



SQuAD-ing Up as a Winning Team: Character Embeddings and Dynamic Coattention Networks

Kevin Guo (kyguo)¹, Brandon Kang (bmkang)¹
Department of Computer Science, Stanford University, Stanford, CA¹

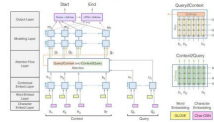


Introduction/Background

- Question Answering is a task in the Natural Language Processing community that has received a lot of attention due to its broad range of implications for issues ranging from web search to data analytics.
- Question Answering has further increase in popularity due in part to the robustness and reasoning-based nature of the SQuAD dataset.^[1]
- We first build upon the given baseline BiDAF model that contains only word-level embeddings by incorporating the **Character-Level Embeddings** as described in Seo et al.^[2]
- We additionally implement **Dynamic Coattention** and create a model combining character-level embeddings with the Dynamic Coattention Network described in Xiong et al.^[3]
- We measure the success of our implementations using EM and F1 scores and further experiment upon our models with **hyperparameter tuning** experiments.
- We find our model performs best with the Coattention and Character-Level Embeddings configuration with a learning rate of 0.6 and a dropout probability of 0.15 and associated EM and F1 scores of 59.189 and 62.787, respectively, on the test set.

Methods

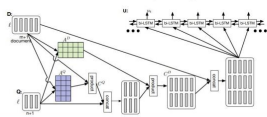
1. Bi-directional Attention Flow (BiDAF) Model



2. Character-Level Embeddings

- Pass preprocessed context/query words through a CNN to produce char embeddings.
- Maxpool, concatenate to word embeddings, pass to Highway Network.

3. Dynamic Coattention Networks (DCN)



Results

Model	lr	dropout	F1	EM
Baseline	0.5	0.2	60.74	57.39
Char-level Embs	0.6	0.2	60.86	58.11
Char-level Embs	0.5	0.25	61.92	58.47
Char-level Embs	0.6	0.15	62.27	59.32
Char-level Embs + Coattention	0.6	0.2	61.78	59.06
Char-level Embs + Coattention	0.2	0.3	57.31	54.17
Char-level Embs + Coattention	0.6	0.15	63.91	60.43

Fig 1: Hyperparameter Tuning on Dev Set

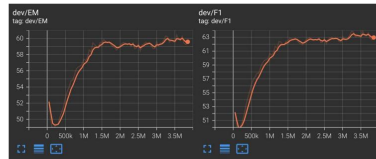


Fig 2: Best Model (Coattention + Char-level Embeddings + Hyperparameter Search) Performance

Model	F1	EM
Baseline on Dev	60.737	57.385
Char-Level Embs + Coattention + Tuned Parameter - Dev	63.914	60.427
Char-Level Embs + Coattention + Tuned Params - Test	62.704	59.138

Fig 3: Model Performance (EM and F1 scores)

Analysis

QUESTION

• Question: Why is Warsaw's flora very rich in species?

• Context: The flora of the city may be considered very rich in species. The species richness is mainly due to the location of Warsaw within the border region of several big forest regions comprising substantial proportions of conifers and deciduous trees (natural forest, vegetation along the Vistula) as well as arable land, meadows and parks. Warsaw's flora is located within the border that, besides its belonging to the Masovian Provincial Forest, Bielski Forest and the Warsaw Forest, is also adjacent to the Lublin Forest. Warsaw has also two botanical gardens: by the Łazienki park (didactic research unit of the University of Warsaw) as well as by the Park of Culture and Recreation in the district of Śródmieście.

• Answer: location of Warsaw

• Prediction: the location of Warsaw within the border region of several big forest regions

- Model pulled more data than necessary from the context; however, the extra information was also relevant to the ground truth answer.

QUESTION

• Question: What does it mean for the dollar to float?

• Context: On August 15, 1971, the United States unilaterally pulled out of the Bretton Woods Accord. The US abandoned the Gold Exchange Standard whereby the value of the dollar had been pegged to the price of gold and all other currencies were pegged to the dollar. The dollar was left to float according to market demand. Shortly thereafter, Britain followed, floating the pound sterling. The other industrialized nations followed suit with their respective currencies. Anticipating that currency values would fluctuate significantly, large firms, the international traders increased their reserves. By expanding their money supplies in amounts far greater than before. The result was a devaluation of the dollar and other international national currencies. Because of was priced in dollars, of production and income decreased. In September 1971, OPEC issued a joint communique stating that from then on, they would price oil in terms of a fixed amount of gold.

• Answer: float

• Prediction: rise and fall according to market demand

- Able to pull answer from context despite human labeling error

Conclusions

- Adding character-level embeddings to the baseline BiDAF increased performance.
- Substituting the attention layer in BiDAF with coattention marginally increased performance.
- We were able to improve and achieve higher accuracy after performing a hyperparameter search.
- Our best model achieved an F1 score of 62.704 and an EM score of 59.138 on the test set.
- For future improvements, we are curious to see what would happen with a feedforward model that separates question types.
- We also suggest implementing data augmentation such as Easy Data Augmentation or more types of attention such as self-attention.

References

- [1] Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for SQuAD. In Association for Computational Linguistics (ACL), 2018.
- [2] Minjoon Seo, Anirudha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension, 2016.
- [3] Caimeing Xiong, Victor Zhong, and Richard Socher. Dynamic coattention networks for question answering. In arXiv, 2018