

CLAIMS

What is claimed is:

1. A computer-implemented data processing method, comprising:

obtaining a continuous feature;

5 performing discretization processing on the continuous feature by using a discretization model, to obtain N discretization probabilities corresponding to the continuous feature, wherein N is an integer greater than 1, and

each of the N discretization probabilities corresponds to a respective one of N preset meta-embeddings; and

10 determining a vector representation value of the continuous feature based on the N discretization probabilities and the N meta-embeddings,

inputting a user feature and an object feature into a recommendation model or a search model to obtain a prediction result, wherein

the user feature or the object feature comprises the vector representation value; and

15 the user feature or the to-be-recommended object feature represents the continuous feature by using the vector representation value,

obtaining an actual result; and

adjusting a weight parameter of the discretization model based on the prediction result and the actual result by using a loss function.

20 2. The method according to claim 1, wherein the performing discretization processing on the continuous feature by using a discretization model, to obtain N discretization probabilities corresponding to the continuous feature comprises:

presetting an initial variable in the discretization model;

25 determining, based on the initial variable, N mapping values corresponding to the continuous feature; and

calculating the N discretization probabilities of the continuous feature based on the mapping value.

30 3. The method according to claim 1 or 2, wherein the discretization model is a multiclass

neural network, an attention network, or linear mapping and softmax.

4. The method according to claim 3, wherein when the discretization model is linear mapping and softmax, the presetting an initial variable in the discretization model comprises:

presetting an initialized linear mapping variable to $W_{logit} = R^{1 \times h}$, wherein the initialized

5 linear mapping variable is the initial variable;

the determining, based on N, N mapping values corresponding to the continuous feature comprises:

determining, according to a linear mapping formula, the N mapping values corresponding to the continuous feature, wherein

10 the linear mapping formula is $cont_{logit} = cont \cdot W_{logit}$; and

the calculating the N discretization probabilities of the continuous feature based on the mapping value comprises:

calculating, according to a discretization formula, a probability corresponding to each of the N mapping values to obtain N probabilities, wherein the discretization formula is

15 $cont_p_k = e^{\frac{1}{\tau} cont_{logit}^k} / \sum_{i=1}^h e^{\frac{1}{\tau} cont_{logit}^i}$, and the N probabilities are used as the N discretization probabilities; and

W_{logit} indicates a linear mapping variable, R indicates a real number field, h indicates a quantity of buckets into which the continuous feature is discretized, h is equal to N, $cont_{logit}$

indicates a representation obtained after linear mapping of the continuous feature, $cont_p_k$

20 indicates a probability that the continuous feature is discretized to a k^{th} bucket, $cont_{logit}^k$

indicates a k^{th} neuron output after linear mapping of the continuous feature, τ indicates a temperature control coefficient of softmax, and $cont_{logit}^i$ indicates an i^{th} neuron output after linear mapping of the continuous feature.

25 5. The method according to any one of claims 1 to 4, wherein the determining a vector representation value of the continuous feature based on the N discretization probabilities and the N meta-embeddings comprises:

determining the vector representation value of the continuous feature by using an aggregate

function and based on the N discretization probabilities and the N meta-embeddings, wherein the aggregate function is Max-Pooling, Top-K-Sum, or Weighted-Average.

5 6. The method according to any one of claims 1 to 5, wherein N is greater than or equal to 20 and less than or equal to 100.

 7. The method according to any one of claims 1 to 6, wherein the continuous feature is feature data having a continuous statistical feature value in sample data.
10

 8. The method according to claim 7, wherein the continuous feature comprises, but is not limited to, an age feature, a click count feature, and a score feature in a recommendation system.

 9. The method according to any one of claims 1 to 8, wherein the continuous feature is a
15 normalized continuous feature.

 10. A data processing apparatus, comprising:
 an obtaining module, configured to obtain a continuous feature; and
 a processing module, configured to: perform discretization processing on the continuous
20 feature by using a discretization model, to obtain N discretization probabilities corresponding to the continuous feature, wherein N is an integer greater than 1, and each of the N discretization probabilities corresponds to a respective one of N preset meta-embeddings; and determine a vector representation value of the continuous feature based on the N discretization probabilities and the N meta-embeddings; and input a user feature and an object feature into a recommendation model
25 or a search model to obtain a prediction result, wherein the user feature or the object feature comprises the vector representation value; and the user feature or the to-be-recommended object feature represents the continuous feature by using the vector representation value;
 wherein the obtaining module is further configured to obtain an actual result; and
 the processing module is further configured to adjust a weight parameter of the discretization
30 model based on the prediction result and the actual result by using a loss function.

11. The apparatus according to claim 10, wherein the processing module is specifically configured to: preset an initial variable in the discretization model, determine, based on the initial variable, N mapping values corresponding to the continuous feature, and calculate the N discretization probabilities of the continuous feature based on the mapping value.

5

12. The apparatus according to claim 10 or 11, wherein the discretization model is a multiclass neural network, an attention network, or linear mapping and softmax.

13. The apparatus according to claim 12, wherein when the discretization model is linear mapping and softmax, the processing module is specifically configured to: preset an initialized linear mapping variable to $W_{logit} = R^{1 \times h}$, wherein the initialized linear mapping variable is the initial variable;

determine, according to a linear mapping formula, the N mapping values corresponding to the continuous feature, wherein

15 the linear mapping formula is $cont_{logit} = cont \cdot W_{logit}$; and

calculate, according to a discretization formula, a probability corresponding to each of the N mapping values to obtain N probabilities, wherein the discretization formula is

$cont_p_k = e^{\frac{1}{\tau} cont_{logit}^k} / \sum_{i=1}^h e^{\frac{1}{\tau} cont_{logit}^i}$, and the N probabilities are used as the N discretization probabilities; and

20 W_{logit} indicates a linear mapping variable, R indicates a real number field, h indicates a quantity of buckets into which the continuous feature is discretized, h is equal to N, $cont_{logit}$ indicates a representation obtained after linear mapping of the continuous feature, $cont_p_k$ indicates a probability that the continuous feature is discretized to a k^{th} bucket, $cont_{logit}^k$ indicates a k^{th} neuron output after linear mapping of the continuous feature, τ indicates a temperature control coefficient of softmax, and $cont_{logit}^i$ indicates an i^{th} neuron output after linear mapping of the continuous feature.

14. The apparatus according to any one of claims 10 to 13, wherein the processing module is

specifically configured to: determine, by using an aggregate function and based on the N discretization probabilities and the N meta-embeddings, the vector representation value corresponding to the continuous feature, wherein the aggregate function is Max-Pooling, Top-K-Sum, or Weighted-Average.

5

15. The apparatus according to any one of claims 10 to 14, wherein N is greater than or equal to 20 and less than or equal to 100.

10 16. The apparatus according to any one of claims 10 to 15, wherein the continuous feature is feature data having a continuous statistical feature value in sample data.

17. The apparatus according to claim 16, wherein the continuous feature comprises, but is not limited to, an age feature, a click count feature, and a score feature in a recommendation system.

15 18. The apparatus according to any one of claims 10 to 16, wherein the continuous feature is a normalized continuous feature.

20 19. A data processing apparatus, comprising at least one processor and a memory, wherein the processor is configured to be coupled to the memory, and the processor invokes instructions stored in the memory to control the data processing apparatus to perform the method according to any one of claims 1 to 9.

25 20. A computer storage medium, wherein the computer-storage medium stores computer instructions, and the computer instructions are used to perform the method according to any one of claims 1 to 9.

30 21. A computer program product comprising instructions, wherein when the computer program product runs on a computer, the computer is enabled to perform the method according to any one of claims 1 to 9.