function can be optimized using a stochastic gradient descent (Stochastic Gradient Descent) method with a minimized mean square error.

[0059] The abovementioned implementation describes an implementation in which a plurality of vehicle insurance risk prediction algorithms can be used to construct a vehicle insurance risk prediction model, so as to more accurately describe the customer risk and help the insurance industry to segment and manage risks, reduce costs, improve efficiency, and provide better services for vehicle insurance consumers.

[0060] In another implementation, at least two candidate vehicle insurance risk prediction algorithms are generated during construction of the vehicle insurance risk prediction algorithm. For example, when using a plurality of generalized linear models that use different construction targets, one or more generalized linear models, and prediction models based on the gradient boosting decision tree algorithm, or a plurality of vehicle insurance risk prediction models based on the deep neural network, the plurality of vehicle insurance risk prediction models can be evaluated by using statistics, to select a model suitable for the current service scenario. Therefore, in one or more implementations of the present disclosure, the method can further include:

Calculate a preset type of statistical indicator, where the statistical indicators include statistics used to indicate impact of output results of the candidate vehicle insurance risk prediction algorithms when the personal attribute information is used as an independent variable of the candidate vehicle insurance risk prediction algorithms; and

Compare the statistical indicators, and then select a vehicle insurance risk prediction algorithm to be used from the candidate vehicle insurance risk prediction algorithms.

[0061] These statistical indicators can specifically include various types of statistics, such as a deviance, a log-likelihood function, an AIC, a BIC, a chi-square, a p-value, etc. These statistics can be obtained through calculation by using a fixed or deformed/transformed formula. In a specific example, processed variable fields can be input to a model, and significance of the variables and that of the model are checked and compared by using statistics such as a Deviance, a log-likelihood function, an AIC, and a BIC. Some of these statistical indicators can be used to determine whether a factor or a model is significant. In an implementation scenario of the present implementation, there can be a plurality of vehicle insurance risk prediction models, including the vehicle insurance risk prediction models

CLAIMS

What is claimed is:

1. A method for predicting a vehicle insurance risk, the method comprising:

obtaining a personal attribute information of a vehicle insurance user, wherein the personal attribute information comprises at least one of a natural attribute information, a social attribute information, and a behavior data of the vehicle insurance user, the natural attribute information being associated with biological characteristics of the vehicle insurance user, the social attribute information comprising a social classification of the vehicle insurance user, and the behavior data comprising statistical information generated by a behavior of the vehicle insurance user (S2);

generating at least two candidate vehicle insurance risk prediction algorithms, wherein a first candidate vehicle insurance risk prediction algorithm of the at least two candidate vehicle insurance risk prediction algorithms comprises a neural network model and a second candidate vehicle insurance risk prediction algorithm of the at least two candidate vehicle insurance risk prediction algorithms comprises a prediction model constructed by modeling at least a first modeling target using a gradient boosting decision tree (GBDT);

inputting the personal attribute information into the first candidate vehicle insurance risk prediction algorithm to generate a first statistical indicator;

inputting the personal attribute information into the second candidate vehicle insurance risk prediction algorithm to generate a second statistical indicator;

selecting a vehicle insurance risk prediction algorithm from the at least two candidate vehicle insurance risk prediction algorithms based on a comparison of the first statistical indicator and the second statistical indicator;

processing the personal attribute information by using the vehicle insurance risk prediction algorithm configured to predict the vehicle insurance risk associated with the vehicle insurance user, the vehicle insurance risk comprising at least one of a score representing a risk of the vehicle insurance user, a type of risk associated with the vehicle insurance user, a probability of a claim, a range of a compensation payout, and a group of a risk population associated with the vehicle insurance user (S4), wherein processing the personal attribute information comprises at least one of

setting a weight of the personal attribute information,

supplementing an incomplete value in the personal attribute information,

determining a data use selection manner of repeated personal attribute information,

processing interaction effect between the personal attribute information, and

generating new personal attribute information based on the personal attribute information; and

generating a vehicle insurance for the vehicle insurance user based on the vehicle insurance risk.

2. The method according to claim 1, wherein the vehicle insurance risk prediction algorithm is configured to perform operations comprising:

collecting a preset type of personal attribute samples (S20);

determining a generalized linear model to be used and a modeling target of the generalized linear model (S22); and

using data that comprises the personal attribute samples as input of the generalized linear model, modeling the modeling target, and determining a vehicle insurance risk prediction model, wherein the vehicle insurance risk prediction model comprises at least one output result that indicates the vehicle insurance risk of the vehicle insurance user (S24).

3. The method according to claim 1, wherein the vehicle insurance risk prediction algorithm is configured to perform operations comprising:

collecting a preset type of personal attribute samples (S30); and

modeling a selected modeling target by using the gradient boosting decision tree, and determining the vehicle insurance risk prediction algorithm , wherein the selected modeling target comprises a difference between an actual modeling target value and a predicted modeling target value, a Gamma regression objective function is used as an objective function during modeling, and the actual modeling target value and the predicted modeling target value are obtained by means of calculation based on the personal attribute samples (S32).

4. The method according to claim 1, wherein the vehicle insurance risk prediction algorithm is configured to perform operations comprising:

collecting a preset type of personal attribute samples (S40); and

modeling a selected modeling target by using the gradient boosting decision tree, and determining the vehicle insurance risk prediction algorithm, wherein the selected

modeling target comprises a quotient of an actual modeling target value and a predicted modeling target value, a Tweedie regression objective function is used as an objective function during modeling, and the actual modeling target value and the predicted modeling target value are obtained by means of calculation based on the personal attribute samples (S42).

5. The method according to claim 1, wherein the vehicle insurance risk prediction algorithm is configured to perform operations comprising:

collecting a preset type of personal attribute samples (S50);

classifying the personal attribute samples into feature samples in different value ranges based on corresponding feature types (S52);

extracting feature data from the feature samples based on N specified feature types, and generating an N-dimensional discrete feature vector (S54);

mapping a single discrete feature vector of the feature sample to an M-dimensional continuous feature vector in a preset manner (S56);

concatenating continuous feature vectors corresponding to the N-dimensional discrete feature vector to form an (N*M)-dimensional continuous feature vector X; and

using the continuous feature vector X as input of a selected deep neural network to construct a vehicle insurance risk prediction model (S58).

6. The method according to any one of claims 2 to 4, wherein the selected modeling target comprises at least one of:

a loss ratio, a claim frequency, and a claim amount of the vehicle insurance user.

7. The method according to any one of claims 2 to 5, wherein the preset type of personal attribute samples comprises at least one of:

a driving habit, a professional characteristic, an identity characteristic, credit history, a consumption habit, and an income stability.

8. The method according to claim 1, wherein the method further comprises:

calculating a preset type of statistical indicator, wherein the first statistical indicator and the second statistical indicator comprise statistics used to indicate impact of output results of the at least two candidate vehicle insurance risk prediction algorithms when the personal attribute information is used as an independent variable of the at least two candidate

vehicle insurance risk prediction algorithms.

9. The method according to any one of claims 1 to 8, wherein the biological characteristics of the vehicle insurance user comprise at least one of an age, a gender, and a physical condition.

10. The method according to any one of claims 1 to 9, wherein the social classification of the vehicle insurance user comprises at least one of a job title, an occupation, a consumption status, and a credit status.

11. An apparatus for predicting a vehicle insurance risk, wherein the apparatus comprises a plurality of modules configured to perform the method according to any one of claims 1 to 10.

12. A system, comprising at least one processor and a memory configured to store an executable instruction of the processor, and the processor executes the instruction to implement data processing steps in the method according to any one of claims 1 to 10.