HIT-SCIR

Harbin Institute of Technology
Research Center for
Social Computing and Information Retrieval

Sequence-to-Sequence Data Augmentation for Dialogue Language Understanding

Yutai Hou, Yijia Liu, Wanxiang Che, Ting Liu

Presenter: Yutai Hou

July 25th, 2018

Beijing, China

Background

Task-oriented Dialogue

- Task-based systems help users achieve goals such as finding restaurants or booking flights.

Language Understanding

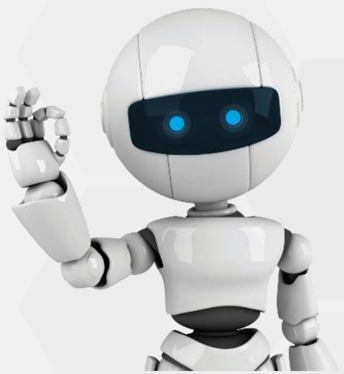
- Language understanding (LU), such as slot filling, is the initial and essential component in the task-oriented dialogue system pipeline (Young et al., 2013).
- Example of slot filling:

User Utterance

show me the closest restaurant.

Slot Filling

show me the [closest]<distance> [restaurant]<poi type>



Challenges

- Handling myriad ways in which users express demands is hard.

  Example of multiple expression for same semantic frame

  – Find me the <distance> route to <poi type>

  – Is there a <distance> <poi type>

  – I'm desiring to eat at some <poi type> is there any in <distance>

  – Give me the <distance> route to <poi type>

- Large-scale labeled data is usually unreachable for new domain.

Solution

Data Augmentation

- Enlarging the size of training data in machine learning systems.
- Data augmentation has been success on a wide range of problem:
  - computer vision (Krizhevsky et al., 2012)
  - speech recognition (Hannun et al., 2014)
  - text classification (Zhang et al., 2015)
  - question answering (Fader et al., 2013)
- However, its application in the task-oriented dialogue system is less studied.
  - The only work on this task (Kurata et al. 2016) augments one single utterance by adding noise
  - Kurata et al. (2016) didn't consider its relation with other utterances.

Our Solution

- Augments one single utterance by considering other utterance
- Add diversity rank into utterance representation to encourage diversity

HIT-SCIR

Harbin Institute of Technology
Research Center for
Social Computing and Information Retrieval

2

Overview

What we did & What we achieved

Overview

- **Task:**
  - We study the problem of data augmentation for language understanding in task-oriented dialogue system.
- **Idea:**
  - ☐ Previous work which augments an utterance without considering its relation with other utterances,
  - ✓ we propose a Seq2Seq generation based data augmentation framework that leverages one utterance's same semantic alternatives in the training data.
  - ☐ Diversity is important to improve augmentation effect
  - ✓ We incorporated a novel diversity rank into the utterance representation to make the model produce diverse utterances
- **Result:**
  - **6.38↑** F-scores on ATIS Dataset (small set)
  - **10.04↑** F-scores on Stanford Multi-Domain-Multi-Turn Dataset(small set)

Overview

- **Task:**
  - We study the problem of data augmentation for language understanding in task-oriented dialogue system.

- **Idea:**
  - ☐ Previous work which augments an utterance without considering its relation with other utterances,
  - ✓ we propose a Seq2Seq generation based data augmentation framework that leverages one utterance's same semantic alternatives in the training data.

  - ☐ Diversity is important to improve augmentation effect
  - ✓ We incorporated a novel diversity rank into the utterance representation to make the model produce diverse utterances

- **Result:**
  - **6.38↑** F-scores on ATIS Dataset (small set)
  - **10.04↑** F-scores on Stanford Multi-Domain-Multi-Turn Dataset(small set)

Overview

- **Task:**
  - We study the problem of data augmentation for language understanding in task-oriented dialogue system.

- **Idea:**
  - ☐ Previous work which augments an utterance without considering its relation with other utterances,
  - ✓ we propose a Seq2Seq generation based data augmentation framework that leverages one utterance's same semantic alternatives in the training data.

  - ☐ Diversity is important to improve augmentation effect
  - ✓ We incorporated a novel diversity rank into the utterance representation to make the model produce diverse utterances

- **Result:**
  - **6.38↑** F-scores on ATIS Dataset (small set)
  - **10.04↑** F-scores on Stanford Multi-Domain-Multi-Turn Dataset(small set)

HIT-SCIR

Harbin Institute of Technology
Research Center for
Social Computing and Information Retrieval

3 Method

Detail of our idea

Data Augmentation Process

One Utterance

show me the [closest]<distance> [restaurant]<poi type>

Delexicalisation

show me the <distance> <poi type>

Diverse Ranks Incorporation

show me the <distance> <poi type> #1
show me the <distance> <poi type> #2
show me the <distance> <poi type> #3

Seq2Seq Generation

where is the <distance> <poi type>
can you find the <distance> <poi type> to me
give me the address to the <distance> <poi type>

Surface Realisation

Augmented Utterance

where is the nearest shopping mall
can you find the nearest rest stop to me
give me the address to the near grocery

Seq2Seq Model Training Process

Build Training Set: match utterances with same semantic frame

find me the <distance> route to <poi type>
(5.0) is there a <distance> <poi type>
(1.0) give me the <distance> route to <poi type>
(4.4) I'm desiring to eat at some <poi type> is there any in <distance>

Ranking Candidates by Diversity Score

find me the <distance> route to <poi type>
#1: is there a <distance> <poi type>
#2: I'm desiring to eat at some <poi type> is there any in <distance>
#3: give me the <distance> route to <poi type>

Filtering and Generating "translation" Pairs

find me the <distance> route to <poi type> #1
→ is there a <distance> <poi type>
find me the <distance> route to <poi type> #2
→ I'm desiring to eat at some <poi type> is there any in <distance>

Training model with filtered pairs

Trained Seq2Seq Model

Details

- **Seq2Seq Generation with Diversity Ranks**
  - For one instance (u, s), we first collect Cs.
  - Then rank each instance (u’, s) ∈ Cs/{(u, s)} by its diversity score against u.
  - The model is formalized as

$$p(\boldsymbol{d}'|\boldsymbol{d},k) = \prod_t p(d'_t|d_1,\ldots,d_n,\#k,d'_1,\ldots,d'_{t-1})$$

- **Diversity score**
  - Diversity score of an utterance pair (u, u’) is calculated by both considering the edit distance and a length difference penalty (LDP) as:

$$SCORE(\boldsymbol{u},\boldsymbol{u}') = EditDistance(\boldsymbol{u},\boldsymbol{u}') \times LDP(\boldsymbol{u},\boldsymbol{u}')$$

  - where LDP is defined as $LDP(\boldsymbol{u},\boldsymbol{u}') = e^{-\frac{|\|\boldsymbol{u}\| - \|\boldsymbol{u}'\||}{\|\boldsymbol{u}\|}}$

4 Experiments

Dataset

- ATIS Dataset
  - The ATIS dataset contains
    - 4978 training utterances
    - 893 testing utterances
  - To simulate the data insufficient situations, we follow Chen et al. (2016) and also evaluate our model on two small proportions of the training data:
    - small (1/40 of the original training set with 129 instances) proportion
    - medium (1/10 of the original training set with 515 instances).
    - In all the experiments, a development set of 500 instances is used.

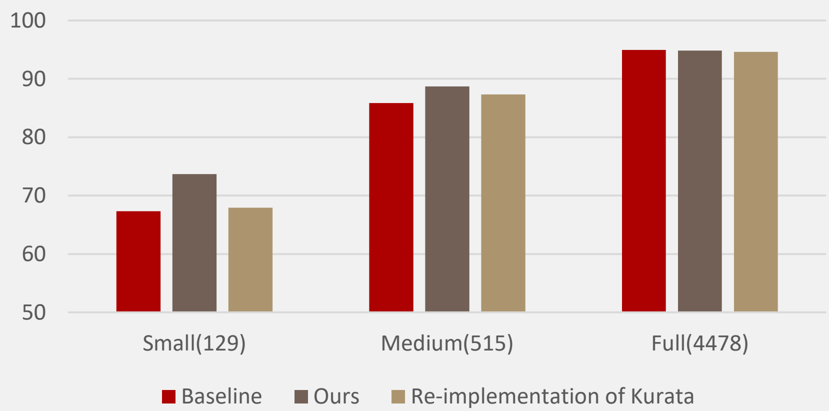
- Stanford Labeled (500 u * 3 d)

	Navigation	Scheduling	Weather
# of training utterances	500	500	500
# of devel. utterances	321	201	262
# of test utterances	337	212	271
Kappa	0.68	0.92	0.90
Agreement	85.05	90.75	95.99

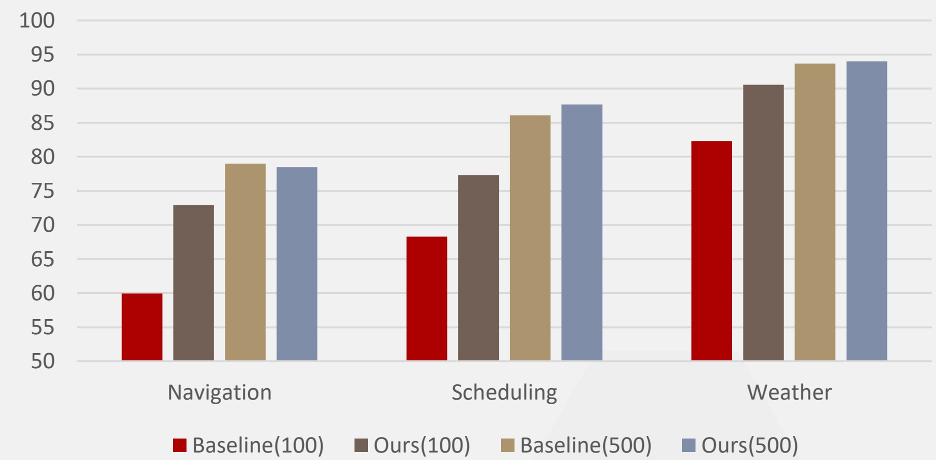
Table 1: Statistics for our annotation.

Main Experiment Setting

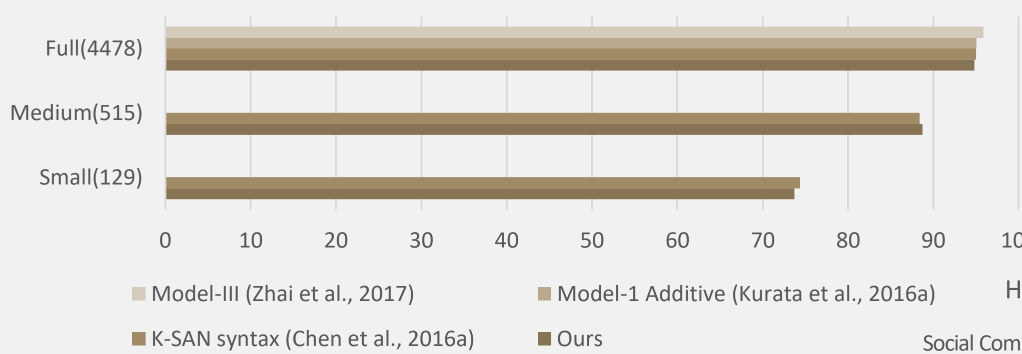
Results on ATIS



Results on Stanford dialogue dataset

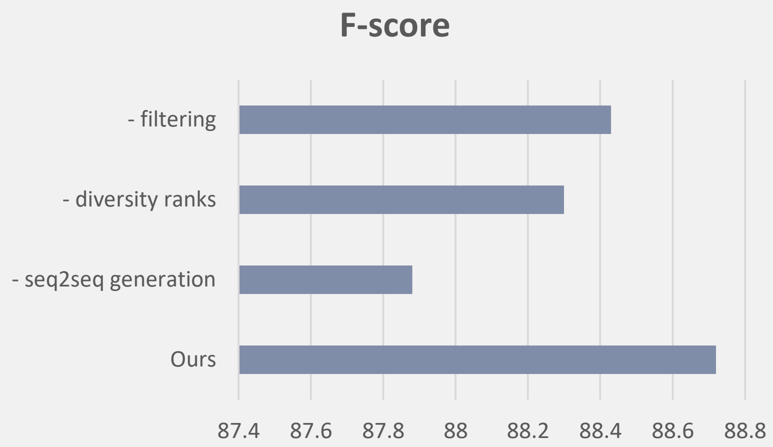


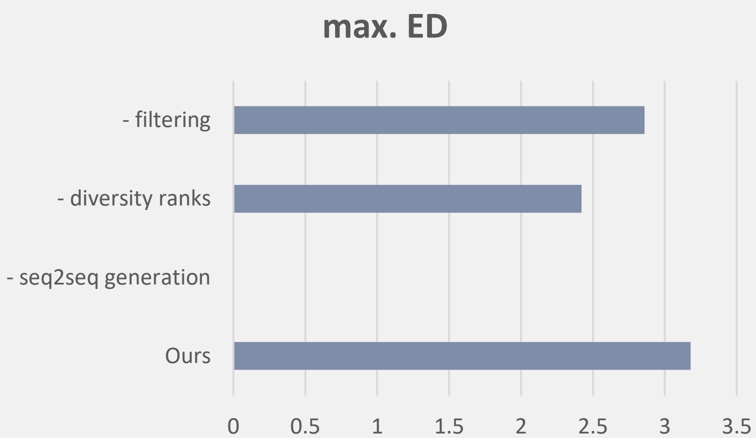
Comparing to ATIS State-of-Art

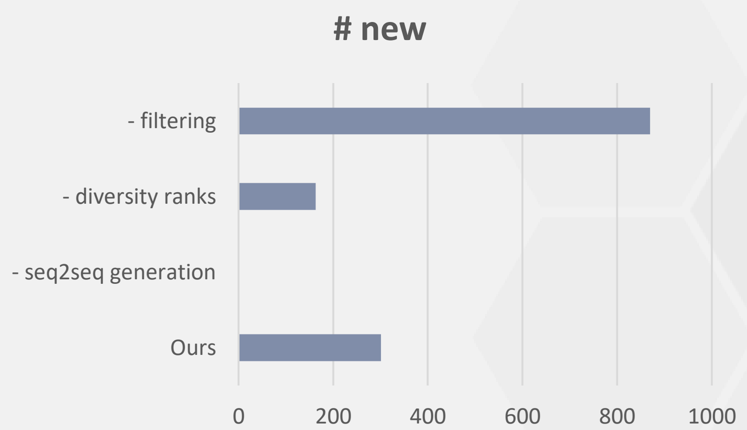


Analysis

Ablation Test







Analysis

- Effect of Training Data Size

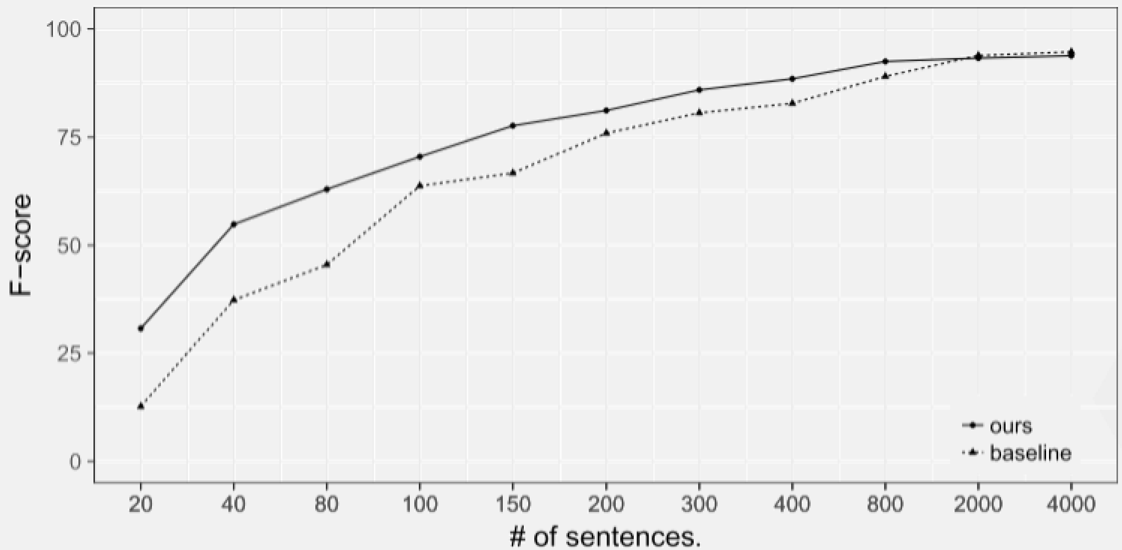


Figure 2: Our method's performances on the ATIS training data of different sizes.

Case Study

Case 1

show me all flights from atlanta to washington with prices		
(delex.) *show me all flights from <from_city> to <to_city> with prices*		
#1	train	let 's look at <from_city> to <to_city> again
	ours	what are all the flights between <from_city> and <to_city>
		(realized) what are all the flights between indianapolis and tampa
#100	train	list types of aircraft that fly between <from_city> and <to_city>
	ours	i 'm looking for a flight from <from_city> to <to_city>
		(realized) i 'm looking for a flight from milwaukee to los angeles
	Kurata16	show me all flights from [atlanta]<from_city> to [washington]<to_city> with airports

Case 2

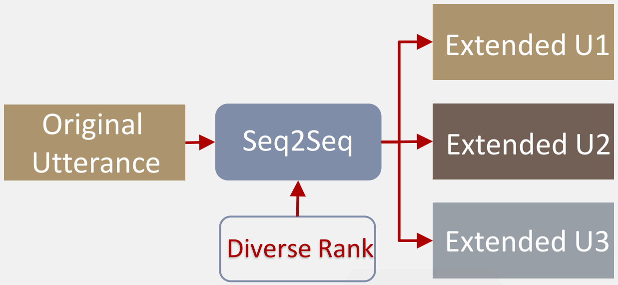
is there a flight between san francisco and boston with a stopover at dallas fort worth		
(delex.) *is there a flight between <from_city> and <to_city> with a stopover at <stop_city>*		
#1	train	which airlines fly from <from_city> to <to_city> and have a stopover in <stop_city>
	ours	is there a flight from <from_city> to<to_city> with a stop in <stop_city>
		(realized) is there a flight from washington to miami with a stop in dallas fort worth
#30	train	do you have any airlines that would stop at <stop_city> on the way from <from_city> to <to_city>
	ours	i 'd like to fly from <from_city> to <to_city> with a stop in <stop_city>
		(realized) i 'd like to fly from memphis to boston with a stop in minneapolis
	Kurata16	is there a flight between [san francisco]<from_city> and [boston]<to_city> with a stopover at [dallas fort worth]<to_city>

5 Conclusion

Conclusion

- Problem
  - Data scarcity of new domain can hinder language understanding

- Method
  - Seq2seq data augmentation framework for LU
  - A novel diversity rank is used to encourage diversity



- Results
  - **6.38↑** F-scores on ATIS Dataset (small set)
  - **10.04↑** F-scores on Stanford Dataset(small set)

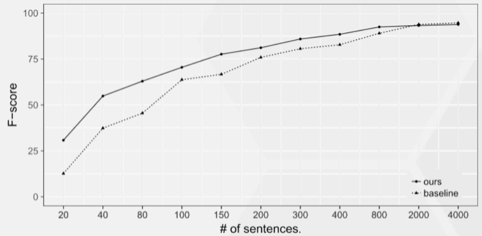


Figure 2: Our method's performances on the ATIS training data of different sizes.

- Code available at:

  https://github.com/AtmaHou/Seq2SeqDataAugmentationForLU

Thanks !

Harbin Institute of Technology
Research Center for
Social Computing and Information Retrieval