**QuAVONet: Answering Questions on SQuAD 2.0 with Neural Networks**

**Jesus Cervantes** (cerjesus@stanford.edu)

---

**PROBLEM DEFINITION**

**Objective:** QuAVONet seeks to answer passage-based reading comprehension questions without the use of Pre-trained Contextual Embeddings (PCEs).

**Motivation:** QANet initially worked fairly well on SQuAD 1.1, with few unanswerable questions. By combining QANet with Answer Verifier (AV) from U-Net, QuAVONet seeks to adapt QANet into the world of determining answerability.

---

**DATA**

The SQuAD 2.0 dataset is a set of 150,000 questions whose answers either lie as a span of the corresponding passage. New to Squad 2.0, 50,000 of these questions, can’t be answered with the corresponding passage.

---

**Answer Verifier**

- Component from U-Net [3]
  - Use logistic regression to determine answerability
- Logistic Regression - $x = (c_0; \theta_m + 1; c_e; c_e) 
  - $J(\theta) = \sum_{i=1}^{y(i)} \log h(x(i)) + (1 - y(i)) \log(1 - h(x(i)))$
  - $c_s = p^1 \cdot B, c_e = p^2 \cdot B$
- Binary Cross-Entropy Loss
  - $L_{AV} \equiv (\sigma \log(p^p) + (1 - \sigma)(1 - \log(p^p)))$
- If P(answerable) < threshold, predict NA
  - Threshold from paper = 0.3

---

**QANet**

- Encoder Block based on Transformer
- Loss function:
  $$L(\theta) = -\frac{1}{N} \sum_{i=1}^{N} (\log(p_i^{c_i}) + \log(p_i^{c_j}))$$

---

**RESULTS**

Various model Performance during training

---

**DISCUSSION**

- 8-Head, d_model=128 QANet does not perform as well as 4-head QANet, d_model =96 QANet, likely due to slightly increased overfitting
- QANet significantly stunts training speed: 7 hours for 3M iters on QAEmb vs 25 hours for 4-head QANet on NV-6
- While adding QANet Embedding significantly boosts baseline performance (dark blue on right), adding the QA Embedding Encoder (bottom orange left) stunts performance
- Replacing the QA Embedding Encoder with RNN Encoder and convolution in QANet (pink left) does not boost speed or performance
- QuAVONet does not surpass QANet and takes more iterations to overfit and achieve comparable performance
- Adding AV to QAEmb (mint left) retains superior speed, but stunts performance

---

**REFERENCES**
