Explainable Recommendation

Lingting Lin

August 2018

1 Experiment Process

1.1 Data Set

We use the list of product recommendation reasons as our dataset. The corpus are generated by high quality person.But the quality of the original dataset is far from ideal, there are many recommended reasons are even the original title of the product, so we need to filter the training data.

1.2 Weak Supervision

Manual labeling is very time consuming, so we use the Snorkel weak supervision method to mark the data without the user to manually mark any training data. Rather than hand-labeling training data, users of Snorkel write labeling functions, which allow them to express various weak supervision sources such as patterns, heuristics, external knowledge bases, and more. We wrote ten labeling functions based on the characteristics of persuasive sentences, as follows

Table 1: Labeling Functions

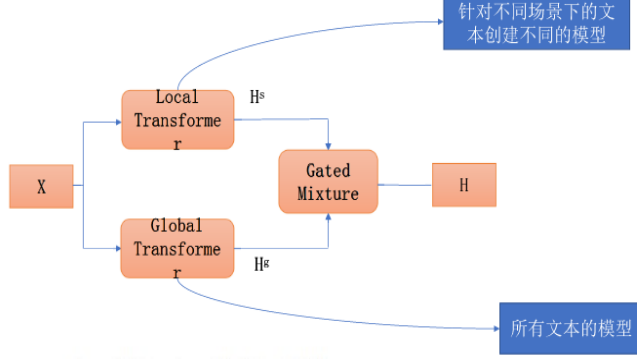| Labeling Functions	Description
is_neat	Sentence is neat
has_adj	Sentence has adjectives
has_adv	Sentence has adverbs
has_modal	Sentence has modal particle
end_exclamation	Sentence ends with an exclamation point
other_words	Sentence contains characters other than Chinese, English, numbers, and specified symbols (。，？！、；：).
four_word	Sentence contains a four-word structure
dot_word	The comma is followed by 〞让/使/为/给〞
dot_v	Comma followed by verbs
tree_depths	the depth of the dependency tree

Figure 1: Global-Locally Transformer

Next, Snorkel automatically learns a generative model over the labeling functions,the output of Snorkel is a set of probabilistic labels. We choose sentences with probabilistic labels are bigger than 0.99 and the words are less than 50 as training set.

1.3 Model

We extract the scene name, product name, cpv in the sentence as the input of model,and the model's output is a textual description of the item with the scene. Cpv is a collection of values of the attributes of the item. Here, only the value of the product attributes is in the sentence it can be extracted. Training data format is as follows:

Table 2: Training data format

Input	Output
家居,沙发,米黄色	米黄色的布艺沙发，质地舒适，让家居环境更为温馨

Global-transformer is a global model on all texts, learning text descriptions of all products, local-transformer creating different models for texts in different scenarios, learning knowledge within the scene.

$$\mathbf{H} = \beta^s \mathbf{H}^s + (1 - \beta^s)\mathbf{H}^g. \tag{1}$$

Here, the scalar s is a learned parameter between 0 and 1 that is specific to the scenario s.