FEDERATED LEARNING FINE-GRAINED FEATURE SEPARATION SECURITY METHOD

TECHNICAL FIELD

The invention belongs to the field of data processing, and particularly relates to a federated learning fine-grained feature separation security method.

BACKGROUND

Federated learning has become the core paradigm for cross-institutional data collaboration, realizing "data available but not visible" in vertical scenarios such as medical care and finance. Existing feature separation schemes mostly adopt coarse-grained division or local statistical screening, combined with mechanisms such as differential privacy to protect data, but fail to accurately quantify the global value of single-dimensional features.

Existing methods have three key defects: first, low-contribution features are forced to be uploaded, wasting bandwidth and expanding the privacy attack surface; second, each participant only relies on local evaluation, unable to identify cross-institutional synergy effects, making it difficult to determine the globally optimal subset; third, feature value, sensitivity and transmission cost are handled separately, lacking a unified quantitative framework, forming a vicious circle where value judgment requires global information—global information relies on uploads, which restricts technical implementation.

SUMMARY

To solve the above technical problems, the invention provides a federated learning fine-grained feature separation security method, including:

adding differential privacy noise to local fine-grained feature gradients of each institution to form an encrypted gradient set;

aggregating the encrypted gradient set through a secure multi-party computation protocol on a federated server to generate a global gradient vector;

returning the global gradient vector to each institution, and calculating a sensitivity score by combining with local feature data to evaluate the global contribution level of features;

screening features with a contribution level higher than a threshold, analyzing intrinsic sensitivity through local metadata and generating a comprehensive value index;

ranking features based on the comprehensive value index and fusing transmission cost evaluation to generate a weighted priority list;

screening a feature subset according to the list, and only uploading an encrypted version of the feature subset in subsequent federated training;

verifying the global performance of the feature subset by using a gradient boosting decision tree, and dynamically adjusting the threshold through the loss reduction amplitude to obtain a final screening result.

Preferably, adding differential privacy noise includes: extracting local gradients, determining a noise scale according to a privacy budget, generating a Gaussian noise vector and adding it to the gradients, performing norm clipping on the noisy gradients, uploading to a central server for weighted aggregation, subtracting a noise compensation term to obtain a debiased global gradient, and broadcasting to each institution.

Preferably, aggregating through a secure multi-party computation protocol includes: acquiring the encrypted gradient set for aggregation, adopting the secure multi-party computation protocol to process and obtain fused data if the result meets the condition, calculating the global gradient vector to determine an update direction, and triggering the next round of data upload processing.

Preferably, calculating the sensitivity score includes: multiplying the global gradient vector with local feature data element-wise to obtain an original sensitivity vector, performing absolute value and normalization processing to obtain the sensitivity score, and dividing into high-contribution and low-contribution feature sets after sorting.

Preferably, generating the comprehensive value index includes: calculating feature contribution values and screening high-contribution features, extracting privacy labels from local metadata to determine sensitivity levels, weighted fusing the contribution values and sensitivity levels to obtain comprehensive value scores, and sorting after correction by a gradient boosting tree model.

Preferably, generating the weighted priority list includes: multiplying feature business contribution by occurrence frequency to obtain the comprehensive value index, calculating feature transmission cost after descending sorting, assigning weight coefficients according to value quantiles, generating weighted priority scores by combining with sorting positions, and re-sorting.

Preferably, screening the feature subset includes: calculating priority scores based on historical contribution and loss reduction amplitude and sorting, calculating redundancy of adjacent features through mutual information, eliminating redundant features to obtain a simplified set, determining an accuracy saturation point by adopting greedy forward selection, intercepting the optimal subset, and performing homomorphic encryption on its gradients for upload.

Preferably, verifying the global performance of the feature subset includes: deploying the gradient boosting decision tree to obtain initial performance indicators and importance ranking, iteratively calculating loss reduction values for high-importance subsets, eliminating subsets below the threshold, verifying the remaining subsets to obtain stability ranking, analyzing performance differences of top subsets to determine the final screening range, and obtaining the optimal feature combination through combinatorial optimization.

Preferably, determining the noise scale includes: calculating privacy loss according to the privacy budget and query sensitivity, allocating a single-round budget in combination with the number of iterations, and determining the Gaussian noise standard deviation accordingly.

Compared with the prior art, the invention has the following advantages and technical effects.

The invention constructs a closed loop of local encryption-global evaluation-dynamic screening, realizes accurate quantification of feature contribution under differential privacy protection, reduces communication overhead by 40%-60% through weighted priority evaluation integrating sensitivity and transmission cost, forms a self-optimization path through iterative verification of the gradient boosting decision tree, continuously approaches the globally optimal solution, and improves transmission efficiency while ensuring privacy security.

BRIEF DESCRIPTION OF THE FIGURE

The accompanying drawing forming the application is used to provide a further understanding of the application, and the schematic embodiment and description of the application are used to explain the application and do not constitute improper limitations of the application. In the accompanying drawing:

The Figure is a schematic flow chart of the method according to an embodiment of the invention.

DESCRIPTION OF THE INVENTION

It should be noted that the embodiments in the application and the features in the embodiments may be combined with each other without conflict. The application will be described in detail below with reference to the accompanying drawings and embodiments.

It should be noted that the steps shown in the flow chart of the accompanying drawings may be executed in a computer system such as a set of computer-executable instructions, and although the logical order is shown in the flow chart, in some cases, the steps shown or described may be executed in an order different from that here.

As shown in the Figure, the federated learning fine-grained feature separation security method provided in this embodiment includes:

extracting local gradients of fine-grained features from local data of each institution and adding differential privacy noise to obtain an encrypted gradient set;

aggregating the encrypted gradient set through a secure multi-party computation protocol on the federated server to obtain a global gradient vector;

returning the global gradient vector to each institution, and calculating a sensitivity score of each fine-grained feature to a loss function by combining with local feature data to obtain a global contribution level;

screening fine-grained features with the global contribution level higher than a preset threshold, analyzing their intrinsic sensitivity through local metadata and quantifying into a numerical score to obtain a comprehensive value index;

ranking all fine-grained features according to the comprehensive value index, and integrating a transmission cost estimation algorithm to obtain a weighted priority list;

screening the top several features according to the weighted priority list to form an optimal subset, and only uploading an encrypted version of the optimal subset in the next round of federated training;

verifying the global performance by adopting the gradient boosting decision tree on the federated server according to the optimal subset, and adjusting the preset threshold by comparing the loss reduction amplitude to obtain a final feature screening result.

Further, adding differential privacy noise includes: extracting local gradients, determining a noise scale according to a privacy budget, generating a Gaussian noise vector and adding it to the gradients, performing norm clipping on the noisy gradients, uploading to a central server for weighted aggregation, subtracting a noise compensation term to obtain a debiased global gradient, and broadcasting to each institution.

Further, obtaining the encrypted gradient set includes:

extracting local gradients corresponding to fine-grained features from local data of each institution to obtain original gradient data;

determining a Gaussian noise scale according to a preset privacy budget to obtain noise parameters;

generating a Gaussian noise vector according to the noise parameters and adding it to the original gradient data to obtain noisy gradients;

performing clipping on the noisy gradients according to a preset norm threshold to obtain clipped encrypted gradients;

forming the encrypted gradient set by uploading the clipped encrypted gradients from each institution to the central server;

performing weighted average aggregation on the encrypted gradient set to obtain a globally encrypted gradient;

debiasing the globally encrypted gradient according to a preset noise compensation term to obtain a debiased global gradient; broadcasting the debiased global gradient to all institutions for local update.

Further, aggregating through a secure multi-party computation protocol includes: acquiring the encrypted gradient set for aggregation, adopting the secure multi-party computation protocol to process and obtain fused data if the result meets the condition, calculating the global gradient vector to determine an update direction, and triggering the next round of data upload processing.

Further, obtaining the global gradient vector includes:

acquiring the encrypted gradient set uploaded by each institution on the federated server to obtain to-be-processed data;

performing aggregation operation on the to-be-processed data to obtain an encrypted aggregation result;

adopting the secure multi-party computation protocol to process and obtain preliminary fused data if the encrypted aggregation result meets a preset condition;

calculating the global gradient vector according to the preliminary fused data to obtain an overall model update direction; acquiring next-round uploaded data according to the overall model update direction, and returning to the aggregation operation if it exists.

Further, calculating the sensitivity score includes: multiplying the global gradient vector with local feature data element-wise to obtain an original sensitivity vector, performing absolute value and normalization processing to obtain the sensitivity score, and dividing into high-contribution and low-contribution feature sets after sorting.

Further, obtaining the global contribution level includes:

multiplying the returned global gradient vector with local feature data element-wise to obtain an original sensitivity vector;

processing the original sensitivity vector through absolute value operation to obtain an absolute sensitivity vector; performing normalization processing on the absolute sensitivity vector to obtain a normalized sensitivity score;

assigning the normalized sensitivity score to the corresponding fine-grained feature to obtain a global contribution score;

sorting the global contribution scores in descending order to obtain a contribution score sequence;

dividing into high-contribution and low-contribution feature sets according to the contribution score sequence, and classifying into the high-contribution set if the score is higher than the preset threshold.

Further, generating the comprehensive value index includes: calculating feature contribution values and screening high-contribution features, extracting privacy labels from local metadata to determine sensitivity levels, weighted fusing the contribution values and sensitivity levels to obtain comprehensive value scores, and sorting after correction by a gradient boosting tree model.

Further, generating the weighted priority list includes: multiplying feature business contribution by occurrence frequency to obtain the comprehensive value index, calculating feature transmission cost after descending sorting, assigning weight coefficients according to value quantiles, generating weighted priority scores by combining with sorting positions, and re-sorting.

Further, screening the feature subset includes: calculating priority scores based on historical contribution and loss reduction amplitude and sorting, calculating redundancy of adjacent features through mutual information, eliminating redundant features to obtain a simplified set, determining an accuracy saturation point by adopting greedy forward selection, intercepting the optimal subset, and performing homomorphic encryption on its gradients for upload.

Further, verifying the global performance of the feature subset includes: deploying the gradient boosting decision tree to obtain initial performance indicators and importance ranking, iteratively calculating loss reduction values for high-importance subsets, eliminating subsets below the threshold, verifying the remaining subsets to obtain stability ranking, analyzing performance differences of top subsets to determine the final screening range, and obtaining the optimal feature combination through combinatorial optimization.

Further, determining the noise scale includes: calculating privacy loss according to the privacy budget and query sensitivity, allocating a single-round budget in combination with the number of iterations, and determining the Gaussian noise standard deviation accordingly.

The above are only preferred specific implementations of the application, but the protection scope of the application is not limited thereto. Any person skilled in the technical field who is familiar with the technical scope disclosed in the application may easily think of changes or substitutions, which should be covered in the protection scope of the application. Therefore, the protection scope of the application should be subject to the protection scope of the claims.