Learning a Matching Model with Co-teaching for Multi-turn Response Selection in Retrieval-based Dialogue Systems

Jiazhan Feng[1*], **Chongyang Tao**[1*], **Wei Wu**[2], **Yansong Feng**[1], **Dongyan Zhao**[1,3] and **Rui Yan**[1,3†]

[1]Institute of Computer Science and Technology, Peking University, Beijing, China
[2]Microsoft Corporation, Beijing, China
[3]Center for Data Science, Peking University, Beijing, China
[1]fengjiazhan@foxmail.com [2]wuwei@microsoft.com
[1,3]{chongyangtao,fengyansong,zhaody,ruiyan}@pku.edu.cn

Abstract

We study learning of a matching model for response selection in retrieval-based dialogue systems. The problem is equally important with designing the architecture of a model, but is less explored in existing literature. To learn a robust matching model from noisy training data, we propose a general co-teaching framework with three specific teaching strategies that cover both teaching with loss functions and teaching with data curriculum. Under the framework, we simultaneously learn two matching models with independent training sets. In each iteration, one model transfers the knowledge learned from its training set to the other model, and at the same time receives the guide from the other model on how to overcome noise in training. Through being both a teacher and a student, the two models learn from each other and get improved together. Evaluation results on two public data sets indicate that the proposed learning approach can generally and significantly improve the performance of existing matching models.

1 Introduction

Human-machine conversation is a long-standing goal of artificial intelligence. Recently, building a dialogue system for open domain human-machine conversation is attracting more and more attention due to both availability of large-scale human conversation data and powerful models learned with neural networks. Existing methods are either retrieval-based or generation-based. Retrieval-based methods reply to a human input by selecting a proper response from a pre-built index (Ji et al., 2014; Zhou et al., 2018b; Yan and Zhao, 2018), while generation-based methods synthesize a response with a natural language model (Shang et al., 2015; Serban et al., 2017). In this work, we study the problem of response selection for retrieval-based dialogue systems, since retrieval-based systems are often superior to their generation-based counterparts on response fluency and diversity, are easy to evaluate, and have powered some real products such as the social bot XiaoIce from Microsoft (Shum et al., 2018), and the E-commerce assistant AliMe Assist from Alibaba Group (Li et al., 2017).

A key problem in response selection is how to measure the matching degree between a conversation context (a message with several turns of conversation history) and a response candidate. Existing studies have paid tremendous effort to build a matching model with neural architectures (Lowe et al., 2015; Zhou et al., 2016; Wu et al., 2017; Zhou et al., 2018b), and advanced models such as the deep attention matching network (DAM) (Zhou et al., 2018b) have achieved impressive performance on benchmarks. In contrary to the progress on model architectures, there is little exploration on learning approaches of the models. On the one hand, neural matching models are becoming more and more complicated; on the other hand, all models are simply learned by distinguishing human responses from some automatically constructed negative response candidates (e.g., by random sampling). Although this heuristic approach can avoid expensive and exhausting human labeling, it suffers from noise in training data, as many negative examples are actually false negatives[1]. As a result, when evaluating a well-trained model using human judgment, one can often observe a significant gap between training and test, as will be seen in our experiments.

In this paper, instead of configuring new architectures, we investigate how to effectively learn existing matching models from noisy training

[*]Equal Contribution.
[†]Corresponding author: Rui Yan (ruiyan@pku.edu.cn).

[1]Responses sampled from other contexts may also be proper candidates for a given context.

data, given that human labeling is infeasible in practice. We propose learning a matching model under a general co-teaching framework. The framework maintains two peer models on two i.i.d. training sets, and lets the two models teach each other during learning. One model transfers knowledge learned from its training set to its peer model to help it combat with noise in training, and at the same time gets updated under the guide of its peer model. Through playing both a role of a teacher and a role of a student, the two peer models evolve together. Under the framework, we consider three teaching strategies including teaching with dynamic margins, teaching with dynamic instance weighting, and teaching with dynamic data curriculum. The first two strategies let the two peer models mutually "label" their training examples, and transfer the soft labels from one model to the other through loss functions; while in the last strategy, the two peer models directly select training examples for each other.

To examine if the proposed learning approach can generally bridge the gap between training and test, we select sequential matching network (SMN) (Wu et al., 2017) and DAM as representative matching models, and conduct experiments on two public data sets with human judged test examples. The first data set is the Douban Conversation benchmark published in Wu et al. (2017), and the second one is the E-commerce Dialogue Corpus published in Zhang et al. (2018b) where we recruit human annotators to judge the appropriateness of response candidates regarding to their contexts on the entire test set[2]. Evaluation results indicate that co-teaching with the three strategies can consistently improve the performance of both matching models over all metrics on both data sets with significant margins. On the Douban data, the most effective strategy is teaching with dynamic margins that brings 2.8% absolute improvement to SMN and 2.5% absolute improvement to DAM on P@1; while on the E-commerce data, the best strategy is teaching with dynamic data curriculum that brings 2.4% absolute improvement to SMN and 3.2% absolute improvement to DAM on P@1. Through further analysis, we also unveil how the peer models get evolved together in learning and how the choice of peer models affects the performance of learning.

Our contributions in the paper are four-folds: (1) proposal of learning matching models for response selection with a general co-teaching framework; (2) proposal of two new teaching strategies as special cases of the framework; and (3) empirical verification of the effectiveness of the proposed learning approach on two public data sets.

2 Problem Formalization

Given a data set $\mathcal{D} = \{(y_i, c_i, r_i)\}_{i=1}^{N}$ where c_i represents a conversation context, r_i is a response candidate, and $y_i \in \{0, 1\}$ denotes a label with $y_i = 1$ indicating r_i a proper response for c_i and otherwise $y_i = 0$, the goal of the task of response selection is to learn a matching model $s(\cdot, \cdot)$ from $\mathcal{D}$. For any context-response pair (c, r), $s(c, r)$ gives a score that reflects the matching degree between c and r, and thus allows one to rank a set of response candidates according to the scores for response selection.

To obtain a matching model $s(\cdot, \cdot)$, one needs to deal with two problems: (1) how to define $s(\cdot, \cdot)$; and (2) how to learn $s(\cdot, \cdot)$. Existing studies concentrate on Problem (1) by defining $s(\cdot, \cdot)$ with sophisticated neural architectures (Wu et al., 2017; Zhou et al., 2018b), and leave Problem (2) in a simple default setting where $s(\cdot, \cdot)$ is optimized with $\mathcal{D}$ using a loss function L usually defined by cross entropy. Ideally, when $\mathcal{D}$ is large enough and has good enough quality, a carefully designed $s(\cdot, \cdot)$ learned using the existing paradigm should be able to well capture the semantics in dialogues. The fact is that since large-scale human labeling is infeasible, $\mathcal{D}$ is established under simple heuristics where negative response candidates are automatically constructed (e.g., by random sampling) with a lot of noise. As a result, advanced matching models only have sub-optimal performance in practice. The gap between ideal and reality motivates us to pursue a better learning approach, as will be presented in the next section.

3 Learning a Matching Model through Co-teaching

In this section, we present co-teaching, a new framework for learning a matching model. We first give a general description of the framework, and then elaborate three teaching strategies as special cases of the framework.

[2]We have released labeled test data of E-commerce Dialogue Corpus at https://drive.google.com/open?id=1HMDHRU8kbbWTsPVr6lKU_-Z2Jt-n-dys.

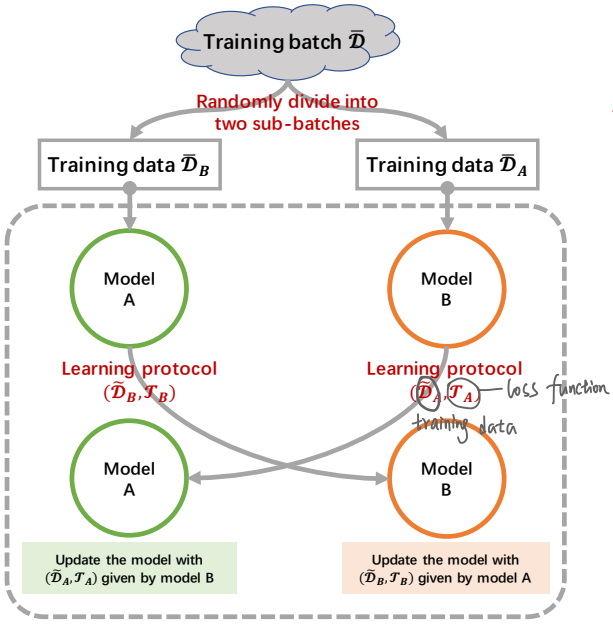


Figure 1: Co-teaching framework.

3.1 Co-teaching Framework

The idea of co-teaching is to maintain two peer models and let them learn from each other by simultaneously acting as a teacher and a student.

Figure 1 gives an overview of the co-teaching framework. The learning program starts from two pre-trained peer models A and B. In each iteration, a batch of training data is equally divided into two sub-batches without overlap as $\bar{\mathcal{D}}_A$ and $\bar{\mathcal{D}}_B$ for B and A respectively. A and B then examine their sub-batches and output learning protocols $(\tilde{\mathcal{D}}_B, \mathcal{J}_B)$ and $(\tilde{\mathcal{D}}_A, \mathcal{J}_A)$ for their peers, where $\tilde{\mathcal{D}}_B$ and $\tilde{\mathcal{D}}_A$ are training data and $\mathcal{J}_B$ and $\mathcal{J}_A$ are loss functions. After that, A and B get updated according to $(\tilde{\mathcal{D}}_A, \mathcal{J}_A)$ and $(\tilde{\mathcal{D}}_B, \mathcal{J}_B)$ respectively, and the learning program moves to the next iteration. Algorithm 1 describes the pseudo code of co-teaching.

The rationale behind the co-teaching framework is that the peer models can gradually obtain different abilities from the different training data as the learning process goes on, even when the two models share the same architecture and the same initial configuration, and thus, they can acquire different knowledge from their training data and transfer the knowledge to their peers to make them robust over the noise in the data. This resembles two peer students who learn from different but related materials. Through knowledge exchange, one can inspire the other to get new insights from his or her material, and thus the two students get improved together. Advantages of the framework reside in various aspects: first, the peer models have their own "judgment" regarding to the quality of the same training example. Thus, one model may guide the other how to pick high quality training examples and circumvent noise; second, since the peer models are optimized with different training sub-batches, knowledge from one sub-batch could be supplementary to the other through exchange of learning protocols; third, the two peer models may have different decision boundaries, and thus are good at recognizing different patterns in data. This may allow one model to help the other rectify errors in learning.

To instantiate the co-teaching framework, one needs to specify initialization of the peer models and teaching strategies that can form the learning protocols. In this work, to simplify the learning program of co-teaching, we assume that model A and model B are initialized by the same matching model pre-trained with the entire training data. We focus on design of teaching strategies, as will be elaborated in the next section.

3.2 Teaching Strategies

We consider the following three strategies that cover teaching with dynamic loss functions and teaching with data curriculum.

Teaching with Dynamic Margins: The strategy fixes $\bar{\mathcal{D}}_A$ and $\bar{\mathcal{D}}_B$ as $\tilde{\mathcal{D}}_A$ and $\tilde{\mathcal{D}}_B$ respectively, and dynamically creates loss functions as the learning protocols. Without loss of generality, the training data $\mathcal{D}$ can be re-organized in a form of $\{(c_i, r_i^+, r_i^-)\}_{i=1}^{N'}$, where r_i^+ and r_i^- refer to a positive response candidate and a negative response candidate regarding to c_i respectively. Suppose that $\bar{\mathcal{D}}_A = \{(c_{A,i}, r_{A,i}^+, r_{A,i}^-)\}_{i=1}^{N_A}$ and $\bar{\mathcal{D}}_B = \{(c_{B,i}, r_{B,i}^+, r_{B,i}^-)\}_{i=1}^{N_B}$, then model A evaluates each $(c_{B,i}, r_{B,i}^+, r_{B,i}^-) \in \bar{\mathcal{D}}_B$ with matching scores $s_A(c_{B,i}, r_{B,i}^+)$ and $s_A(c_{B,i}, r_{B,i}^-)$, and form a margin for model B as

$$\Delta_{B,i} = \max\left(0, \lambda(s_A(c_{B,i}, r_{B,i}^+) - s_A(c_{B,i}, r_{B,i}^-))\right), \quad (1)$$

where λ is a hyper-parameter. Similarly, $\forall(c_{A,i}, r_{A,i}^+, r_{A,i}^-) \in \bar{\mathcal{D}}_A$, the margin provided by model B for model A can be formulated as

$$\Delta_{A,i} = \max\left(0, \lambda(s_B(c_{A,i}, r_{A,i}^+) - s_B(c_{A,i}, r_{A,i}^-))\right), \quad (2)$$

where $s_B(c_{A,i}, r_{A,i}^+)$ and $s_B(c_{A,i}, r_{A,i}^-)$ are matching scores calculated with model B. Loss functions

Algorithm 1: The proposed co-teaching framework

Input: model parameters θ_A, θ_B, learning rate η, number of epochs n_T, number of iterations n_K;

for $T = 1, 2, ..., T_{n_T}$ **do**
  Shuffle training set $\mathcal{D}$;
  for $K = 1, 2, ..., K_{n_K}$ **do**
    Fetch a batch of training data $\bar{\mathcal{D}}$;
    Distributes $\bar{\mathcal{D}}$ equally to two sub-batches of training data $\bar{\mathcal{D}}_A, \bar{\mathcal{D}}_B$; ▷ $\bar{\mathcal{D}}_A, \bar{\mathcal{D}}_B \subset \bar{\mathcal{D}}$
    Obtain learning protocol $(\tilde{\mathcal{D}}_B, \mathcal{J}_B)$ from model A and $\bar{\mathcal{D}}_B$;
    Obtain learning protocol $(\tilde{\mathcal{D}}_A, \mathcal{J}_A)$ from model B and $\bar{\mathcal{D}}_A$;
    Update $\theta_A = \theta_A - \eta \nabla \mathcal{J}_A(\tilde{\mathcal{D}}_A)$; ▷ Update model A by $(\tilde{\mathcal{D}}_A, \mathcal{J}_A)$.
    Update $\theta_B = \theta_B - \eta \nabla \mathcal{J}_B(\tilde{\mathcal{D}}_B)$; ▷ Update model B by $(\tilde{\mathcal{D}}_B, \mathcal{J}_B)$.
  end
end

Output: θ_A, θ_B.

$\mathcal{J}_A$ and $\mathcal{J}_B$ are then defined as

$$\mathcal{J}_A = \sum_{i=1}^{N_A} \max\{0, \Delta_{A,i} - s_A(c_{A,i}, r^+_{A,i}) + s_A(c_{A,i}, r^-_{A,i})\}, \quad (3)$$

$$\mathcal{J}_B = \sum_{i=1}^{N_B} \max\{0, \Delta_{B,i} - s_B(c_{B,i}, r^+_{B,i}) + s_B(c_{B,i}, r^-_{B,i})\}. \quad (4)$$

Intuitively, one model may assign a small margin to a negative example if it identifies the example as a false negative. Then, its peer model will pay less attention to such an example in its optimization. This is how the two peer models help each other combat with noise under the strategy of teaching with dynamic margins.

Teaching with Dynamic Instance Weighting: Similar to the first strategy, this strategy also defines the learning protocols with dynamic loss functions. The difference is that this strategy penalizes low-quality negative training examples with weights. Formally, let us represent $\bar{\mathcal{D}}_B$ as $\{(y_{B,i}, c_{B,i}, r_{B,i})\}_{i=1}^{N'_B}$, then $\forall (y_{B,i}, c_{B,i}, r_{B,i}) \in \bar{\mathcal{D}}_B$, its weight from model A is defined as

$$w_{B,i} = \begin{cases} 1 & y_{B,i} = 1 \\ 1 - s_A(c_{B,i}, r_{B,i}) & y_{B,i} = 0 \end{cases} \quad (5)$$

Similarly, $\forall (y_{A,i}, c_{A,i}, r_{A,i}) \in \bar{\mathcal{D}}_A$, model B assign a weight as

$$w_{A,i} = \begin{cases} 1 & y_{A,i} = 1 \\ 1 - s_B(c_{A,i}, r_{A,i}) & y_{A,i} = 0 \end{cases} \quad (6)$$

Then, loss functions $\mathcal{J}_A$ and $\mathcal{J}_B$ can be formulated as

$$\mathcal{J}_A = \sum_{i=1}^{N'_A} w_{A,i} L(y_{A,i}, s_A(c_{A,i}, r_{A,i})), \quad (7)$$

$$\mathcal{J}_B = \sum_{i=1}^{N'_B} w_{B,i} L(y_{B,i}, s_B(c_{B,i}, r_{B,i})), \quad (8)$$

where $L(\cdot, \cdot)$ is defined by cross entropy:

$$-y \log(s(c, r)) + (1 - y) \log(1 - s(c, r)). \quad (9)$$

In this strategy, negative examples that are identified as false negatives by one model will obtain small weights from the model, and thus be less important than other examples in the learning process of the other model.

Teaching with Dynamic Data Curriculum: In the first two strategies, knowledge is transferred mutually through "soft labels" defined by the peer matching models. In this strategy, we directly transfer data to each model. During learning, $\mathcal{J}_A$ and $\mathcal{J}_B$ are fixed as cross entropy, and the learning protocols vary by $\tilde{\mathcal{D}}_A$ and $\tilde{\mathcal{D}}_B$. Inspired by Han et al. (2018), we construct $\tilde{\mathcal{D}}_A$ and $\tilde{\mathcal{D}}_B$ with small-loss instances. These instances are far from decision boundaries of the two models, and thus are more likely to be true positives and true negatives. Formally, $\tilde{\mathcal{D}}_A$ and $\tilde{\mathcal{D}}_B$ are defined as

$$\tilde{\mathcal{D}}_B = argmin_{|\tilde{\mathcal{D}}_B| = \delta |\bar{\mathcal{D}}_B|, \tilde{\mathcal{D}}_B \subset \bar{\mathcal{D}}_B} \mathcal{J}_A(\tilde{\mathcal{D}}_B),$$
$$\tilde{\mathcal{D}}_A = argmin_{|\tilde{\mathcal{D}}_A| = \delta |\bar{\mathcal{D}}_A|, \tilde{\mathcal{D}}_A \subset \bar{\mathcal{D}}_A} \mathcal{J}_B(\tilde{\mathcal{D}}_A), \quad (10)$$

where $|\cdot|$ measures the size of a set, $\mathcal{J}_A(\tilde{\mathcal{D}}_B)$ and $\mathcal{J}_B(\tilde{\mathcal{D}}_A)$ stand for accumulation of loss on the corresponding data sets, and δ is a hyper-parameter. Note that we do not shrink δ as in Han et al. (2018), since fixing δ as a constant yields a simple yet effective learning program, as will be seen in our experiments.

4 Experiments

We test our learning schemes on two public data sets with human annotated test examples.

4.1 Experimental Setup

The first data set we use is Douban Conversation Corpus (Douban) (Wu et al., 2017) which is a multi-turn Chinese conversation data set crawled from Douban group[3]. The data set consists of 1 million context-response pairs for training, 50 thousand pairs for validation, and 6,670 pairs for test. In the training set and the validation set, the last turn of each conversation is regarded as a positive response and negative responses are randomly sampled. The ratio of the positive and the negative is 1:1 in training and validation. In the test set, each context has 10 response candidates retrieved from an index whose appropriateness regarding to the context is judged by human annotators. The average number of positive responses per context is 1.18. Following Wu et al. (2017), we employ $R_{10}@1$, $R_{10}@2$, $R_{10}@5$, mean average precision (MAP), mean reciprocal rank (MRR), and precision at position 1 (P@1) as evaluation metrics.

In addition to the Douban data, we also choose E-commerce Dialogue Corpus (ECD) (Zhang et al., 2018b) as an experimental data set. The data consists of real-world conversations between customers and customer service staff in Taobao[4], which is the largest e-commerce platform in China. There are 1 million context-response pairs in the training set, and 10 thousand pairs in both the validation set and the test set. Each context in the training set and the validation set corresponds to one positive response candidate and one negative response candidate, while in the test set, the number of response candidates per context is 10 with only one of them positive. In the released data, human responses are treated as positive responses, and negative ones are automatically collected by ranking the response corpus based on conversation history augmented messages using Apache Lucene[5]. Thus, we recruit 3 active users of Taobao as human annotators, and ask them to judge each context-response pair in the test data (i.e., in total 10 thousand pairs are judged). If a response can naturally reply to a message given the conversation history before it, then the context-response pair is labeled as 1, otherwise, it is labeled as 0. Each pair receives three labels and the majority is taken as the final decision. On average, each context has 2.5 response candidates labeled as positive. There are only 33 contexts with all responses labeled as positive or negative, and we remove them from test. Fleiss' kappa (Fleiss, 1971) of the labeling is 0.64, indicating substantial agreement among the annotators. We employ the same metrics as in Douban for evaluation.

Note that we do not choose the Ubuntu Dialogue Corpus (Lowe et al., 2015) for experiments, because (1) the test set of the Ubuntu data is constructed by randomly sampling; and (2) conversations in the Ubuntu data are in a casual style and too technical, and thus it is very difficult for us to find qualified human annotators to label the data.

4.2 Matching Models

We select the following two models that achieve superior performance on benchmarks to test our learning approach.

SMN: (Wu et al., 2017) first lets each utterance in a context interact with a response, and forms a matching vector for the pair through CNNs. Matching vectors of all the pairs are then aggregated with an RNN as a matching score.

DAM: (Zhou et al., 2018b) performs matching under a representation-matching-aggregation framework, and represents a context and a response with stacked self-attention and cross-attention.

Both models are implemented with TensorFlow according to the details in Wu et al. (2017) and Zhou et al. (2018b). To implement co-teaching, we pre-train the two models using the training sets of Douban and ECD, and tune the models with the validation sets of the two data. Each pre-trained model is used to initialize both model A and model B. After co-teaching, the one in A and B that performs better on the validation sets is picked for comparison. We denote models learned with the teaching strategies in Section 3.2

[3] https://www.douban.com/group
[4] https://www.taobao.com
[5] http://lucene.apache.org/

	Douban						ECD					
	MAP	MRR	P@1	R_{10}@1	R_{10}@2	R_{10}@5	MAP	MRR	P@1	R_{10}@1	R_{10}@2	R_{10}@5
SMN (Wu et al., 2017)	0.529	0.569	0.397	0.233	0.396	0.724	-	-	-	-	-	-
SMN-Pre-training	0.527	0.570	0.396	0.236	0.392	0.734	0.662	0.742	0.598	0.302	0.464	0.757
SMN-Margin	**0.559***	**0.601***	**0.424***	**0.260***	**0.426***	**0.764***	0.674	0.750	0.615	0.318	0.481	0.765
SMN-Weighting	0.550*	0.593*	0.414	0.253	0.413	0.762*	0.666	0.745	0.601	0.311	0.475	0.775
SMN-Curriculum	0.548	0.594*	0.418*	0.254*	0.411	0.763*	**0.678**	**0.762***	**0.622***	**0.323***	**0.487***	**0.778***
DAM (Zhou et al., 2018b)	0.550	0.601	0.427	0.254	0.410	0.757	-	-	-	-	-	-
DAM-Pre-training	0.552	0.605	0.426	0.258	0.408	0.766	0.685	0.756	0.621	0.325	0.491	0.772
DAM-Margin	**0.583***	0.628*	0.451*	**0.276***	**0.454***	0.806*	0.692	**0.777***	0.652*	0.337	0.506	0.778
DAM-Weighting	0.579*	**0.629***	**0.453***	0.272	**0.454***	**0.809***	0.695	0.775	0.651*	0.343	0.497	**0.789**
DAM-Curriculum	0.580*	0.623*	0.442	0.269	0.459*	0.804*	**0.696**	**0.777***	**0.653***	**0.345***	**0.506**	0.781

Table 1: Evaluation results on the two data sets. Numbers marked with $*$ mean that the improvement is statistically significant compared with the best baseline (t-test with p-value < 0.05). Numbers in bold indicate the best strategies for the corresponding models on specific metrics.

as Model-Margin, Model-Weighting, and Model-Curriculum respectively, where "Model" refers to either SMN or DAM. These models are compared with the pre-trained model denoted as Model-Pre-training, and those reported in Wu et al. (2017); Zhou et al. (2018b); Zhang et al. (2018b).

4.3 Implementation Details

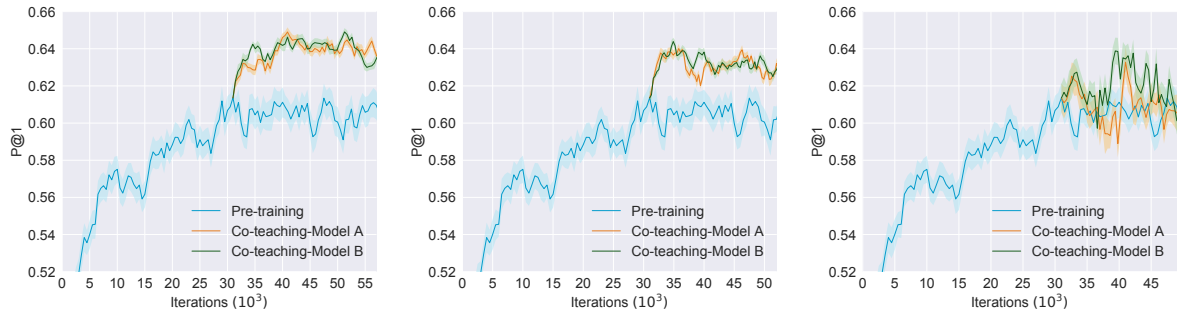
We limit the maximum number of utterances in each context as 10 and the maximum number of words in each utterance and response as 50 for computational efficiency. Truncation or zero-padding are applied when necessary. Word embedding is pre-trained with Word2Vec (Mikolov et al., 2013) on the training sets of Douban and ECD, and the dimension of word vectors is 200. The co-teaching framework is implemented with TensorFlow. In co-teaching, learning rates (i.e., η in Algorithm 1) in dynamic margins, dynamic instance weighting, and dynamic data curriculum are set as 0.001, 0.0001, and 0.0001 respectively. We choose 200 in co-teaching with SMN and 50 in co-teaching with DAM as the size of mini-batches. Optimization is conducted using stochastic gradient descent with Adam algorithm (Kingma and Ba, 2015). In teaching with dynamic margins, we vary λ in $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{5}, \frac{1}{10}, \frac{1}{15}, \frac{1}{20}\}$, and choose $\frac{1}{10}$ for SMN on Douban, $\frac{1}{2}$ for SMN on ECD, $\frac{1}{3}$ for DAM on Douban, and $\frac{1}{2}$ for DAM on ECD. In teaching with dynamic data curriculum, we select δ in $\{0.1, 0.2, ..., 0.9, 1.0\}$, and find that 0.9 is the best choice for both models on both data sets.

4.4 Evaluation Results

Table 1 reports evaluation results of co-teaching with the three teaching strategies on the two data sets. We can see that all teaching strategies can improve the original models on both data sets, and improvement from the best strategy is statistically significant (t-test with p-value < 0.05) on most metrics. On Douban, the best strategy for SMN is teaching with dynamic margins, and it is comparable with teaching with dynamic instance weighting for DAM, while on ECD, for both SMN and DAM, the best strategy is teaching with dynamic data curriculum. The difference may stem from the nature of training sets of the two data. The training set of Douban is built from random sampling, while the training set of ECD is constructed through response retrieval that may contain more false negatives. Thus, in training, Douban could be cleaner than ECD, making "hard data filtering" more effective than "soft labeling" on ECD. It is worth noting that on ECD, there are significant gaps between the results of SMN (pre-trained) reported in Table 1 and those reported in Zhang et al. (2018b), since SMN in this paper is evaluated on the human-judged test set while SMN in Zhang et al. (2018b) is evaluated on the automatically constructed test set that is homogeneous with the training set. This somehow indicates the gap between training and test in real applications for the existing research on response selection, and thus demonstrates the merits of this work.

4.5 Discussions

In addition to efficacy of co-teaching as a learning approach, we are also curious about **Q1:** if model A and model B can "co-evolve" when they are initialized with one network; **Q2:** if co-teaching is still effective when model A and model B are initialized with different networks; and **Q3:** if the teaching strategies are sensitive to the hyper-parameters (i.e., λ in Equations (1)-(2) and δ in Equation (10)).



(a) Dynamic data curriculum (b) Dynamic instance weighting (c) Dynamic margins

Figure 2: Test P@1 of DAM with the three teaching strategies on ECD. All curves are smoothed by exponential moving average[6] for beauty.

	Douban (Margin)						ECD (Curriculum)					
	MAP	MRR	P@1	R_{10}@1	R_{10}@2	R_{10}@5	MAP	MRR	P@1	R_{10}@1	R_{10}@2	R_{10}@5
SMN-Pre-training	0.527	0.570	0.396	0.236	0.392	0.734	0.662	0.742	0.598	0.302	0.464	0.757
SMN-Co-teaching	0.558	0.602	0.420	0.255	0.431	0.787	0.674	0.765	0.626	0.322	0.485	0.779
DAM-Pre-training	0.552	0.605	0.426	0.258	0.408	0.766	0.685	0.756	0.621	0.325	0.491	0.772
DAM-Co-teaching	0.570	0.617	0.438	0.270	0.455	0.781	0.696	0.775	0.652	0.341	0.499	0.784

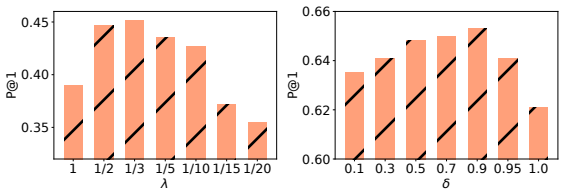
Table 2: Evaluation results of co-teaching initialized with different networks.

Answer to Q1: Figure 2 shows P@1 of DAM vs. number of iterations on the test set of ECD under the three teaching strategies. Co-teaching with any of the three strategies can improve both the performance of model A and the performance of model B after pre-training, and the peer models move with almost the same pace. The results verified our claim that "by learning from each other, the peer models can get improved together". Curves of dynamic margins oscillate more fiercely than others, indicating that optimization with dynamic margins is more difficult than optimization with the other two strategies.

Answer to Q2: as a case study of co-teaching with two networks in different capabilities, we initialize model A and model B with DAM and SMN respectively, and select teaching with dynamic margins for Douban and teaching with dynamic data curriculum for ECD (i.e., the best strategies for the two data sets when co-teaching is initialized with one network). Table 2 shows comparison between models before/after co-teaching. We find that co-teaching is still effective when starting from two networks, as both SMN and DAM get improved on the two data sets. Despite the improvement, it is still better to learn the two networks one by one, as co-teaching with two networks cannot bring more improvement than co-teaching with one network, and the performance of the stronger one between the two networks could also drop (e.g., DAM on Douban). We guess this is because the stronger model cannot be well taught by the weaker model, especially in teaching via soft labels, and as a result, it is not able to transfer more knowledge to the weaker one as well.

Answer to Q3: finally, we check the effect of hyper-parameters to co-teaching. Figure 3(a) illustrates how the performance of DAM varies under different λs in teaching with dynamic margins on Douban. We can see that both small λs and large λs will cause performance drop. This is because small λs will reduce the effect of margins, making clean examples and noisy examples indifferent in learning, while with large λs, some errors from the "soft labels" might be magnified, and thus hurt the performance of the learning approach. Figure 3(b) shows the performance of DAM under different δs in teaching with dynamic data curriculum on ECD. Similarly, DAM gets worse when δ becomes small or large, since a smaller δ means fewer data will be involved in training, while a larger δ brings more risks to introducing noise into training. Thus, we conclude that the teaching strategies are sensitive to the choice of hyper-parameters.

[6] https://en.wikipedia.org/wiki/Moving_average#Exponential_moving_average



(a) Dynamic margins on Douban (b) Data curriculum on ECD

Figure 3: Effects of λ and δ to co-teaching. Experiments are conducted with DAM on the two data sets.

5 Related Work

So far, methods used to build an open domain dialogue system can be divided into two categories. The first category utilize an encoder-decoder framework to learn response generation models. Since the basic sequence-to-sequence models (Vinyals and Le, 2015; Shang et al., 2015; Tao et al., 2018) tend to generate generic responses, extensions have been made to incorporate external knowledge into generation (Mou et al., 2016; Xing et al., 2017), and to generate responses with specific personas or emotions (Li et al., 2016; Zhang et al., 2018a; Zhou et al., 2018a). The second category design a discriminative model to measure the matching degree between a human input and a response candidate for response selection. At the beginning, research along this line assumes that the human input is a single message (Lu and Li, 2013; Wang et al., 2013; Hu et al., 2014; Wang et al., 2015). Recently, researchers begin to make use of conversation history in matching. Representative methods include the dual LSTM model (Lowe et al., 2015), the deep learning to respond architecture (Yan et al., 2016), the multi-view matching model (Zhou et al., 2016), the sequential matching network (Wu et al., 2017, 2018c), the deep attention matching network (Zhou et al., 2018b), and the multi-representation fusion network (Tao et al., 2019).

Our work belongs to the second group. Rather than crafting a new model, we are interested in how to learn the existing models with a better approach. Probably the most related work is the weakly supervised learning approach proposed in Wu et al. (2018b). However, there is stark difference between our approach and the weak supervision approach: (1) weak supervision employs a static generative model to teach a discriminative model, while co-teaching dynamically lets two discriminative models teach each other and evolve together; (2) weak supervision needs pre-training a generative model with extra resources and pre-building an index for training data construction, while co-teaching does not have such request; and (3) in terms of multi-turn response selection, weak supervision is only tested on the Douban data with SMN and the multi-view matching model, while co-teaching is proven effective on both the Douban data and the E-commerce data with SMN and DAM which achieves state-of-the-art performance on benchmarks. Moreover, improvement to SMN on the Douban data from co-teaching is bigger than that from weak supervision, when the ratio of the positive and the negative is 1:1 in training[7].

Our work, in a broad sense, belongs to the effort on learning with noisy data. Previous studies including curriculum learning (CL) (Bengio et al., 2009) and self-paced learning (SPL) (Jiang et al., 2014, 2015) tackle the problem with heuristics, such as ordering data from easy instances to hard ones (Spitkovsky et al., 2010; Tsvetkov et al., 2016) and retaining training instances whose losses are smaller than a threshold (Jiang et al., 2015). Recently, Fan et al. (2018) propose a deep reinforcement learning framework in which a simple deep neural network is used to adaptively select and filter important data instances from the training data. Jiang et al. (2017) propose a MentorNet which learns a data-driven curriculum with a Student-Net to mitigate overfitting on corrupted labels. In parallel to curriculum learning, several studies explore sample weighting schemes where training samples are re-weighted according to their label-quality (Wang et al., 2017; Dehghani et al., 2018; Wu et al., 2018b). Instead of considering data quality, Wu et al. (2018a) employ a parametric model to dynamically create appropriate loss functions.

The learning approach in this work is mainly inspired by the work of Han et al. (2018) for handling extremely noisy labels. However, with substantial extensions, our work is far beyond that work. First, we generalize the concept of "co-teaching" to a framework, and now the method in Han et al. (2018) becomes a special case of the framework. Second, Han et al. (2018) only exploits data curriculum, while in addition to data

[7]Our results are 0.559 (MAP), 0.601 (MRR), and 0.424 (P@1), while results reported in (Wu et al., 2018b) are 0.542 (MAP), 0.588 (MRR), and 0.408 (P@1).

curriculum, we also propose two new strategies for teaching with dynamic loss functions as special cases of the framework. Third, unlike Han et al. (2018) who only use one network to initialize the peer models in co-teaching, we studied co-teaching with both one network and two different networks. Finally, Han et al. (2018) verified that the special co-teaching method is effective in some computer vision tasks, while we demonstrate that the co-teaching framework is generally useful for building retrieval-based dialogue systems.

6 Conclusions

We propose learning a matching model for response selection under a general co-teaching framework with three specific teaching strategies. The learning approach lets two matching models teach each other and evolve together. Empirical studies on two public data sets show that the proposed approach can generally improve the performance of existing matching models.

Acknowledgement

We would like to thank the anonymous reviewers for their constructive comments. This work was supported by the National Key Research and Development Program of China (No. 2017YFC0804001), the National Science Foundation of China (NSFC Nos. 61672058 and 61876196).

References

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM.

Mostafa Dehghani, Arash Mehrjou, Stephan Gouws, Jaap Kamps, and Bernhard Schölkopf. 2018. Fidelity-weighted learning. In *International Conference on Learning Representations*.

Yang Fan, Fei Tian, Tao Qin, Xiang-Yang Li, and Tie-Yan Liu. 2018. Learning to teach. In *International Conference on Learning Representations*.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor W. Tsang, and Masashi Sugiyama. 2018. Co-sampling: Training robust networks for extremely noisy supervision. *CoRR*, abs/1804.06872.

Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional neural network architectures for matching natural language sentences. In *Advances in Neural Information Processing Systems*, pages 2042–2050.

Zongcheng Ji, Zhengdong Lu, and Hang Li. 2014. An information retrieval approach to short text conversation. *arXiv preprint arXiv:1408.6988*.

Lu Jiang, Deyu Meng, Shoou-I Yu, Zhenzhong Lan, Shiguang Shan, and Alexander Hauptmann. 2014. Self-paced learning with diversity. In *Advances in Neural Information Processing Systems*, pages 2078–2086.

Lu Jiang, Deyu Meng, Qian Zhao, Shiguang Shan, and Alexander G Hauptmann. 2015. Self-paced curriculum learning. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. 2017. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *Proceedings of the 35-th International Conference on Machine Learning,*, pages 2304–2313.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.

Feng-Lin Li, Minghui Qiu, Haiqing Chen, Xiongwei Wang, Xing Gao, Jun Huang, Juwei Ren, Zhongzhou Zhao, Weipeng Zhao, Lei Wang, et al. 2017. Alime assist: An intelligent assistant for creating an innovative e-commerce experience. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2495–2498.

Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. In *Association for Computational Linguistics*, pages 994–1003.

Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2015. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 285–294.

Zhengdong Lu and Hang Li. 2013. A deep architecture for matching short texts. In *Advances in Neural Information Processing Systems*, pages 1367–1375.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Lili Mou, Yiping Song, Rui Yan, Ge Li, Lu Zhang, and Zhi Jin. 2016. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3349–3358.

Iulian Vlad Serban, Tim Klinger, Gerald Tesauro, Kartik Talamadupula, Bowen Zhou, Yoshua Bengio, and Aaron Courville. 2017. Multiresolution recurrent neural networks: An application to dialogue response generation. In *AAAI*, pages 3288–3294.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1577–1586.

Heung-Yeung Shum, Xiaodong He, and Di Li. 2018. From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of IT & EE*, 19(1):10–26.

Valentin I. Spitkovsky, Hiyan Alshawi, and Daniel Jurafsky. 2010. From baby steps to leapfrog: How "less is more" in unsupervised dependency parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 751–759.

Chongyang Tao, Shen Gao, Mingyue Shang, Wei Wu, Dongyan Zhao, and Rui Yan. 2018. Get the point of my utterance! learning towards effective responses with multi-head attention mechanism. In *IJCAI*, pages 4418–4424.

Chongyang Tao, Wei Wu, Can Xu, Wenpeng Hu, Dongyan Zhao, and Rui Yan. 2019. Multi-representation fusion network for multi-turn response selection in retrieval-based chatbots. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 267–275. ACM.

Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Brian MacWhinney, and Chris Dyer. 2016. Learning the curriculum with bayesian optimization for task-specific word representation learning. *arXiv preprint arXiv:1605.03852*.

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Hao Wang, Zhengdong Lu, Hang Li, and Enhong Chen. 2013. A dataset for research on short-text conversations. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 935–945.

Mingxuan Wang, Zhengdong Lu, Hang Li, and Qun Liu. 2015. Syntax-based deep matching of short texts. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 1354–1361.

Yixin Wang, Alp Kucukelbir, and David M Blei. 2017. Robust probabilistic modeling with bayesian data reweighting. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3646–3655. JMLR. org.

Lijun Wu, Fei Tian, Yingce Xia, Yang Fan, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2018a. Learning to teach with dynamic loss functions. *CoRR*, abs/1810.12081.

Yu Wu, Wei Wu, Zhoujun Li, and Ming Zhou. 2018b. Learning matching models with weak supervision for response selection in retrieval-based chatbots. *arXiv preprint arXiv:1805.02333*.

Yu Wu, Wei Wu, Chen Xing, Can Xu, Zhoujun Li, and Ming Zhou. 2018c. A sequential matching framework for multi-turn response selection in retrieval-based chatbots. *Computational Linguistics*, 45(1):163–197.

Yu Wu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li. 2017. Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 496–505.

Chen Xing, Wei Wu, Yu Wu, Jie Liu, Yalou Huang, Ming Zhou, and Wei-Ying Ma. 2017. Topic aware neural response generation. In *AAAI*, pages 3351–3357.

Rui Yan, Yiping Song, and Hua Wu. 2016. Learning to respond with deep neural networks for retrieval-based human-computer conversation system. In *SIGIR*, pages 55–64.

Rui Yan and Dongyan Zhao. 2018. Coupled context modeling for deep chit-chat: towards conversations between human and computer. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2574–2583. ACM.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018a. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*.

Zhuosheng Zhang, Jiangtong Li, Pengfei Zhu, Hai Zhao, and Gongshen Liu. 2018b. Modeling multi-turn conversation with deep utterance aggregation. *CoRR*, abs/1806.09102.

Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018a. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *The Thirty-Second AAAI Conference on Artificial Intelligence*, pages 730–738.

Xiangyang Zhou, Daxiang Dong, Hua Wu, Shiqi Zhao, Dianhai Yu, Hao Tian, Xuan Liu, and Rui Yan. 2016. Multi-view response selection for human-computer conversation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 372–381.

Xiangyang Zhou, Lu Li, Daxiang Dong, Yi Liu, Ying Chen, Wayne Xin Zhao, Dianhai Yu, and Hua Wu. 2018b. Multi-turn response selection for chatbots with deep attention matching network. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1118–1127.