Distant Supervision for Question Summarization

Tatsuya Ishigaki[†], Kazuya Machida[†], Hayato Kobayashi[‡], Hiroya Takamura[†§], Manabu Okumura[†]
[†]Tokyo Institute of Technology / [‡]Yahoo Japan Corporation / [§]AIST

ECIR2020

1. Introduction

Motivation:
Questions tend to be lengthy and hard to understand.
We aim to convert them easy-to-understand shorter questions.

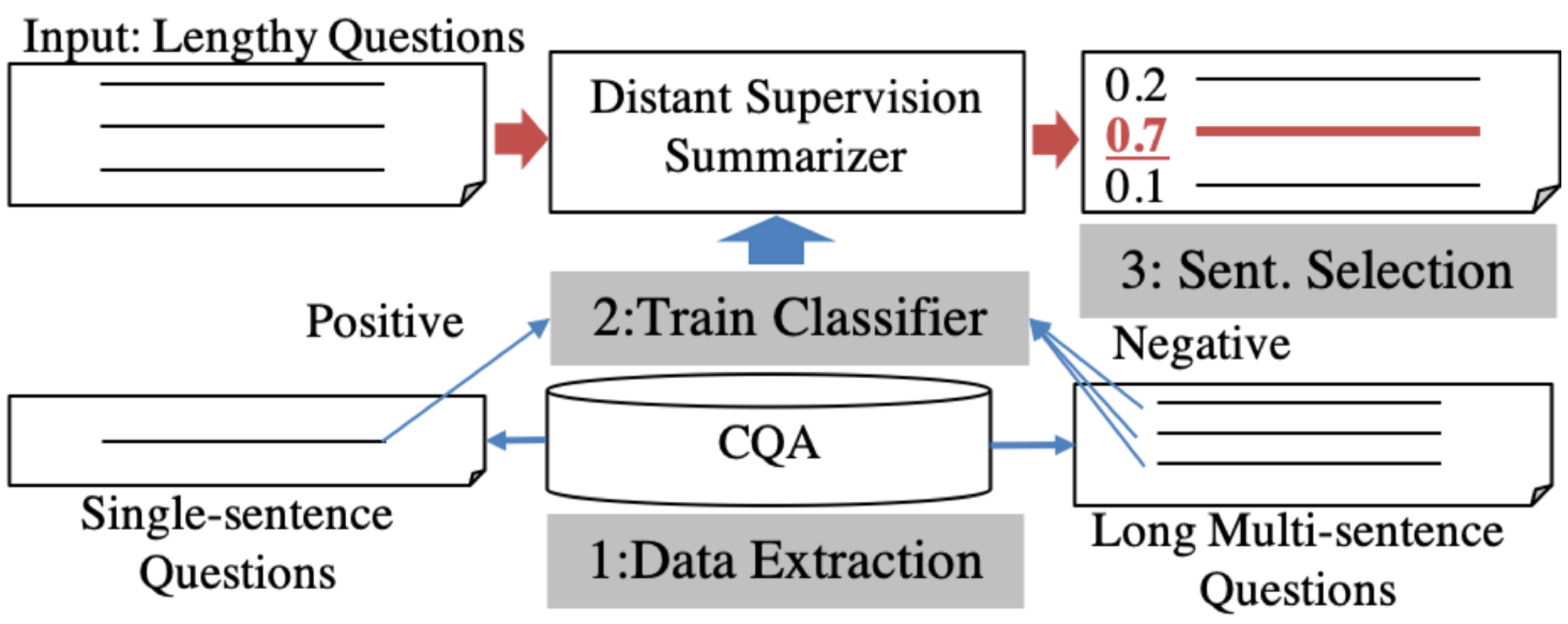
Task: Extractive Question Summarization
Input : multi-sentence question
Output : extracted single-sentence summary

Existing Approaches (Extractive):
Supervised: - Classification/Regression [Ishigaki+,2017, Tamura+2007]
- learning-to-rank [Higurashi+,2018]
→ Supervised methods require costly labeled data
Unsupervised: - Graph-based (e.g. LexRank) [Erkan+2004]
- Semantic similarity [Kobayashi+,2018]
→ Major unsupervised methods do not perform well (See our experiments.)

Our Approach:
This paper describes a distant supervision that creates pseudo labeled data for training a summarizer w/o labeled data.

Contributions:
1. We propose a distant supervision approach to create a pseudo labeled data for training a question summarizer.
2. Our models w/o any supervision performs competitively with respect to supervised models.
3. We release a large dataset including 2.5M sentences with pseudo labels.

2. Proposed Framework

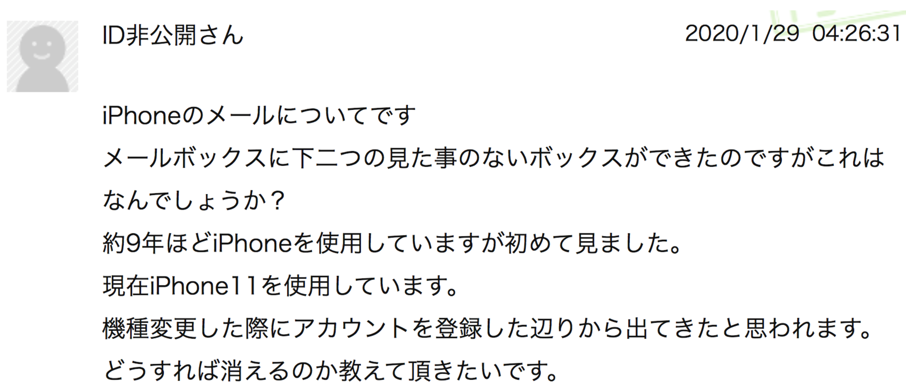


1. Data Extraction
We extracted 2.5M sentences from a corpus of CQA.
All sentences are labeled by our proposed heuristics:
Pseudo positive labels : single-sentence questions.


au_********さん 2020/1/29 01:10:49
iPadpro11inchのSIMカードをiPhoneに差し込んだら使えますか？

Single-sentence questions have summary-like properties: basically they are self-contained questions.
(= similar to ones that we want to include in the summary).

Pseudo negative labels : individual sentences extracted from extremely long post.


ID非公開さん 2020/1/29 04:26:31
iPhoneのメールについてです
メールボックスに下二つの見た事のないボックスができたのですがこれは
なんでしょうか？
約9年ほどiPhoneを使用していますが初めて見ました。
現在iPhone11を使用しています。
機種変更した際にアカウントを登録した辺りから出てきたと思われます。
どうすれば消えるのか教えて頂きたいです。

Individual sentences in long post are not summary-like: basically they are not self-contained and often not a question.
(=we need information from other sentences to understand.)

2. Train Classifier
We trained a binary classifier that outputs a score that represents how likely the sentence is summary-like.
3. Sentence Selection
We score every sentence in an input. We propose several sentence selection strategies that use the scores as explained in Sec.3.

3. Experiment

Datasets:
1. Dataset with pseudo labels (2.5M sentences)
 - Labeled data created by our framework.
2. Dataset with manually annotated labels (10K sentences)
 - We used a crowdsourcing to annotate the sentences.

Compared Models:
- Our Models (trained on our data with pseudo labels)
 - DistNet: NN-based sentence tagger (LSTM + Softmax)
 - DistReg: Logistic Regression with N-gram, POS features.
- Unsupervised Models
 - Lead : Simply selects the initial sentence.
 - LexRank: A graph-based algorithm for sentence selection.
 - SimEmb: Selects the sentence that has the minimum Word Movers' Distanse from the input.
 - TfIdf : Selects the sentence that has the highes Tf-Id in the input.
- Supervised Models (trained on the manually annotated data)
 - SupNet: NN-based sentence tagger (LSTM + Softmax)
 - SupReg: Logistic Regression with N-gram, POS features.

Sentence Selection Strategies:
- Greedy: Simply selects the highest scored sentence.
- Init : Selects the initial sentnece that has higher score than a specific threthold (tuned on validation data.)
- Q : Selects the highest scored question sentence.

4. Result

Accuracy = correctly selected sentences / total sentences.

	Greedy	Init	Q	Best
DistNet	**87.38**	**90.45**	**87.38**	**90.45**
DistReg	86.17	89.05	86.17	89.05
Lead	81.79	81.79	88.08	88.08
LexRank	78.49	81.79	84.95	84.95
SimEmb	59.46	81.79	71.17	81.79
TfIdf	52.03	81.79	69.68	81.79
SupNet	81.67	86.31	81.67	86.31
SupReg	87.89	91.21	87.89	91.21

- Our distant supervision approach outperformed all unsupervised baselines.
- Using our pseudo data improved the performance of NN-based approach (DistNet).
- There is no statistically significant difference between the best performed model of our distant supervision approach and the best model of supervised models.

5. Conclusion

- We proposed a distant supervision for extractive summarization task.
- Our approach outperformed unsupervised baselines and performed competitively with supervised baselines.
- The data is publicly available: http://lr-www.pi.titech.ac.jp/~ishigaki/chiebukuro/