CLAIMS

1. A method for cross-modal information retrieval, comprising:

acquiring first modal information and second modal information;

determining a first semantic feature of the first modal information and a first attention feature of the first modal information according to a modal feature of the first modal information;

determining a second semantic feature of the second modal information and a second attention feature of the second modal information according to a modal feature of the second modal information; and

determining a similarity between the first modal information and the second modal information based on the first attention feature, the second attention feature, the first semantic feature, and the second semantic feature.

2. The method of claim 1,

wherein the first semantic feature comprises a first branch semantic feature and a first overall semantic feature, wherein the first attention feature comprises a first branch attention feature and a first overall attention feature,

wherein the second semantic feature comprises a second branch semantic feature and a second overall semantic feature, wherein the second attention feature comprises a second branch attention feature and a second overall attention feature.

3. The method of claim 2, wherein determining the first semantic feature of the first modal information and the first attention feature of the first modal information according to the modal feature of the first modal information comprises:

dividing the first modal information into at least one information unit;

determining a first modal feature of each information unit of the at least one information unit by performing first modal feature extraction on the each information unit;

extracting the first branch semantic feature in a semantic feature space based on the first modal feature of the each information unit; and

extracting the first branch attention feature in an attention feature space based on the first modal feature of the each information unit.

4. The method of claim 3, further comprising:

determining the first overall semantic feature of the first modal information according to the first branch semantic feature of the each information unit; and

determining the first overall attention feature of the first modal information according to the first branch attention feature of the each information unit.

5. The method of claim 2, wherein determining the second semantic feature of the second modal information and the second attention feature of the second modal information according to the modal feature of the second modal information comprises:

dividing the second modal information into at least one information unit;

determining a second modal feature of each information unit of the at least one information unit by performing second modal feature extraction on the each information unit;

extracting the second branch semantic feature in a semantic feature space based on the second modal feature of the each information unit; and

extracting the second branch attention feature in an attention feature space based on the second modal feature of the each information unit.

6. The method of claim 5, further comprising:

determining the second overall semantic feature of the second modal information according to the second branch semantic feature of the each information unit; and

determining the second overall attention feature of the second modal information according to the second branch attention feature of the each information unit.

7. The method of claim 2, wherein determining the similarity between the first modal information and the second modal information based on the first attention feature, the second attention feature, the first semantic feature, and the second semantic feature comprises:

determining first attention information according to the first branch attention feature

of the first modal information and the first branch semantic feature of the first modal information and the second overall attention feature of the second modal information;

determining second attention information according to the second branch attention feature of the second modal information and the second branch semantic feature of the second modal information and the first overall attention feature of the first modal information; and

determining the similarity between the first modal information and the second modal information according to the first attention information and the second attention information.

8. The method of claim 7, wherein determining the first attention information according to the first branch attention feature of the first modal information and the first branch semantic feature of the first modal information and the second overall attention feature of the second modal information comprises:

determining attention information of the second modal information for each information unit of the first modal information according to the first branch attention feature of the first modal information and the second overall attention feature of the second modal information; and

determining the first attention information of the second modal information for the first modal information according to the attention information of the second modal information for the each information unit of the first modal information and the first branch semantic feature of the first modal information.

9. The method of claim 7, wherein determining the second attention information according to the second branch attention feature of the second modal information and the second branch semantic feature of the second modal information and the first overall attention feature of the first modal information comprises:

determining attention information of the first modal information for each information unit of the second modal information according to the second branch attention feature of the second modal information and the first overall attention feature of the first modal information; and

determining the second attention information of the first modal information for the second modal information according to the attention information of the first modal information for the each information unit of the second modal information and the second branch semantic feature of the second modal information.

10. The method of any one of claims 1 to 9, wherein the first modal information is to-be-retrieved information of a first mode, wherein the second modal information is pre-stored information of a second mode, wherein the method further comprises:

in response to the similarity meeting a preset condition, determining the second modal information as a retrieval result for the first modal information.

11. The method of claim 10, wherein there are multiple pieces of the second modal information, wherein in response to the similarity meeting the preset condition, determining the second modal information as the retrieval result for the first modal information comprises:

acquiring a ranking result by ranking the multiple pieces of the second modal information according to a similarity between the first modal information and each of the multiple pieces of the second modal information;

determining second modal information meeting the preset condition according to the ranking result; and

determining the second modal information meeting the preset condition as the retrieval result for the first modal information.

12. The method of claim 11, wherein the preset condition comprises any one of:

the similarity being greater than a preset value, or a rank of the similarity acquired by ranking similarities in an ascending order being greater than a preset rank.

13. The method of claim 10, further comprising: after determining the second modal information as the retrieval result for the first modal information,

outputting the retrieval result to a user terminal.

14. The method of any one of claims 1 to 13, wherein the first modal information comprises any one of text information or image information, wherein the second modal information comprises any one of the text information or the image information.

15. The method of any one of claims 1 to 14, wherein the first modal information is training sample information of a first mode, wherein the second modal information is training sample information of a second mode, wherein each piece of the training sample information of the first mode and each piece of the training sample information of the second mode form a training sample pair.

16. A device for cross-modal information retrieval, comprising an acquiring module, a first determining module, a second determining module, and a similarity determining module,

wherein the acquiring module is adapted to acquiring first modal information and second modal information,

wherein the first determining module is adapted to determining a first semantic feature of the first modal information and a first attention feature of the first modal information according to a modal feature of the first modal information,

wherein the second determining module is adapted to determining a second semantic feature of the second modal information and a second attention feature of the second modal information according to a modal feature of the second modal information; and

wherein the similarity determining module is adapted to determining a similarity between the first modal information and the second modal information based on the first attention feature, the second attention feature, the first semantic feature, and the second semantic feature.

17. The device of claim 16,

wherein the first semantic feature comprises a first branch semantic feature and a first overall semantic feature, wherein the first attention feature comprises a first branch attention feature and a first overall attention feature,

wherein the second semantic feature comprises a second branch semantic feature and a second overall semantic feature, wherein the second attention feature comprises a second branch attention feature and a second overall attention feature.

18. A device for cross-modal information retrieval, comprising a processor and memory,

wherein the memory is adapted to storing an instruction executable by the processor,

wherein the processor is adapted to executing the executable instruction stored in the memory to implement the method of any one of claims 1 to 15.

19. A non-transitory computer-readable storage medium, having stored therein computer program instructions which, when executed by a processor, implement the method of any one of claims 1 to 15.

20. A computer program, when executed by a processor, implementing the method of any one of claims 1 to 15.