Improve Query-Focused Summarization using NLI

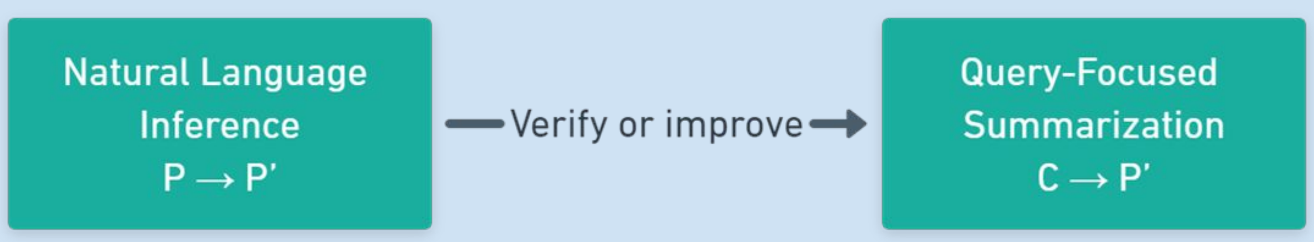
Team NLImitless

Chao Zhai, Dat Nguyen, Hitesh Manivannan, Lihan Zhu, Zhongheng He

USC Viterbi School of Engineering
Department of Computer Science

Motivation

- **Query-focused summarization** (QFS) is a special type of text summarization focusing on generating summaries conditioned to a specific query.
- Inspired by the work of Chen et al (2021), we try to improve QFS using NLI.
- **Natural language inference** (NLI) can be used to generate **an evaluation score** for the generated summary given the context and specific query, which can be used as a supervision signal when training the QFS model.



Dataset

- **Debatepedia**: an encyclopedia of pro and con arguments and quotes on critical debate topics.
- It contains 663 debates, from which 12695 **{query, document, summary}** triples are put into a dataset.

Average number of words per	Document	66.4
	Summary	11.16
	Query	9.97

Table 1: Summary of Debatepedia

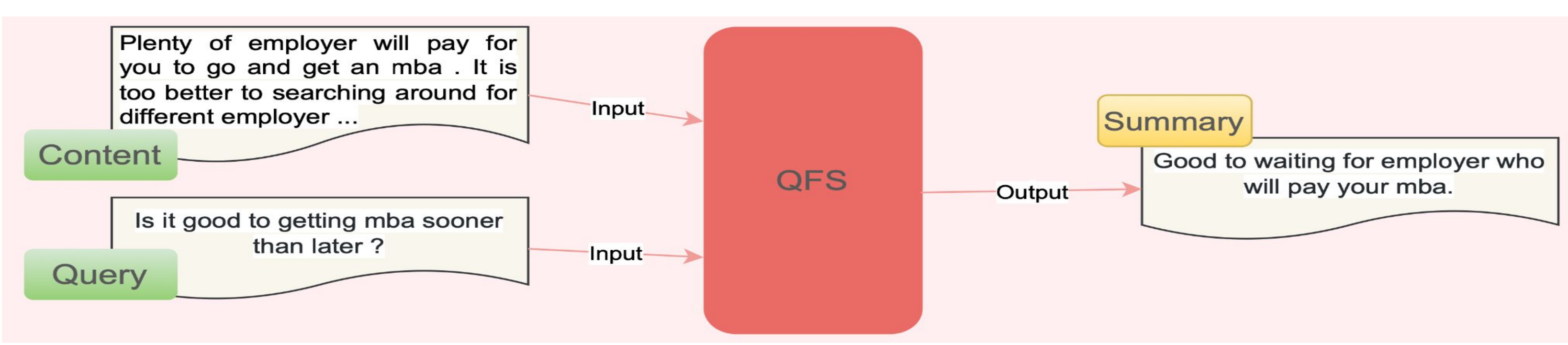
- **QMSum**: a new human-annotated benchmark for query-based multidomain meeting summarization task.
- It consists of 1,808 query-summary pairs over 232 meetings in multiple domains.

Average number of words per	Document	9590
	Summary	70
	Query	11

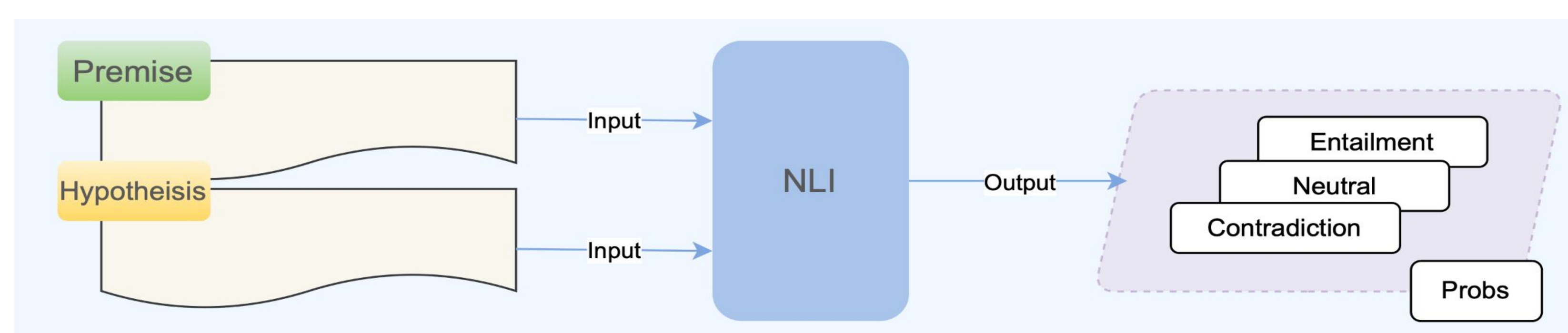
Table 2: Summary of QMSum

Method

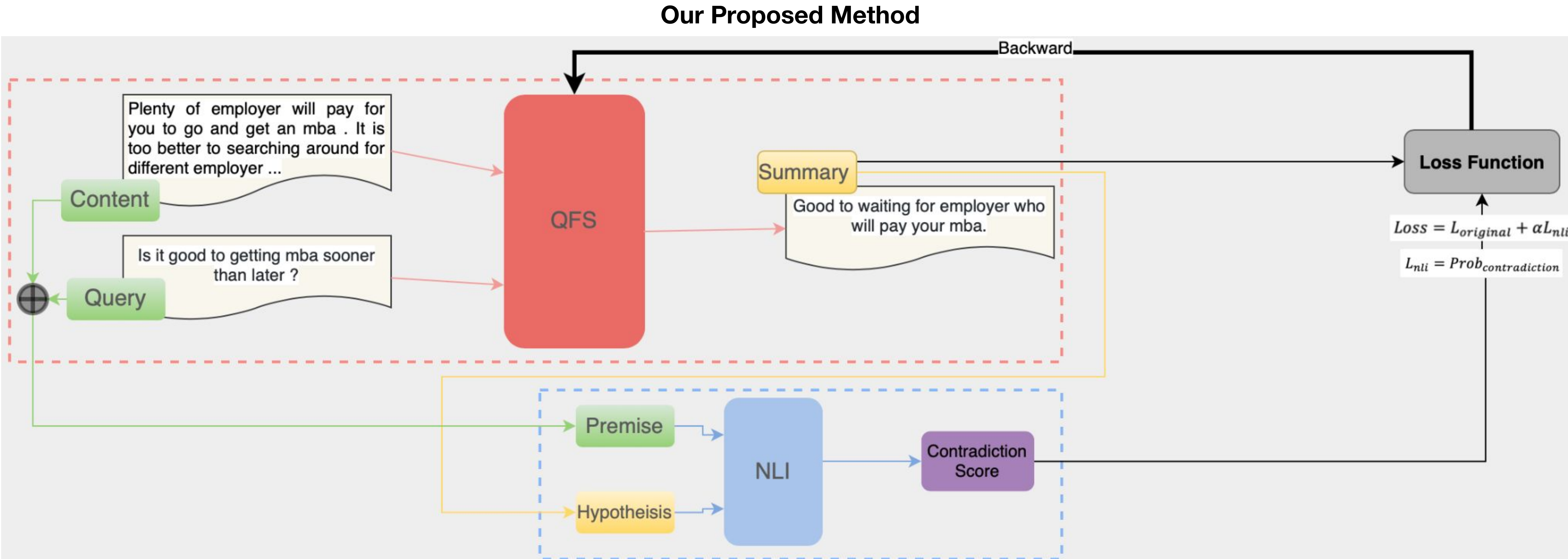
Query-Focused Summarization (QFS)



Natural Language Inference (NLI)



Our Proposed Method



Result

Dataset	Model	Rouge-1	Rouge-2	Rouge-L
Debate-pedia	Baseline	57.1623	45.3881	55.8273
	NLI loss	57.3192	45.6343	55.9098
QMSum	Baseline	31.6814	10.4602	21.0466
	NLI loss	31.9101	10.8455	21.6838

Table 3: Rouge scores using different NLI loss

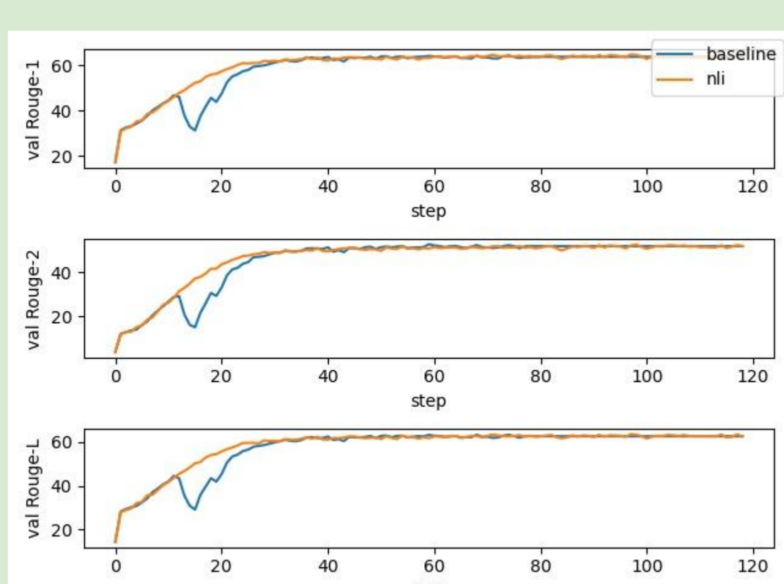


Figure 1: Rouge-2 curve of valid set

- NLI brings small improvements on all Rogue scores.
- These losses have similar performance on above metrics.
- Using NLI allows for a steadier learning curve.

Analysis

- Hyperparameter tuning seemed to converge at same point.
- We believe that the loss signal from NLI was not strong enough.

Avg Probability	entailment	neutral	contradiction
Groundtruth	0.4835	0.4003	**0.1383**
Baseline	0.4895	0.3660	0.1644
NLI loss	0.4963	0.3458	0.1579

Table 4: Distribution of entailment, neutral and contradiction samples, in Debatepedia

- Net contradiction reduce to 15.79%.
- Need stronger training samples of negative examples to train.
- Alternative method: Contrastive Learning
  - Pull a random sentence from Context as Negative Sample.

Future Works

- Reinforcement learning (Policy Gradient)
- Entailment Generation
- Data Augmentation: for testing on other datasets
- Expand dataset to include more negative samples