CLAIMS

What is claimed is:

1. A~~n~~ method for evaluating an unsupervised model for word processing ~~evaluation method~~, wherein the method comprises:

determining S node vectors corresponding to S nodes from N node vectors obtained by using an unsupervised model, wherein both N and S are positive integers, wherein the N node vectors are obtained through providing input data to the unsupervised model, and wherein the input data comprises a quantity of vectors used to represent words;

determining a neighboring node and a non-neighboring node of each of the S nodes;

determining, based on a node vector of each node and a node vector of the neighboring node of each node, a similarity between each node and the neighboring node of each node as a positive sample predicted value;

determining, based on the node vector of each node and a node vector of the non-neighboring node of each node, a similarity between each node and the non-neighboring node of each node as a negative sample predicted value; and

evaluating the unsupervised model based on the positive sample predicted value and the negative sample predicted value.

2. The ~~unsupervised model evaluation~~ method according to claim 1, wherein the evaluating the unsupervised model based on the positive sample predicted value and the negative sample predicted value comprises:

constructing a first-type 2-tuple sample based on the positive sample predicted value and a positive sample label value;

constructing a second-type 2-tuple sample based on the negative sample predicted value and a negative sample label value; and

evaluating the unsupervised model based on the first-type 2-tuple sample and the second-type 2-tuple sample.

3. The ~~unsupervised model evaluation~~ method according to claim 2, wherein the evaluating the unsupervised model based on the first-type 2-tuple sample and the second-type 2-tuple sample comprises:

determining a target receiver operating characteristic curve based on the first-type

2-tuple sample and the second-type 2-tuple sample; and

obtaining an area under the target receiver operating characteristic curve to evaluate the unsupervised model.

4. The ~~unsupervised model evaluation~~ method according to claim 1, wherein the determining a neighboring node and a non-neighboring node of each of the S nodes comprises:

performing the following steps for each of the S nodes:

determining a neighboring node set and a non-neighboring node set of the node from N nodes based on a predetermined neighbor generation algorithm;

determining one or more neighboring nodes from the neighboring node set; and

determining one or more non-neighboring nodes from the non-neighboring node set.

5. The ~~unsupervised model evaluation~~ method according to claim 1, wherein the determining S node vectors corresponding to S nodes from N node vectors obtained by using an unsupervised model comprises:

determining a node set, wherein the node set comprises N nodes, and the N nodes correspond to the N node vectors; and

randomly selecting one node from the node set as a sampling node, and performing selections for a total of S times to obtain the S nodes.

6. An apparatus for evaluating an unsupervised model for word processing ~~evaluation apparatus~~, wherein the apparatus comprises:

a first determining module, configured to determine S node vectors corresponding to S nodes from N node vectors obtained by using an unsupervised model, wherein both N and S are positive integers, wherein the N node vectors are obtained through providing input data to the unsupervised model, and wherein the input data comprises a quantity of vectors used to represent words;

a second determining module, configured to determine a neighboring node and a non-neighboring node of each of the S nodes;

a positive sample predicted value determining module, configured to determine, based on a node vector of each node and a node vector of the neighboring node of each

node, a similarity between each node and the neighboring node of each node as a positive sample predicted value;

a negative sample predicted value determining module, configured to determine, based on the node vector of each node and a node vector of the non-neighboring node of each node, a similarity between each node and the non-neighboring node of each node as a negative sample predicted value; and

an evaluation module, configured to evaluate the unsupervised model based on the positive sample predicted value and the negative sample predicted value.

7. The ~~unsupervised model evaluation~~ apparatus according to claim 6, wherein the evaluation module is configured to:

construct a first-type 2-tuple sample based on the positive sample predicted value and a positive sample label value;

construct a second-type 2-tuple sample based on the negative sample predicted value and a negative sample label value; and

evaluate the unsupervised model based on the first-type 2-tuple sample and the second-type 2-tuple sample.

8. The ~~unsupervised model evaluation~~ apparatus according to claim 7, wherein the evaluation module is configured to:

determine a target receiver operating characteristic curve based on the first-type 2-tuple sample and the second-type 2-tuple sample; and

obtain an area under the target receiver operating characteristic curve to evaluate the unsupervised model.

9. The ~~unsupervised model evaluation~~ apparatus according to claim 6, wherein the second determining module is configured to:

perform the following steps for each of the S nodes:

determine a neighboring node set and a non-neighboring node set of the node from N nodes based on a predetermined neighbor generation algorithm;

determine one or more neighboring nodes from the neighboring node set; and

determine one or more non-neighboring nodes from the non-neighboring node set.

10. The ~~unsupervised model evaluation~~ apparatus according to claim 6, wherein

the first determining module is configured to:

determine a node set, wherein the node set comprises N nodes, and the N nodes correspond to the N node vectors; and

randomly select one node from the node set as a sampling node, and perform selections for a total of S times to obtain the S nodes.

11. A server, comprising a memory, a processor, and a computer program that is stored in the memory and that can run on the processor, wherein when executing the program, the processor implements steps of the method according to any one of claims 1 to 5.

12. A computer-readable storage medium, wherein the computer-readable storage medium stores a computer program, and when executed by a processor, the program implements steps of the method according to any one of claims 1 to 5.