Neural Domain Adaptation for Biomedical Question Answering

Georg Wiese[1,2], Dirk Weissenborn[2], Mariana Neves[1]

[1]Hasso Plattner Institute, August Bebel Strasse 88, Potsdam, Germany [2]Language Technology Lab, DFKI, Alt-Moabit 91c, Berlin, Germany

Motivation

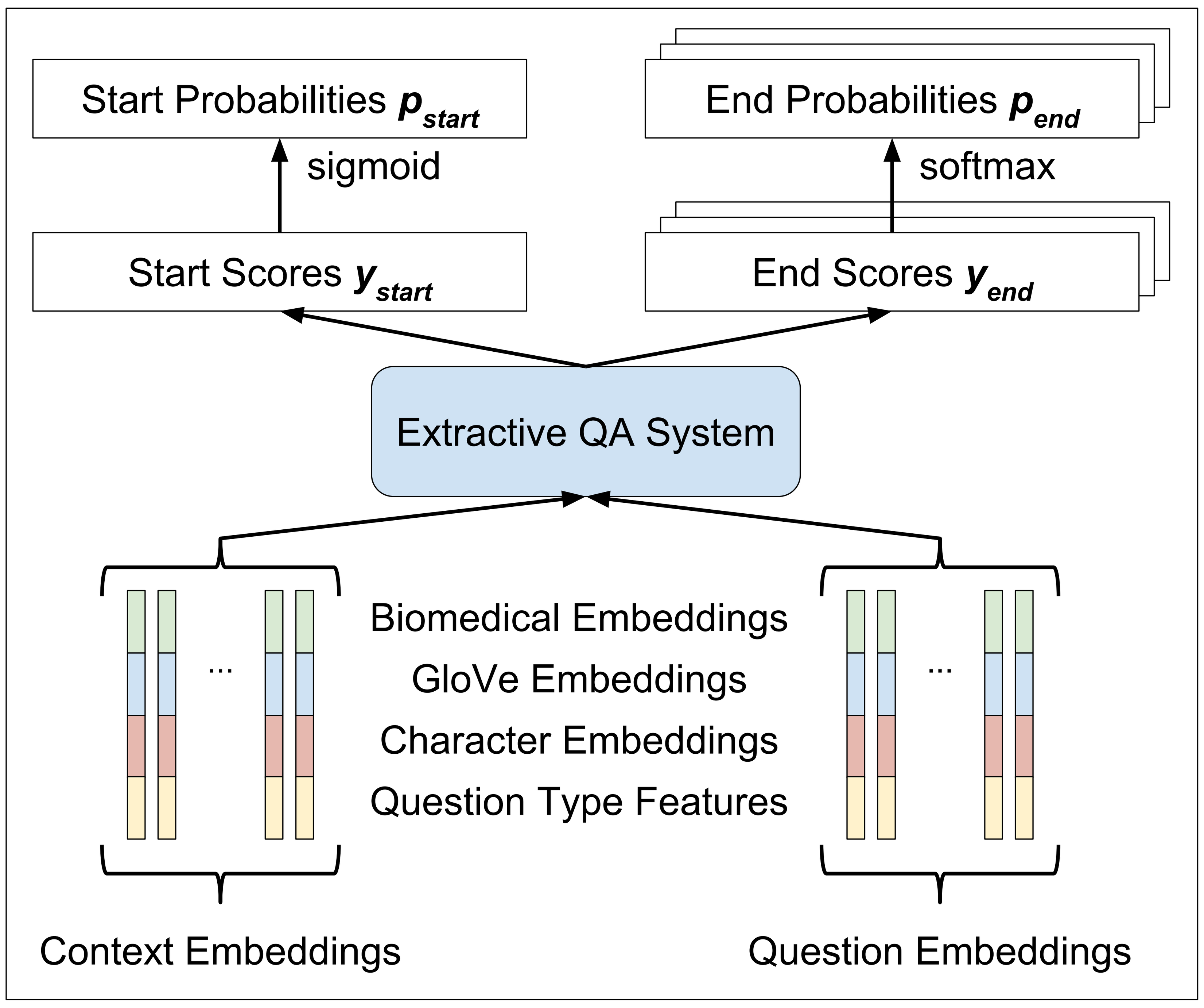
- **Neural question answering** (QA) systems outperform traditional methods in open-domain **factoid QA**.
- In biomedicine, datasets are **too small** to apply deep learning directly.
- Can we bridge this gap via **domain adaptation**?

Architecture & Training

- Our architecture wraps an existing **neural QA** system (FastQA [1]), with the following changes:
  - **Input Layer**: In addition to GloVe embeddings and character embeddings, we feed biomedical token embeddings and question type features.
  - **Output Layer**: We generalize our activation and decoding process to support list questions in addition to factoid questions.
- During **training**, we explore several domain adaptation techniques, including mere fine-tuning, joint training, and forgetting cost regularization [2].

Domain Adaptation

- Our system is pre-trained on **SQuAD**, a large-scale (10^5) open-domain factoid QA dataset.
- Then, we adapt the system to the biomedical domain, using **BioASQ**, a small (10^3) biomedical QA dataset.



Results

- **Pre-training** on SQuAD and **fine-tuning** on BioASQ already improves performance significantly over training on BioASQ only.
- The **forgetting cost** improves results slightly for factoid questions.

Experiment	Factoid MRR	List F1
Training on BioASQ only	17.9%	19.1%
Training on SQuAD only	20.0%	8.1%
Fine-tuning on BioASQ	24.6%	**23.6%**
Fine-tuning on BioASQ w/ forgetting cost	**26.2%**	21.1%

Comparison to state of the art

- In order to compare our system to the state of the art in biomedical QA, we tested it on the **2016 BioASQ** challenge.
- We compared a **single** model and model **ensemble**.
- Our system achieves **state-of-the-art results on factoid** questions and **competitive results on list** questions.

Experiment	Factoid MRR	List F1
Single model	24.8%	27.8%
Ensemble model	**27.5%**	26.5%
Best competitor	24.0%	**28.1%**

German Research Center for Artificial Intelligence

software campus

Federal Ministry of Education and Research

[1] Weissenborn et al., “Making Neural QA as Simple as Possible but not Simpler”

[2] Riemer et al., “Representation Stability as a Regularizer for Improved Text Analytics Transfer Learning”

HPI Hasso Plattner Institut

IT Systems Engineering | Universität Potsdam