Multi-task Transfer Learning for Sentimentally consistent Summarization

Arth Dharaskar - Sai Ashish Somayajula

Motivation

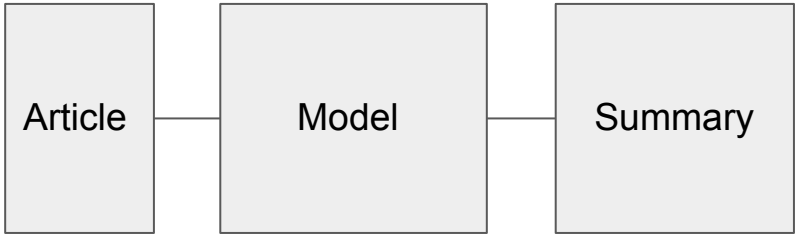
- Internet consists of a large corpus of **Text**

- We are often limited on **Time**

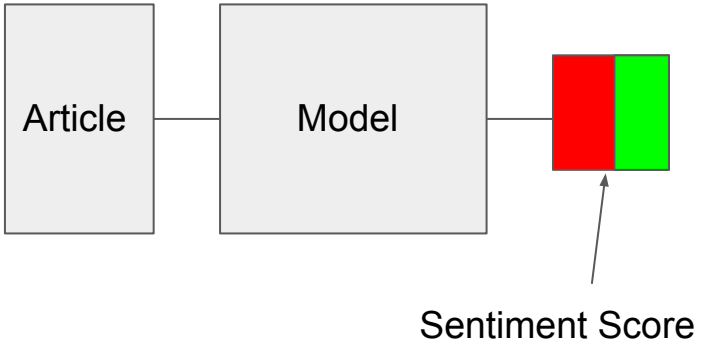
- Human beings respond to **Strong Sentiments**

Targeted Tasks

Text Summarization



Sentiment classification



Dataset

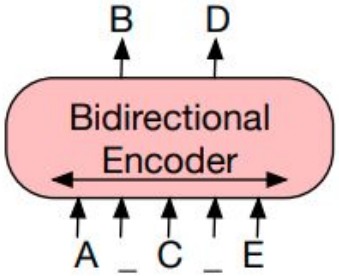
Fine-Tuning BART

- CNN Dailymail ~ 300k unique articles + summary

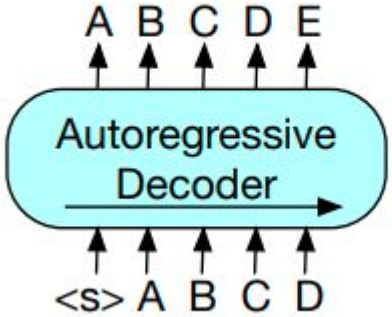
Pre-training Classifier Network

- IMDB Dataset ~ 25k polar movie reviews

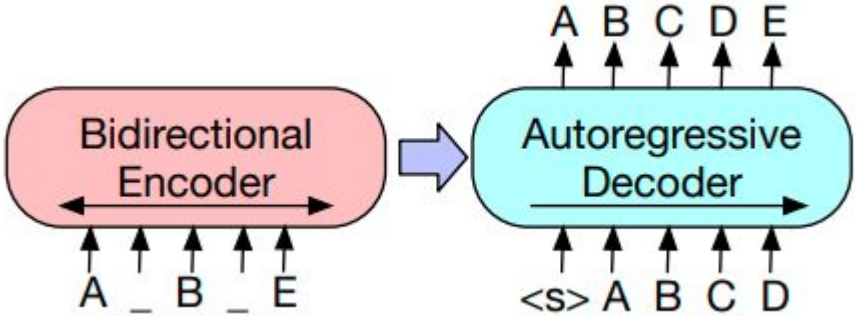
BART



BERT

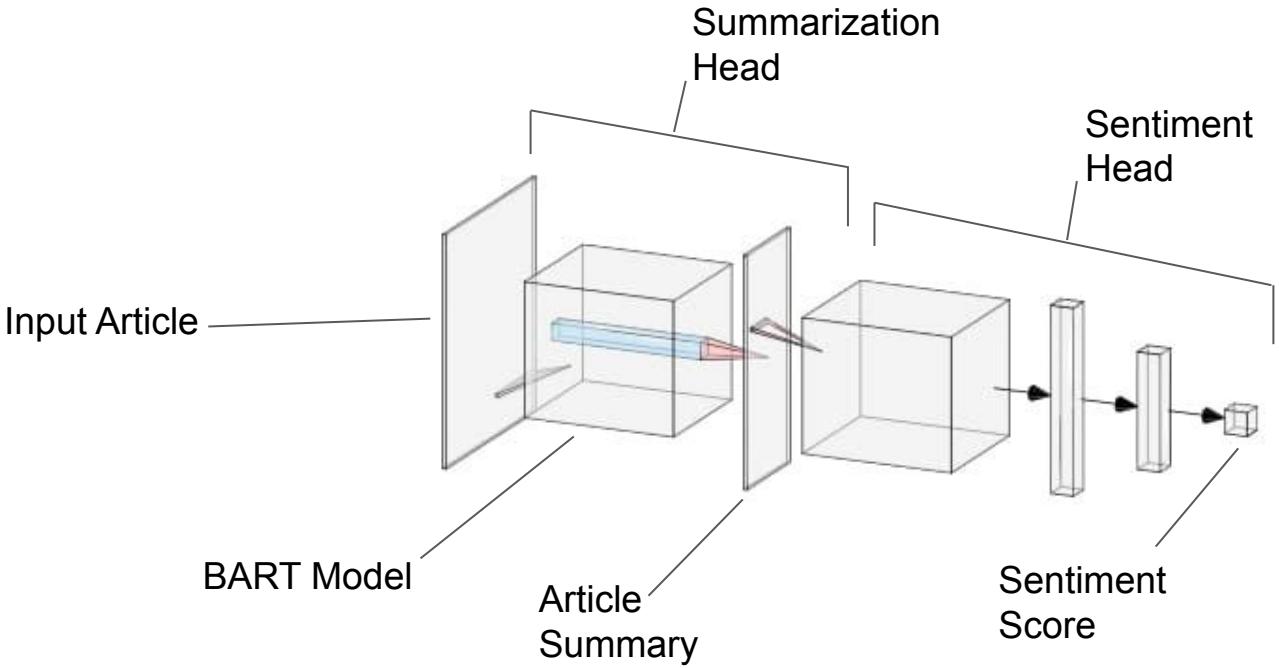


GPT

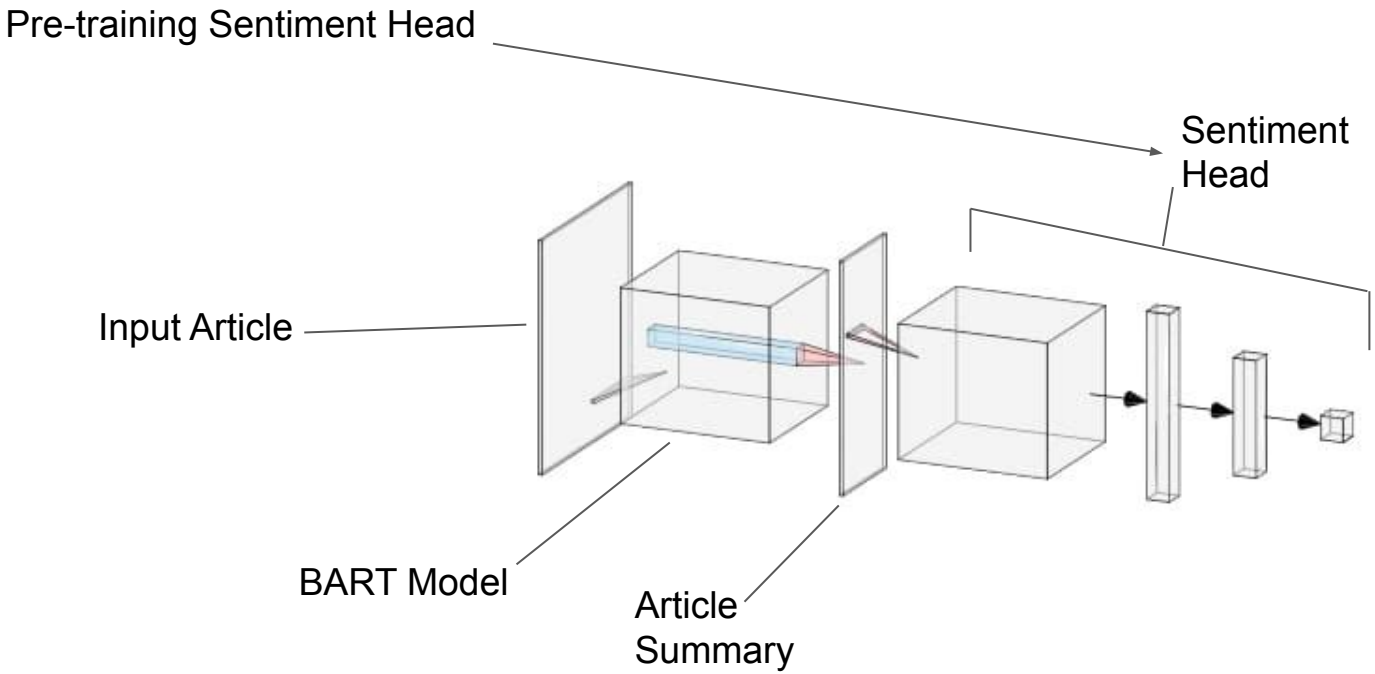


BART

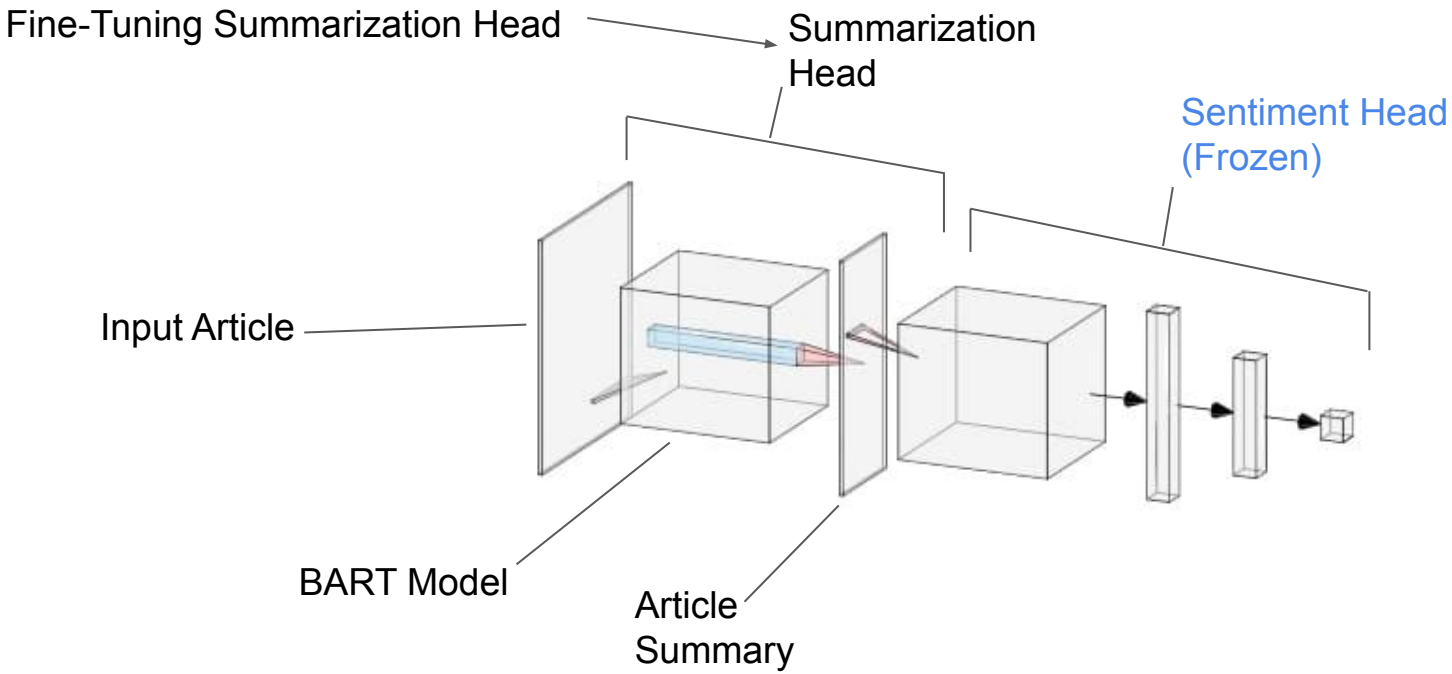
Model Architecture



Training - Part 1

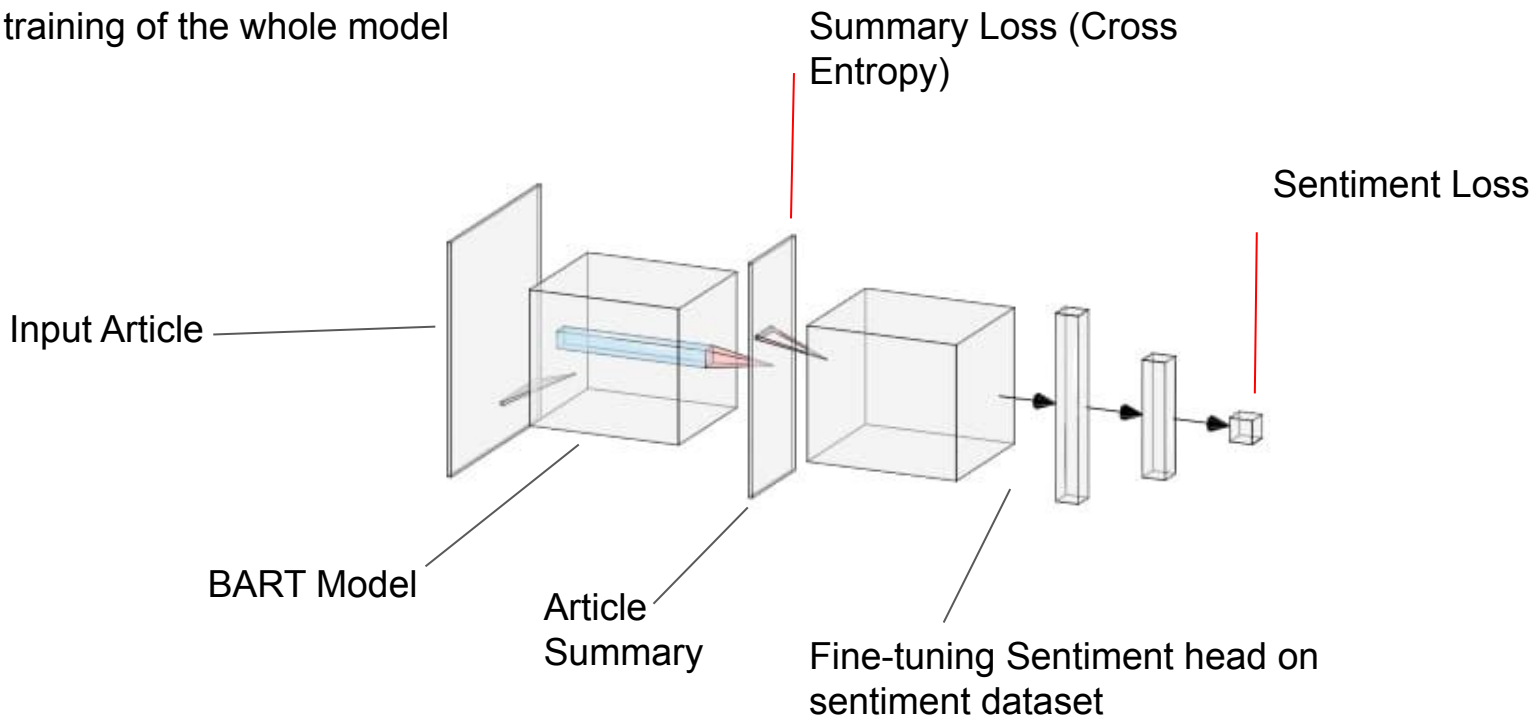


Training - Part 2



Training - Part 3

Joint training of the whole model



Sentiment Head

- Sentiment head acts as a discriminator that aids in the generation of summaries that are “sentimentally” consistent with the original article.

- We freeze the weights of the sentiment head else it will overfit to make the sentiment loss minimum.

Experimentation with losses

After we derive the sentiment scores of the article and the summary, we will experiment with the following losses defines over them,

1) MSE - The sentiment scores are vectors defined on R^K , where K is the number of sentiment classes. In our case, it is a positive or a negative sentiment. We want these vectors to be as close as possible to each other. Hence, we use MSE loss.
2) KL-Divergence - The sentiment scores are probability vectors over the sentiment classes, here positive or negative. We want the approximated probability density (via summaries) to match the true probability density over sentiment classes (via original articles). Hence, we also experiment with this loss.

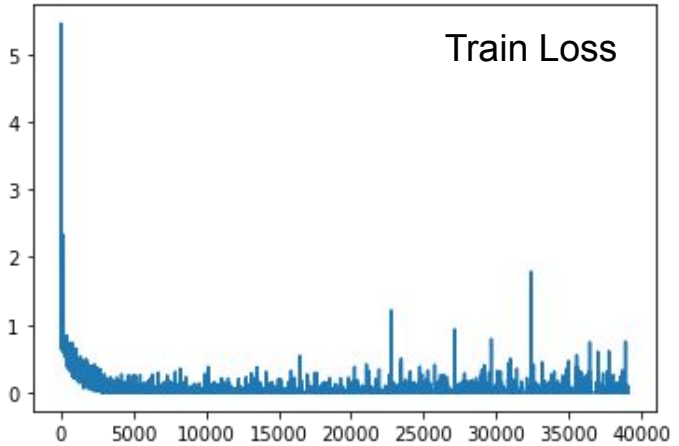
Code contributions to PyTorch

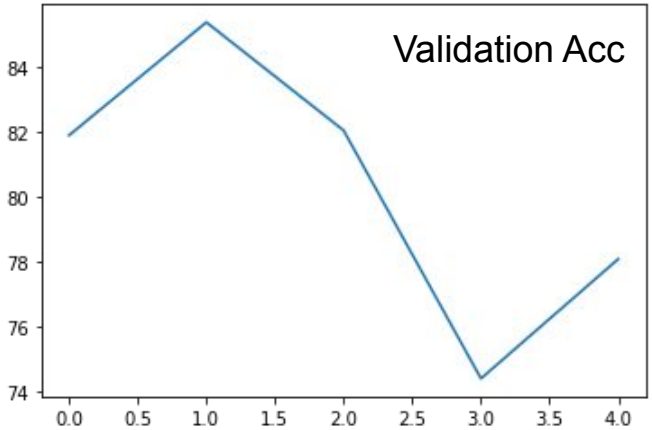
- We tackled the non-differentiability of the argmax function via implementing a custom function for argmax in PyTorch that is backward compatible. We derived inspiration from the Gumbel softmax trick.

- We cannot use the embedding layer from PyTorch in the middle layers of a model since it renders the model non-differentiable due to its lookup table implementation. We implemented a custom function that utilizes the nn.embedding implementation from the PyTorch but is backward compatible even though used in the middle layers of the model

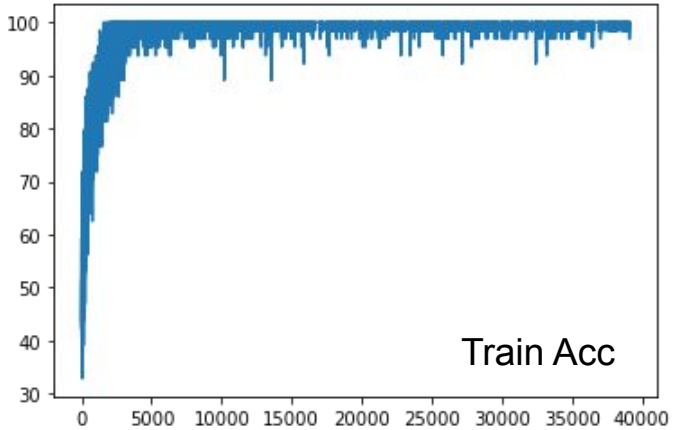
Preliminary Results

- Sentiment Classifier
  - Validation Accuracy **84%**





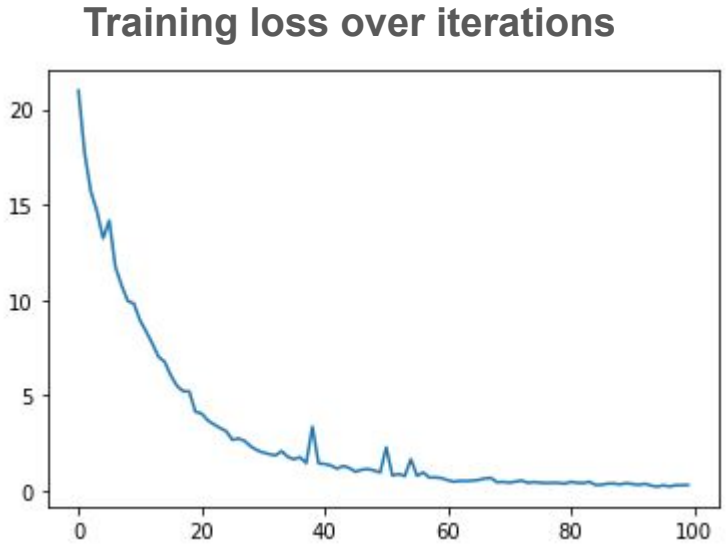
Epochs



Steps

BART Training

- Used Adam optimizer with weights decay, learning rate 1e-5.
- Fine tuning on pre-trained BART model by facebook. Transfer learning in NLP is possible after the introduction of BERT. BERT model is a part of BART.
- Needed 100 epochs to converge.



Article:

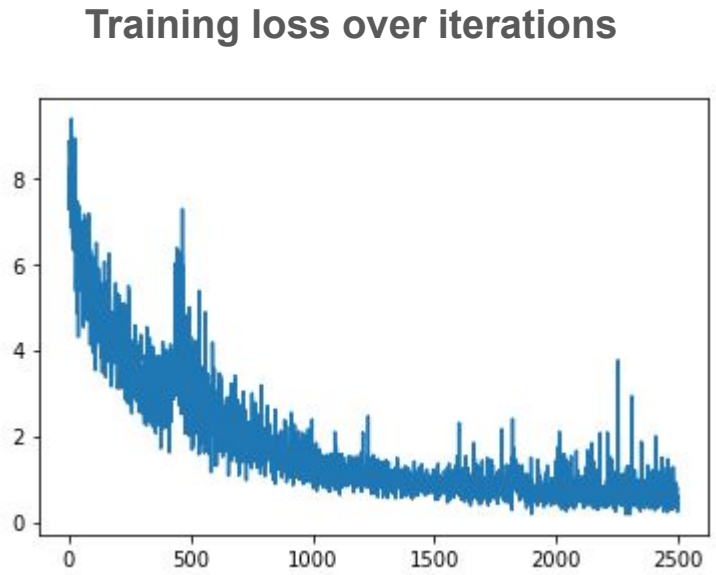
'(CNN)Arizona investigators have released dramatic video of a Walmart parking lot brawl that left a police officer wounded, one man dead, and reportedly involved members of a Christian family band. Enoch Gaver, 21, was killed in the fight in the town of Cottonwood, and suspect David Gaver, 28, was shot in the stomach and taken into custody. Police Sergeant Jeremy Daniels was hit in the leg by a bullet fired during the melee. The police dashcam video, released Friday, shows Cottonwood Police approaching the group of eight people -- all identified as members of the Gaver family -- around a large SUV in a Walmart parking lot on March 21. Officers wanted to question them about the alleged assault of a Walmart employee who was going into the store bathroom. The police were accompanied by another Walmart employee. On the video, an officer tells the group that they "need to separate these folks and talk to them." Someone then responds, "No, you are not going to separate me from my parents," and, "don\'t touch me." The video then shows a police officer being put in a headlock and knocked to the ground. The sound of Taser fire is heard. Police say pepper spray was deployed and that at least three shots were fired in an apparent struggle for an officer\'s gun. Several times on the video the group appears to surrender, but starts fighting again. The melee goes for several minutes until backup officers arrive and make arrests. Police charged four members of the family with assaulting an officer and resisting arrest. Two minors were also taken into custody and are being held at a juvenile detention facility. At least three members of the family are reportedly in a Christian band named "Matthew 24 Now," which is a Bible verse that refers to the end times, according to CNN affiliate KPHO. The family was living out of its Chevy Suburban. CNN\'s Greg Morrison contributed to this report.'

Summary:

'A man who was a member of the Gaver family was arrested, charged with assaulting a police officer . Police dashcam shows the group entering a Walmart parking lot and attacking a police employee .'

Merged model training with KL divergence loss

- Performed the experiment over 100 data-points.



Conclusion

- Introduced a Sentimentally consistent Summarization model
- Introduced a Differentiable embedding using softmax trick

Weighted KL