

CLAIMS

1. Method for creating avatars from an initial sensitive data set stored in a database of a computer system, said initial data comprising attributes relating to a plurality of individuals, the method being of the type that identifies the closest neighbors of an individual among the other individuals in the initial data set, characterized in that it comprises:

(a) choosing, centered on a given individual, for attributes relating to said individual, a number (k) of nearest neighbors to be used from all the individuals in the initial data set,

(b) identifying, for that individual, the k nearest neighbors from among the other individuals in the data set,

(c) generating, for at least one attribute relating to said individual, a new attribute value from quantities which are characteristic of said attribute in the identified k nearest neighbors and weighted by a coefficient, and

(d) creating avatar data comprising the new attribute value(s), so as to ensure the sensitive data relating to the individual are non-identifiable.

2. Method according to claim 1, wherein the number (k) of nearest neighbors is chosen according to a sensitivity factor of the initial data and/or a confidence factor in the intended recipient of the avatar data, the higher the sensitivity factor and/or the lower the confidence factor, the higher the number of nearest neighbors, and the lower the sensitivity factor and/or the higher the confidence factor, the lower the number of nearest neighbors.

3. Method according to either claim 1 or claim 2, wherein a multivariate analysis, in particular a principal component analysis, is carried out on the initial data, prior to identifying the k nearest neighbors of the individual, the attribute values thus modified being then in particular projected in the same Euclidean space.

4. Method according to any one of the preceding claims, wherein the initial data comprising at least one attribute that has a numeric value, the new attribute value corresponds to a stochastically weighted barycenter (G) calculated from the values of that same attribute for the k nearest neighbors, each weighted by a randomly chosen coefficient.

5. Method according to the preceding claim, wherein a different coefficient is chosen randomly for each value of the attribute of each of the k nearest neighbors.

6. Method according to either claim 4 or claim 5, wherein the stochastically weighted barycenter (G) is calculated in a number (N) of dimensions corresponding to the number of attributes that have a numeric value.
7. Method according to any one of the preceding claims, wherein the initial data comprising at least one attribute belonging to a class representing a characteristic of the individual, the new attribute value is chosen randomly according to the weighted representation of said class among the k nearest neighbors, calculated according to the number of values belonging to said class in the k nearest neighbors, weighted by the number of values belonging to said class in the entire initial data set.
8. Method according to any one of the preceding claims, wherein the number (k) of nearest neighbors is variable, being different for some individuals, or from one individual to another, for the same initial data set.
9. Method according to any one of the preceding claims, wherein at least one constraint function is applied to the avatar data in order to limit aberrant values or a change in the relationship between attributes relating to the same individual, in particular to restrict the distributions of numerical attributes or to respect a logical relationship between class attributes.
10. Method according to any one of the preceding claims, wherein the avatar data are recorded in a database of the same computer system and/or of a remote server.
11. Method according to any one of the preceding claims, wherein the recipients of the avatar data are advised that said data has been made available by a hybrid authentication system comprising at least two steps, based in particular on generating and transmitting a unique download link, for example by email, then the communicating a limited-life key, for example by text message.
12. Device for creating avatars from an initial sensitive data set stored in a database of a computer system, said initial data comprising attributes relating to a plurality of individuals, the device being of the type that identifies the nearest neighbors of an individual among the other individuals in the initial data set, characterized in that it comprises:

(a) a choice module configured to choose, being centered on a given individual, for attributes relating to said individual, a number (k) of nearest neighbors to be used from all the individuals in the initial data set,

(b) an identification module configured to identify, for said individual, the k nearest neighbors from among the other individuals in the data set,

(c) a generation module configured to generate, for at least one attribute relating to said individual, a new attribute value only from quantities which are characteristic of said attribute in the identified k nearest neighbors and weighted by a coefficient, and

(d) an avatar creation module configured to create avatar data comprising the new attribute value(s), so as to ensure the sensitive data relating to the individual are non-identifiable.

13. Device according to the preceding claim, which also comprises a recording module configured to record the avatar data in a database of the same computer system and/or of a remote server.

14. Device according to either claim 12 or claim 13, comprising an advice module configured to advise the recipients of the avatar data that said data have been made available by a hybrid authentication system comprising at least two steps, based in particular on generating and transmitting a unique downloading link, for example by email, then communicating a limited-life key, for example by text message.

15. Computer program product for implementing the method for creating avatars from an initial sensitive data set stored in a database of a computer system, according to any one of claims 1 to 12, said initial data containing attributes relating to a plurality of individuals, the computer program product comprising a support and instructions recorded on said support that can be read by a processor in order, when executed, to:

(a) choose, being centered on a given individual, for attributes relating to said individual, a number (k) of nearest neighbors to be used from all the individuals in the initial data set,

(b) identify, for said individual, the k nearest neighbors from among the other individuals in the initial data set,

(c) generate, for at least one attribute relating to said individual, a new attribute value only from quantities which are characteristic of said attribute in the identified k nearest neighbors and weighted by a coefficient, and

(d) create avatar data comprising the new attribute value(s), so as to ensure the sensitive data relating to the individual are non-identifiable.

CLAIMS

1. A method for creating avatars from an initial sensitive data set stored in a database of a computer system, the initial data comprising attributes relating to a plurality of individuals, the method being of the type that identifies the nearest neighbours of an individual among the other individuals in the initial data set, the method comprising:
 - (a) choosing, centered on a given individual, for attributes relating to the individual, a number (k) of nearest neighbours to be used from all the individuals in the initial data set,
 - (b) identifying, for that individual, the k nearest neighbours from among the other individuals in the data set,
 - (b) generating, for at least one attribute relating to the individual, a new attribute value from quantities which are characteristic of the attribute in the identified k nearest neighbours and weighted by a coefficient, and
 - (c) creating avatar data comprising the new attribute value(s), so as to ensure the sensitive data relating to the individual are non-identifiable.
2. A method according to Claim 1, in which the number (k) of nearest neighbours is chosen according to a sensitivity factor of the initial data and/or a confidence factor in an intended recipient of the avatar data, the higher the sensitivity factor and/or the lower the confidence factor, the higher the number of nearest neighbours, and the lower the sensitivity factor and/or the higher the confidence factor, the lower the number of nearest neighbours.
3. A method according to Claim 1 or 2, in which a multivariate analysis, in particular a principal component analysis, is carried out on the initial data, prior to identifying the k nearest neighbours of the individual, the attribute values thus modified being then in particular projected in the same Euclidean space.
4. A method according to Claim 1, 2 or 3, in which the initial data comprising at least one attribute that has a numeric value, the new attribute value corresponds to a stochastically weighted barycenter (G) calculated from the values of that same attribute for the k nearest neighbours, each weighted by a randomly chosen coefficient.

5. A method according to Claim 4, in which a different coefficient is chosen randomly for each value of the attribute of each of the k nearest neighbours.
6. A method according to Claim 4 or 5, in which the stochastically weighted barycenter (G) is calculated in a number (N) of dimensions corresponding to the number of attributes that have a numeric value.
7. A method according to any preceding claim, in which the initial data comprising at least one attribute belonging to a class representing a characteristic of the individual, the new attribute value is chosen randomly according to the weighted representation of the class among the k nearest neighbours, calculated according to the number of values belonging to the class in the k nearest neighbours, weighted by the number of values belonging to the class in the entire initial data set.
8. A method according to any preceding claim, in which the number (k) of nearest neighbours is variable, being different for some individuals, or from one individual to another, for the same initial data set.
9. A method according to any preceding claim, in which at least one constraint function is applied to the avatar data in order to limit aberrant values or a change in the relationship between attributes relating to the same individual, in particular to restrict the distributions of numerical attributes or to respect a logical relationship between class attributes.
10. A method according to any preceding claim, in which the avatar data are recorded in a database of the same computer system and/or of a remote server.
11. A method according to any preceding claim, in which the recipients of the avatar data are advised that the data has been made available by a hybrid authentication system comprising at least two steps, based in particular on generating and transmitting a unique download link, for example by email, then the communicating a limited-life key, for example by text message.
12. A device for creating avatars from an initial sensitive data set stored in a database of a computer system, the initial data comprising attributes relating to a plurality of individuals, the device being of the type that identifies the nearest neighbours of an individual among the other individuals in the initial data set, the device comprising:

(a) a choice module configured to choose, being centered on a given individual, for attributes relating to the individual, a number (k) of nearest neighbours to be used from all the individuals in the initial data set,

(b) an identification module configured to identify, for the individual, the k nearest neighbours from among the other individuals in the data set,

(c) a generation module configured to generate, for at least one attribute relating to the individual, a new attribute value only from quantities which are characteristic of the attribute in the identified k nearest neighbours and weighted by a coefficient, and

(d) an avatar creation module configured to create avatar data comprising the new attribute value(s), so as to ensure the sensitive data relating to the individual are non-identifiable.

13. A device according to Claim 12, which also comprises a recording module configured to record the avatar data in a database of the same computer system and/or of a remote server.

14. A device according to Claim 12 or 13, comprising an advice module configured to advise the recipients of the avatar data that the data have been made available by a hybrid authentication system comprising at least two steps, based in particular on generating and transmitting a unique downloading link, for example by email, then communicating a limited-life key, for example by text message.

15. A computer program product for implementing the method for creating avatars from an initial sensitive data set stored in a database of a computer system, according to any of Claims 1 to 12, the initial data containing attributes relating to a plurality of individuals, the computer program product comprising a support and instructions recorded on the support that can be read by a processor in order, when executed, to:

(a) choose, being centered on a given individual, for attributes relating to the individual, a number (k) of nearest neighbours to be used from all the individuals in the initial data set,

(b) identify, for the individual, the k nearest neighbours from among the other individuals in the initial data set,

(c) generate, for at least one attribute relating to the individual, a new attribute value only from quantities which are characteristic of the attribute in the identified k nearest neighbours and weighted by a coefficient, and

(d) create avatar data comprising the new attribute value(s), so as to ensure the sensitive data relating to the individual are non-identifiable.