CLAIMS

1. An information processing system comprising:

a first acquisition unit configured to acquire a first data set relating to first entities, the first data set describing a feature of each of the first entities;

a second acquisition unit configured to acquire a second data set relating to second entities, the second data set describing a feature of each of the second entities;

a dimension reduction unit configured to perform dimension reduction processing on a group of first feature vectors identified from the first data set, each of the first feature vectors representing a feature of a corresponding one of the first entities, and a group of second feature vectors identified from the second data set, each of the second feature vectors representing a feature of a corresponding one of the second entities, to generate a group of first low-dimensional feature vectors corresponding to the group of the first feature vectors and a group of second low-dimensional feature vectors corresponding to the group of second feature vectors, the group of the second low-dimensional feature vectors having the same number of dimensions as the group of the first low-dimensional feature vectors; and

an association unit configured to associate each of the first entities with at least one of the second entities based on the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors, wherein

the association unit associates each of the first entities with at least one of the second entities based on similarity between the first entities specified from the group of the first low-dimensional feature vectors and similarity between the second entities specified from the group of the second low-dimensional feature vectors such that mutual relationship between the first entities regarding the similarity matches mutual relationship between the second entities.

2. An information processing system comprising:

a first acquisition unit configured to acquire a first data set relating to first entities, the first data set describing a feature of each of the first entities;

a second acquisition unit configured to acquire a second data set relating to second entities, the second data set describing a feature of each of the second entities;

a dimension reduction unit configured to perform dimension reduction processing on a group of first feature vectors identified from the first data set, each of the first feature vectors representing a feature of a corresponding one of the first entities, and a group of second feature vectors identified from the second data set, each of the second feature vectors representing a feature of a corresponding one of the second entities, to generate a group of first low-dimensional feature vectors corresponding to the group of the first feature vectors and a group of second low-dimensional feature vectors corresponding to the group of second feature vectors, the group of the second low-dimensional feature vectors having the same number of dimensions as the group of the first low-dimensional feature vectors; and

an association unit configured to associate each of the first entities with at least one of the second entities based on the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors, wherein:

the group of the first low-dimensional feature vectors is defined by a first feature space;

the group of the second low-dimensional feature vectors is defined by a second feature space; and

the association unit searches for a mapping for mapping the first entities on the first feature space to the second feature space such that distribution of the first entities in the first feature space specified from the group of the first low-dimensional feature vectors matches distribution of the second entities in the second feature space specified from the group of the second low-dimensional feature vectors, and associates each of the first entities with at least one of the second entities based on the mapping.

3. An information processing system comprising:

a first acquisition unit configured to acquire a first data set relating to first entities, the first data set describing a feature of each of the first entities;

a second acquisition unit configured to acquire a second data set relating to second entities, the second data set describing a feature of each of the second entities;

a dimension reduction unit configured to perform dimension reduction processing on a group of first feature vectors identified from the first data set, each of the first feature vectors representing a feature of a corresponding one of the first entities, and a group of second feature vectors identified from the second data set, each of the second feature vectors representing a feature of a corresponding one of the second entities, to generate a group of first low-dimensional feature vectors corresponding to the group of the first feature vectors and a group of second low-dimensional feature vectors corresponding to the group of second feature vectors, the group of the second low-dimensional feature vectors having the same number of dimensions as the group of the first low-dimensional feature vectors; and

an association unit configured to associate each of the first entities with at least one of the second entities based on the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors, wherein:

the association unit searches, as a matrix Ω^*, a matrix Ω that maximizes a value $Z(\Omega)$ according to the following formula including a matrix K, a matrix L, and a matrix H,

[Mathematical formula 1]

$$\Omega^* = \mathrm{argmax}_{\Omega} Z(\Omega)$$
$$Z(\Omega) = \mathrm{trace}(K'\Omega^{T}L'\Omega)$$
$$K' = HKH$$
$$L' = HLH$$

and is configured to associate each of the first entities with at least one of the second entities based on the matrix Ω^*;

the number of the first entities is N;

the number of the second entities is the same as that of the first entities;

the matrix K is a first similarity matrix with N rows and N columns in which a value of an element in an i-th row and a j-th column represents similarity between an i-th entity and a j-th entity among the first entities, and is calculated based on a first low-dimensional feature vector of the i-th entity among the first entities and a first low-dimensional feature vector of the j-th entity among the first entities;

the matrix L is a second similarity matrix with N rows and N columns in which a value of an element in an i-th row and a j-th column represents similarity between an i-th entity and a j-th entity among the second entities, and is calculated based on a second low-dimensional feature vector of the i-th entity among the second entities and a second low-dimensional feature vector of the j-th entity among the second entities; and

the matrix H is a matrix with N rows and N columns in which a value of an element in an i-th row and a j-th column indicates a value of 1-1/N when i = j and indicates a value of 0 when $i \neq j$.

4. The information processing system according to claim 3, wherein

the association unit is configured to improve the matrix Ω* by repeatedly performing a re-search processing for the matrix Ω* until a predetermined condition is satisfied, and associate each of the first entities with at least one of the second entities based on the improved matrix Ω*, and

the re-search processing includes:

changing a dimension reduction method in the dimension reduction processing based on the matrix Ω*; and

causing the dimension reduction unit to perform the dimension reduction processing based on the dimension reduction method changed, and

re-searching the matrix Ω* based on the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors newly obtained by the dimension reduction processing.

5. The information processing system according to claim 4, wherein the association unit changes the dimension reduction method such that a distance in a feature space between a first low-dimensional feature vector and a second low-dimensional feature vector corresponding to each other in the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors is shortened.

6. The information processing system according to any one of claims 1 to 5, wherein:

the first data set includes two or more first feature data, and each of the two or more first feature data represents a feature of a corresponding one of the first entities;

the second data set includes two or more second feature data, and each of the two or more second feature data represents a feature of a corresponding one of the the second entities; and

the information processing system further comprises

a data fusion unit that generates an extended data set including two or more extended data, each of the two or more extended data being a combined data of one of the two or more first feature data and a corresponding one of the two or more second feature data, by combining each of the two or more first feature data with one of the two or more second feature data based on association between the first entities and the second entities by the association unit.

7. The information processing system according to any one of claims 1 to 6, wherein the first entities and the second entities are persons,

the first data set is a data set describing a first feature of each person belonging to a first group, and

the second data set is a data set describing a second feature of each person belonging to a second group.

8. The information processing system according to claim 7, wherein a combination of the first feature and the second feature is a combination of a feature related to purchasing behavior, and a feature related to movement in at least one of an online space and an offline space and/or a feature related to visits to two or more points in at least one of the online space and the offline space.

9. The information processing system according to any one of claims 1 to 8, wherein

the first entities and the second entities are persons,

identification information of an information terminal corresponding to each of the second entities is associated with the second data set, and

the information processing system further comprises:

a selection unit that selects, as delivery destinations of information content, at least a part of a set of second entities associated with any one of the first entities by the association unit; and

a delivery unit configured to deliver the information content to a set of information terminals corresponding to the delivery destinations of the information content based on the identification information.

10. The information processing system according to claim 9, wherein the selection unit selects, as delivery destinations of the information content, a first set that is the set of second entities associated by the association unit with any one of the first entities, and a second set having similar features to the first set among the second entities.

11. The information processing system according to any one of claims 1 to 10, wherein the first entities and the second entities are persons,

the second data set describes a feature related to behavior of each of the second entities,

the information processing system further includes

an estimation unit that calculates, for each of one or more entities of interest, an estimation value regarding behavior of a corresponding entity of interest,

the one or more entities of interest are at least some of the first entities, and

the estimation value is calculated based on a feature related to the behavior of at least one of the second entities associated with the corresponding entity of interest.

12. A computer-implemented information processing method comprising:

acquiring a first data set relating to first entities, the first data set describing a feature of each of the first entities;

acquiring a second data set relating to second entities, the second data set describing a feature of each of the second entities;

performing dimension reduction processing on a group of first feature vectors identified from the first data set, each of the first feature vectors representing a feature of a corresponding one of the first entities, and a group of second feature vectors identified from the second data set, each of the second feature vectors representing a feature of a corresponding one of the second entities, to generate a group of first low-dimensional feature vectors corresponding to the group of the first feature vectors and a group of second low-dimensional feature vectors corresponding to the group of second feature vectors, the group of the second low-dimensional feature vectors having the same number of dimensions as the group of the first low-dimensional feature vectors; and

associating each of the first entities with at least one of the second entities based on the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors, wherein

the associating includes associating each of the first entities with one of the second entities based on similarity between the first entities specified from the group of the first low-dimensional feature vectors and similarity between the second entities specified from the group of the second low-dimensional feature vectors such

that mutual relationship between the first entities regarding the similarity matches mutual relationship between the second entities.

13. A computer-implemented information processing method comprising:

acquiring a first data set relating to first entities, the first data set describing a feature of each of the first entities;

acquiring a second data set relating to second entities, the second data set describing a feature of each of the second entities;

performing dimension reduction processing on a group of first feature vectors identified from the first data set, each of the first feature vectors representing a feature of a corresponding one of the first entities, and a group of second feature vectors identified from the second data set, each of the second feature vectors representing a feature of a corresponding one of the second entities, to generate a group of first low-dimensional feature vectors corresponding to the group of the first feature vectors and a group of second low-dimensional feature vectors corresponding to the group of second feature vectors, the group of the second low-dimensional feature vectors having the same number of dimensions as the group of the first low-dimensional feature vectors; and

associating each of the first entities with at least one of the second entities based on the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors, wherein:

the group of the first low-dimensional feature vectors is defined by a first feature space;

the group of the second low-dimensional feature vectors is defined by a second feature space; and

the associating includes searching for a mapping for mapping the first entities on the first feature space to the second feature space such that distribution of the first entities in the first feature space specified from the group of the first low-dimensional feature vectors matches distribution of the second entities in the second feature space specified from the group of the second low-dimensional feature vectors,

and associating each of the first entities with at least one of the second entities based on the mapping.

14. A computer-implemented information processing method comprising:

acquiring a first data set relating to first entities, the first data set describing a feature of each of the first entities;

acquiring a second data set relating to second entities, the second data set describing a feature of each of the second entities;

performing dimension reduction processing on a group of first feature vectors identified from the first data set, each of the first feature vectors representing a feature of a corresponding one of the first entities, and a group of second feature vectors identified from the second data set, each of the second feature vectors representing a feature of a corresponding one of the second entities, to generate a group of first low-dimensional feature vectors corresponding to the group of the first feature vectors and a group of second low-dimensional feature vectors corresponding to the group of second feature vectors, the group of the second low-dimensional feature vectors having the same number of dimensions as the group of the first low-dimensional feature vectors; and

associating each of the first entities with at least one of the second entities based on the group of the first low-dimensional feature vectors and the group of the second low-dimensional feature vectors, wherein:

the associating includes searching, as a matrix Ω^*, a matrix Ω that maximizes a value $Z(\Omega)$ according to the following formula including a matrix K, a matrix L, and a matrix H,

[Mathematical formula 2]

$$\Omega^* = \mathrm{argmax}_{\Omega} Z(\Omega)$$
$$Z(\Omega) = \mathrm{trace}(K'\Omega^T L'\Omega)$$
$$K' = HKH$$
$$L' = HLH$$

and associating each of the first entities with at least one of the second entities based on the matrix Ω*;

the number of the first entities is N;

the number of the second entities is the same as that of the first entities;

the matrix K is a first similarity matrix with N rows and N columns in which a value of an element in an i-th row and a j-th column represents similarity between an i-th entity and a j-th entity among the first entities, and is calculated based on a first low-dimensional feature vector of the i-th entity among the first entities and a first low-dimensional feature vector of the j-th entity among the first entities;

the matrix L is a second similarity matrix with N rows and N columns in which a value of an element in an i-th row and a j-th column represents similarity between an i-th entity and a j-th entity among the second entities, and is calculated based on a second low-dimensional feature vector of the i-th entity among the second entities and a second low-dimensional feature vector of the j-th entity among the second entities; and

the matrix H is a matrix with N rows and N columns in which a value of an element in an i-th row and a j-th column indicates a value of 1-1/N when i = j and indicates a value of 0 when $i \neq j$.

15. A computer-readable recording medium storing a computer program including instructions for causing a computer to perform the information processing method according to any one of claims 12 to 14.