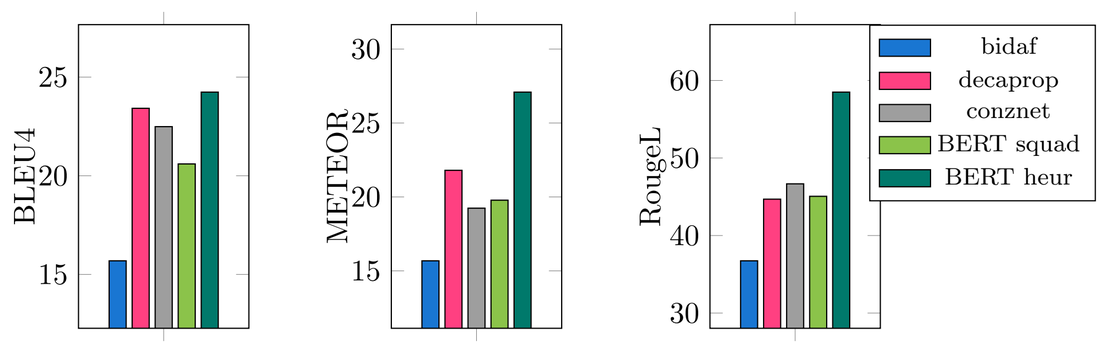
- maximize log-likelihood of true start and end positions

Book passage retrieval

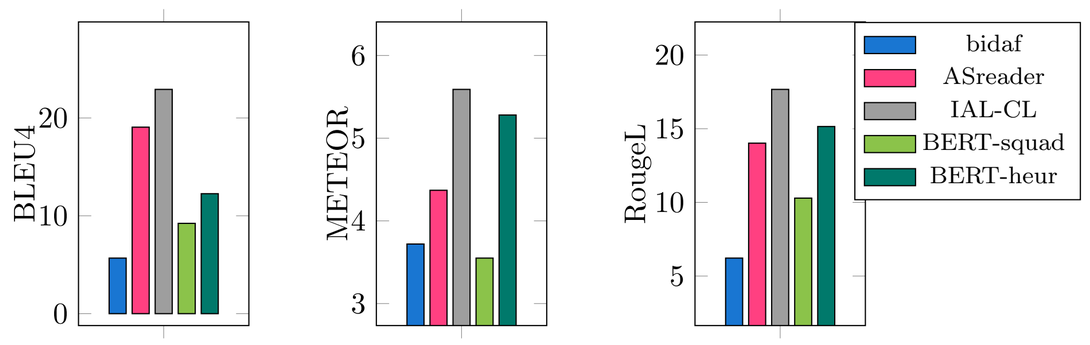
- exhaustive span prediction over full book text is infeasible
- train a separate BERT model for sentence relevance prediction
- $\mathbf{c} = n = 100$ most relevant sentences in ± 5 sentence context

Results

Summary-level NarrativeQA



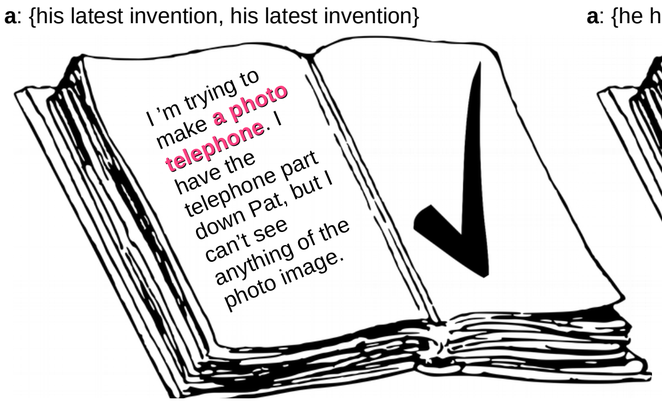
Book-level NarrativeQA



Examples and Error Analysis (Book-level)

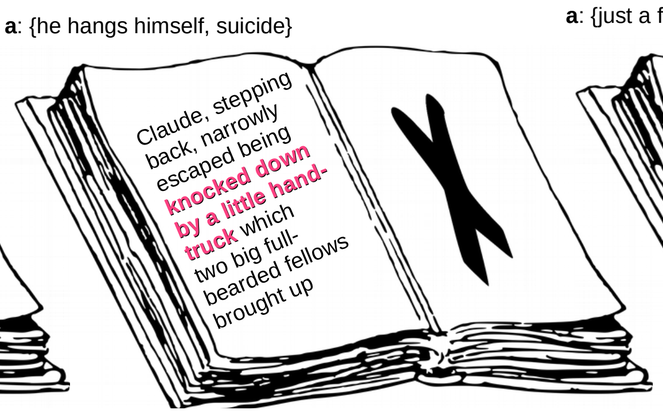
q: What is Tom trying to desperately get working?

a: {his latest invention, his latest invention}



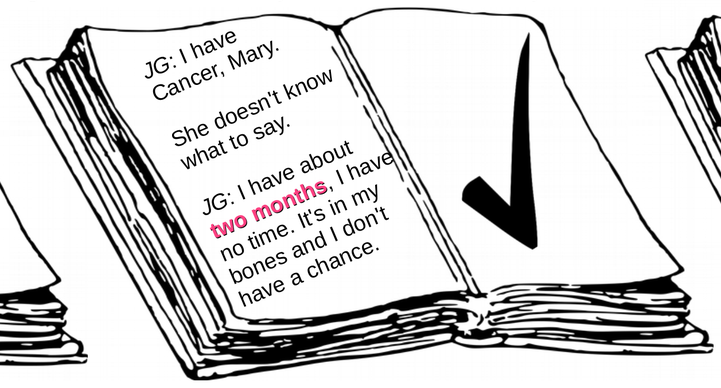
q: How does Claude die?

a: {he hangs himself, suicide}



q: How long does Jimmy Gator have to live?

a: {just a few months, he only has a few months to live}



q: How does Mr. Peters feel about Tom not allowing him access to the patents?

a: {he gets angry, mad}