

Claims

1. A machine learning system for processing data associated with a transaction, the machine learning system comprising:

a first processing stage comprising a recurrent neural network architecture, the recurrent neural network architecture comprising a forget gate to modify state data for a previous iteration based on data representing a time difference between a proposed transaction and a prior transaction, the recurrent neural network architecture generating output data using the modified state data and data for the proposed transaction; and

a second processing stage comprising an attention neural network architecture communicably coupled to the first processing stage, the attention neural network architecture comprising neural network layers to apply attention weights to an input feature tensor to generate output data, the input feature tensor being generated from at least the data representing a time difference and historical output data from the first processing stage, the attention weights being computed based on the input feature tensor and current output data from the first processing stage,

wherein the machine learning system is configured to map the output data from the second processing stage to a scalar value representative of a likelihood that the proposed transaction presents an anomaly within a sequence of actions,

wherein the scalar value is used to determine whether to approve or decline the proposed transaction.

2. The machine learning system of claim 1, wherein the recurrent neural network architecture comprises:

a time difference interface to receive at least data useable to output the data representing a time difference between the proposed transaction and the prior transaction;

a transaction data input interface to receive at least data for the proposed transaction;

a state input interface to receive the state data for the previous iteration, the state data being generated via a previous application of the recurrent neural network architecture to data for a set of previous transactions that include the prior transaction; and

combinatory logic to combine the modified state data and data for the proposed transaction to generate the output data, the output data comprising state data for the present iteration.

3. The machine learning system of claim 1 or claim 2, wherein the attention neural network architecture comprises:

a time difference interface to receive time difference data comprising at least the data representing a time difference and one or more time differences between the proposed transaction and one or more further prior transactions;

a historical input interface to receive historical output data from the first processing stage;
and

a current input interface to receive current output data from the first processing stage.

4. The machine learning system of claim 3, wherein the attention neural network architecture comprises:

combinatory logic to combine data derived from the time difference interface and the historical input interface to generate the input feature tensor;

at least one neural network layer to compute at least one attention query vector from the data received at the current input interface;

at least one neural network layer to compute at least one attention key vector from the input feature tensor; and

at least one neural network layer to compute at least one attention value vector from the input feature tensor.

5. The machine learning system of claim 4, wherein the attention neural network architecture comprises:

at least one time encoding layer to apply a parameterised function to the time difference data,

wherein an output of the at least one time encoding layer is used to generate the attention weights together with the at least one attention key vector and the at least one attention query vector.

6. The machine learning system of any one of claims 1 to 5, wherein the forget gate comprises a time difference encoding that implements a parameterised exponential function that is applied to the data representing a time difference to output an activation vector, the activation vector modifying the state data by way of an element-wise multiplication.

7. The machine learning system of claim 6, wherein the parameters of the parameterised exponential function are fixed to represent different time decay periods.

8. The machine learning system of any one of claims 1 to 7, wherein the attention neural network architecture comprises:

a time encoding layer to encode time difference data representing a time difference between the proposed transaction and each of a set of prior transaction,

wherein an output of the time encoding layer is combined with the historical output data from the first processing stage for use in computing value input data for computation of the attention weights.

9. The machine learning system of any one of claims 1 to 8, wherein the second processing stage comprises a plurality of attention neural network architectures that form a multi-headed self-attention mechanism, wherein the attention neural network architectures are applied in parallel and the outputs are combined to generate output data for the second processing stage.

10. The machine learning system of any one of claims 1 to 9, comprising:

a fully-connected neural network architecture comprising one or more neural network layers to receive at least output data from the second processing stage and to map the output data to the scalar value.

11. The machine learning system of claim 10, wherein the fully-connected neural network architecture also receives output data from the first processing stage to generate the scalar value.

12. The machine learning system of any one of claims 1 to 11, comprising:

a fully-connected neural network architecture comprising one or more neural network layers to pre-process the data for the proposed transaction prior to at least the first processing stage.

13. The machine learning system of any one of claims 1 to 12, wherein the scalar output is a normalised value between 0 and 1 representing the probability of the proposed transaction being an anomaly within a sequence of actions.

14. The machine learning system of any one of claims 1 to 13, comprising one or more skip connections that respectively bypass one or more of the first processing stage and the second processing stage.

15. A method of processing data associated with a proposed transaction, the method comprising:

receiving an incoming event from a client transaction processing system, the incoming event being associated with a request for an approval decision for the proposed transaction;

parsing the incoming event to extract data for the proposed transaction, including determining a time difference between the proposed transaction and a prior transaction;

applying the machine learning system of any one of claims 1 to 10 to output the scalar value representative of a likelihood that the proposed transaction presents an anomaly within a sequence of actions based on the extracted data and the time difference, said applying comprising accessing a storage repository to retrieve at least the state data for a previous iteration, historical output data from the first processing stage, and time differences between the proposed transaction and one or more further historical transactions;

determining a binary output based on the scalar value output by the machine learning system, the binary output representing whether the proposed transaction is approved or declined; and

returning the binary output to the client transaction processing system.

16. The method of claim 15, wherein the latency from receipt of the incoming event to the return of the binary output is less than a second.