## CLAIMS

What is claimed is:

1. A model training method, comprising:

obtaining an original text, a positive sample, and a negative sample, wherein the positive sample and the negative sample are rewritten texts of the original text;

determining, by using an encoder, a first distribution parameter corresponding to the original text, a second distribution parameter corresponding to the positive sample, and a third distribution parameter corresponding to the negative sample, wherein the first distribution parameter is used to describe a first semantic distribution, the second distribution parameter is used to describe a second semantic distribution, and the third distribution parameter is used to describe a third semantic distribution;

determining an original keyword, a positive sample keyword, and a negative sample keyword respectively based on the original text, the positive sample, and the negative sample; and

optimizing a model parameter of the encoder by using a loss function based on the first distribution parameter, the second distribution parameter, the third distribution parameter, the original keyword, the positive sample keyword, and the negative sample keyword, wherein the loss function comprises a first term and a second term, wherein the first term is used to constrain a degree of similarity between the first semantic distribution and the second semantic distribution to be higher than a degree of similarity between the first semantic distribution and the third semantic distribution, and wherein the second term is used to constrain a degree of similarity between the original keyword and the positive sample keyword to be higher than a degree of similarity between the original keyword and the negative sample keyword.

2. The method according to claim 1, wherein the encoder comprises a first encoder and a second encoder, the first encoder is an encoder in a text generative model, and the second encoder is used to assist in training of the text generative model; and

the step of obtaining the first distribution parameter, the second distribution parameter, and the third distribution parameter comprises the following:

obtaining, by using the first encoder, the first distribution parameter corresponding to the original text; and

obtaining, by using the second encoder, the second distribution parameter

corresponding to the positive sample and the third distribution parameter corresponding to the negative sample.

3. The method according to claim 2, wherein the step of obtaining the first distribution parameter comprises the following:

concatenating a keyword in the original text and the original text, and performing input into the first encoder to obtain the first distribution parameter; and

the step of obtaining the second distribution parameter and the third distribution parameter comprises the following:

concatenating the original text, a label corresponding to the positive sample, and the positive sample, and performing input into the second encoder to obtain the second distribution parameter; and

concatenating the original text, a label corresponding to the negative sample, and the negative sample, and performing input into the second encoder to obtain the third distribution parameter.

4. The method according to claim 1, wherein the step of determining the negative sample keyword comprises the following:

extracting a keyword from the negative sample as a reference keyword;

constructing a keyword network based on the reference keyword and the positive sample keyword;

selecting one or more reference keywords adjacent to the positive sample keyword from the keyword network; and

constructing the negative sample keyword based on the selected one or more reference keywords.

5. The method according to claim 1, wherein the loss function further comprises a third term, and the third term is used to constrain a distance between the first semantic distribution and the positive sample keyword to be smaller than a distance between the first semantic distribution and the negative sample keyword.

6. The method according to claim 1, wherein the first distribution parameter, the second distribution parameter, and the third distribution parameter each comprises a mean value and a variance, and the first semantic distribution, the second semantic distribution, and the third semantic distribution are normal distributions.

7. The method according to claim 1, wherein the method further comprises the following:

performing sampling on the first semantic distribution to obtain a latent variable;

inputting the latent variable into a decoder; and

optimizing a model parameter of the decoder by using the loss function based on an output of the decoder and the positive sample, wherein the loss function further comprises a fourth term, and the fourth term is used to constrain a degree of difference between the output of the decoder and the positive sample.

8. A text processing method, comprising:

inputting an original text into an encoder to obtain a distribution parameter, wherein the encoder is obtained through training by using the method according to any one of claims 1 to 6, and the distribution parameter is used to describe a semantic distribution;

performing sampling on the semantic distribution to obtain a latent variable; and

inputting the latent variable into a decoder to obtain a rewritten text.

9. The method according to claim 8, wherein the decoder is obtained through training by using the method according to claim 7.

10. A model training apparatus, comprising:

an acquisition unit, configured to obtain an original text, a positive sample, and a negative sample, wherein the positive sample and the negative sample are rewritten texts of the original text;

a determining unit, configured to determine, by using an encoder, a first distribution parameter corresponding to the original text, a second distribution parameter corresponding to the positive sample, and a third distribution parameter corresponding to the negative sample, wherein the first distribution parameter is used to describe a first semantic distribution, the second distribution parameter is used to describe a second semantic distribution, and the third distribution parameter is used to describe a third semantic distribution; and

an optimization unit, configured to optimize a model parameter of the encoder by using a loss function based on the first distribution parameter, the second distribution parameter, the third distribution parameter, the original keyword, the positive sample keyword, and the negative sample keyword, wherein the loss function comprises a first term and a second term, wherein the first term is used to constrain a degree of similarity between the first semantic distribution and the second semantic distribution to be higher than a degree of similarity between the first semantic distribution and the third semantic

distribution, and wherein the second term is used to constrain a degree of similarity between the original keyword and the positive sample keyword to be higher than a degree of similarity between the original keyword and the negative sample keyword.

11. A text processing apparatus, comprising:

a first input unit, configured to input an original text into an encoder to obtain a distribution parameter, wherein the encoder is obtained through training by using the method according to any one of claims 1 to 6, and the distribution parameter is used to describe a semantic distribution;

a sampling unit, configured to perform sampling on the semantic distribution to obtain a latent variable; and

a second input unit, configured to input the latent variable into a decoder to obtain a rewritten text.

12. A computer device, comprising:

at least one processor; and

a memory storing program instructions, wherein the program instructions are configured to be suitable for execution by the at least one processor, and the program instructions comprise instructions for performing the method according to any one of claims 1 to 9.