Dual Supervised Learning

Abstract

Two dual tasks have intrinsic connections with each other due to the probabilistic correlation between their models.

In this work, we propose training the models of two dual tasks simultaneously, and explicitly exploiting the probabilistic correlation between them to regularize the training process.

1. Introduction

Interestingly, we find that many of the aforementioned AI tasks are emerged in dual forms, i.e., the input and output of one task are exactly the output and input of the other task respectively.

To exploit the duality, we formulate a new learning scheme, which involves two tasks: a primal task and its dual task. The primal task takes a sample from space X as input and maps to space Y, and the dual task takes a sample from space Y as input and maps to space X.

P(x; y) = P(x)P(y|x) = P(y)P(x|y)

P(x)P(y|x;Theta(xy)) = P(y)P(x|y;Theta(yx)) (1)

The basic idea of DSL is to jointly learn the two models and by Theta(xy) and Theta(yx) minimizing their loss functions subject to the constraint of equation (1).

To demonstrate the effectiveness of DSL, we apply it to three artificial intelligence applications:

(1) Neural Machine Translation (NMT) We first apply DSL to NMT, which formulates machine translation as a sequence-to-sequence learning problem, with the sentences in the source language as inputs and those in the target language as outputs.

(2) Image Processing We then apply DSL to image processing, in which the primal task is image classification and the dual task is image generation conditioned on category labels.

(3) Sentiment Analysis Finally, we apply DSL to sentiment analysis, in which the primal task is sentiment classification (i.e., to predict the sentiment of a given sentence) and the dual one is sentence generation with given sentiment polarity.

2. Framework

2.1. Problem Formulation

To exploit the duality, we formulate a new learning scheme, which involves two tasks: a primal task that takes a sample from space X as input and maps to space Y, and a dual task takes a sample from space Y as input and maps to space X.

two tasks:

(1) the primal learning task aims at finding a function f : X -> Y such that the prediction of f for x is similar to its real counterpart y;

(2) the dual learning task aims at finding a function g : Y -> X such that the prediction of g for y is similar to its real counterpart x.

2.2. Algorithm Description

2.3. Discussions

The duality between tasks has been used to enable learning from unlabeled data. For example, in the application of machine translation, the primal task/model first translates an unlabeled English sentence to a French sentence y0; then, the dual task/model x translates y0 back to an English sentence x0; finally, both the primal and the dual models get optimized by minimizing the difference between x0 with x.

DSL is different from the following three learning schemes: (1) Co-training focuses on single-task learning and assumes that different subsets of features can provide enough and complementary information about data, while DSL targets at learning two tasks with structural duality simultaneously

and does not yield any prerequisite or assumptions on features. (2) Multi-task learning requires that different tasks share the same input space and coherent feature representation while DSL does not. (3) Transfer Learning uses auxiliary tasks to boost the main task, while there is no difference between the roles of two tasks in DSL, and DSL enables them to boost the performance of each other simultaneously.

1. Application to Machine Translation
2. Application to Images Processing
3. Application to Sentiment Analysis
4. Conclusions and Future Work

There are multiple directions to explore in the future. First, we will test dual supervised learning on more dual tasks, such as speech recognition and speech synthesis. Second, we will enrich theoretical study to better understand dual supervised learning. Third, it is interesting to combine dual supervised learning with unsupervised dual learning (He et al., 2016a) to leverage unlabeled data so as to further improve the two dual tasks. Fourth, we will combine dual supervised learning with dual inference (Xia et al., 2017) so as to leverage structural duality to enhance both the training and inference procedures.