1 Problem Definition

We are given a set of strings that represents the Chinese document written by the gradute students. We formally define the set of documents as D. For each document $d \in D$, we use a k-dimensional word vetor $x_d =< x_{d1}, ..., x_{dK} >$ ($\forall_{k,x_{dk}} \in \mathbb{R}$), where x_{dk} indicates the k-th feature of document. More specifically, the feature can be abstract like Word Enbedding, or interpretable like Bag of Words model.

For each $d \in D$, we are given a number s_d and a set of string e_d, which represents the score and the evaluation of document d.

Definition 1 ***Document evaluation function*** Given the set of documents D, and score S, our goal is to learn a function f_d, which can caculate the score of a given document:

$$f_d(D) \to S \tag{1.1}$$

Definition 2 ***Document evaluation classification function*** Given the set of documents D, and evaluation E, our goal is to learn a function f_e, which can caculate the score of a given document on several categories.

We divide the evaluation into $< T, N, A, I >$, which indicates the topic, norm, achievement and innovation. We train a divide function f_{divE}, which can express the evaluation with a 4-dimensional feature vector $e_i =< e_{iT}, e_{iN}, e_{iA}, e_{iI} >$.

We use the set of documents D, and the evaluation vector E to learn a function f_e, which can caculate the value on the four feature.

Definition 3 ***Sentence evaluation function*** In this part, we want to make the score more fine-grained. Since we don't have the score of every sentence in document. We can extract the first sentence of each paragraph and the total abstract, which has a good represention of the document. We give these sentences in d_i the score $f_{score}(sen_{ij})$, where sen_{ij} indicates the j-th sentence in d_i and f_{score} indicates a sentence score function base on the given score s_i of d_i (e.g., the abstract have a higher weight of the represention of a document, so its value is closer to s_i).

We ues the set of sentence Sen, and the score of sentences S_{sen} to learn a function f_{sen}, which can caculate the score of a sentence.