

METHODS AND SYSTEMS FOR IDENTITY AUTHENTICATION

FIELD OF INVENTION

[001] The present invention relates broadly, but not exclusively, to methods and systems for identity authentication.

BACKGROUND

[002] Electronic-Know Your Customer (eKYC) is a digital due diligence process performed by a business to verify the identity of its clients and to assess potential risks of illegal intentions towards the business relationship. In the context of eKYC, as the cost to falsify an identity (e.g. identity number, name, address) on the Internet is low, fraud is getting more common and identity authentication has become an important topic.

[003] An identity can have multiple features from different aspects. Some aspects of an identity are changeable, such as contact number, user identifier (UID) and device used. Some aspects of an identity are relatively objective and less changeable, such as name, identity number and how the user looks like. These features of identities can be utilized in identity authentication.

[004] Traditional techniques of detecting fraudsters focus on attempted times of a single feature, and use simple rules to detect fraudsters (e.g. counting by time and cross counting). However, most of these techniques do not involve the use of face recognition, although face can be viewed as the most important feature of an identity due to its high-cost for falsification. Further, the relationship between different features is complex, and simple rules used by these existing techniques are insufficient to account of the complexity.

SUMMARY

[005] Described embodiments provide methods and systems for identity authentication with a high precision and recall rate. In some embodiments, the present disclosure uses identity information from different aspects, including face, document, device, to contribute to identity authentication and fraud detection. In particular, with the help of face recognition techniques, the methods and systems

include face as a feature, which is high-cost for a fraudster to falsify. Further, in some embodiments, identity network is built based on identities and their features as nodes, and relationships (e.g., equality, similarity) as edges, which takes the complex relationship between features into consideration. By representing biometric and document information as vectors, similarity or equality can be calculated more efficiently and accurately. Additionally, in some embodiments, graph mining techniques are used to replace traditional rules of fraud detection, which is more powerful, accurate, and comprehensive. Advantageously, fraudsters who use different information for the same person, or use the information from another identity can be detected with a high precision and recall rate.

[006] According to another embodiment, there is provided a method for identity authentication. The method includes: building an identity network based on history data of identities, wherein the identity network comprises nodes corresponding to each of the identities and features of each identity, and edges connecting related nodes; developing a graph mining model associated with the identity network; receiving a new identity and adding the new identity to the identity network; and detecting one or more fraud identities within the identity network including the new identity using the graph mining model.

[007] In some implementations, the graph mining model is based on one or more rules. In some implementations, the graph mining model is trained from the identity network. In some implementations, building the identity network based on the history data of identities includes: receiving at least one biometric feature and at least one document feature for each identity in the history data; training a biometric recognition model and a document embedding model from the history data; generating, by the trained biometric recognition model and the trained document embedding model, vectors representing the at least one biometric feature and the at least one document feature received; and creating nodes corresponding to each identity and features of each identity, and creating edges connecting related nodes. In some implementations, the biometric feature comprises a face feature and the biometric recognition model comprises a face recognition model.

[008] According to other embodiments, one or more of these general and specific embodiments may be implemented using an apparatus including a plurality of modules, a system, a method, or a computer-readable media, or any combination of devices, systems, methods, and computer-readable media. The foregoing and other

described embodiments can each, optionally, include some, none or all of the following embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[009] Embodiments and implementations are provided by way of example only, and will be better understood and readily apparent to one of ordinary skill in the art from the following written description, read in conjunction with the drawings, in which:

[0010] Figure 1 is a flow chart illustrating an example of a method for identity authentication, according to an embodiment.

[0011] Figure 2A is a schematic of an example of a fraud type showing a fraudster using one face with different information, according to an embodiment.

[0012] Figure 2B is a schematic of an example of a fraud type showing a fraudster using an existing user's information in a different identity, according to an embodiment.

[0013] Figure 3 is a flow chart illustrating an example of a method for building an identity network, according to an embodiment.

[0014] Figures 4A-4B are schematics of an example of an identity network including nodes and edges, according to an embodiment.

[0015] Figure 5 is a block diagram of an example of a computer system suitable for executing at least some steps of the example methods shown in Figures 1 and 3, according to an embodiment.

[0016] Figure 6 is a schematic diagram of an example of a system for identity authentication, according to an embodiment.

[0017] Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been depicted to scale. For example, the dimensions of some of the elements in the illustrations, block diagrams or flowcharts may be exaggerated in respect to other elements to help to improve understanding of the present embodiments.

DETAILED DESCRIPTION

[0018] Embodiments will be described, by way of example only, with reference to the drawings. Like reference numerals and characters in the drawings refer to like elements or equivalents.

[0019] Some portions of the description which follows are explicitly or implicitly presented in terms of algorithms and functional or symbolic representations of operations on data within a computer memory. These algorithmic descriptions and functional or symbolic representations are the means used by those skilled in the data processing arts to convey most effectively the substance of their work to others skilled in the art. An algorithm is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. The steps are those requiring physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transferred, combined, compared, and otherwise manipulated.

[0020] Unless specifically stated otherwise, and as apparent from the following, it will be appreciated that throughout the present specification, discussions utilizing terms such as “receiving”, “obtaining”, “training”, “determining”, “building”, “generating”, “calculating”, “detecting”, “indicating”, “converting”, “adding”, “comparing”, “updating”, “correcting”, “representing”, “outputting”, or the like, refer to the action and processes of a computer system, or similar electronic device, that manipulates and transforms data represented as physical quantities within the computer system into other data similarly represented as physical quantities within the computer system or other information storage, transmission or display devices.

[0021] The present specification also discloses apparatuses for performing the operations of the methods. Such apparatuses may be specially constructed for the required purposes, or may comprise a computer or other device selectively activated or reconfigured by a computer program stored in the computer. The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various machines may be used with programs in accordance with the teachings herein. Alternatively, the construction of more specialized apparatus to perform the required method steps may be

appropriate. The structure of a computer suitable for executing the various methods / processes described herein will appear from the description below.

[0022] In addition, the present specification also implicitly discloses a computer program, in that it would be apparent to the person skilled in the art that the individual steps of the method described herein may be put into effect by computer code. The computer program is not intended to be limited to any particular programming language and implementation thereof. It will be appreciated that a variety of programming languages and coding thereof may be used to implement the teachings of the disclosure contained herein. Moreover, the computer program is not intended to be limited to any particular control flow. There are many other variants of the computer program, which can use different control flows without departing from the spirit or scope of the invention.

[0023] Furthermore, one or more of the steps of the computer program may be performed in parallel rather than sequentially. Such a computer program may be stored on any computer readable medium. The computer readable medium may include storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a computer. The computer readable medium may also include a hard-wired medium such as exemplified in the Internet system, or wireless medium such as exemplified in the GSM mobile telephone system. The computer program when loaded and executed on such a computer effectively results in an apparatus that implements the steps of the preferred method.

[0024] The present invention may also be implemented as hardware modules. More particularly, in the hardware sense, a module is a functional hardware unit designed for use with other components or modules. For example, a module may be implemented using discrete electronic components, or it can form a portion of an entire electronic circuit such as an Application Specific Integrated Circuit (ASIC) or Field Programmable Gate Array (FPGA). Numerous other possibilities exist. Those skilled in the art will appreciate that the system can also be implemented as a combination of hardware and software modules.

[0025] Identity authentication is the act or process of verifying that identity, and can be considered as a form of fraud detection in which legitimacy of users are verified and potential fraudsters may be detected before fraudulent acts are

carried out. Effective authentication can enhance data security of systems by permitting only authenticated users to access its protected resources. Embodiments seek to provide methods and systems for identity authentication which detect fraudsters falsifying identities by using different information for the same person, or using the information from another identity.

[0026] The techniques described in this specification produce one or more technical effects. In some embodiments, face feature is used for identity authentication, because face is an important feature in an identity and involves higher cost to falsify. In some embodiments, an identity network is created by representing identities and their features as nodes and relationships as edges, which takes the complex relationship between features into consideration. In some embodiments, face recognition models and document embedding models are used to convert faces and documents into vectors as representation, which can improve the accuracy of comparing the faces and information of the documents. In some embodiments, graph mining techniques are used to extract patterns of interest from graphs, which are more efficient and accurate in detecting fraudsters than traditional rules.

[0027] Figure 1 is a flow chart 100 illustrating a method for identity authentication including the following steps:

- 110: building an identity network with history data of identities, wherein the identity network comprises nodes corresponding to each of the identities and features of each identity, and edges connecting related nodes;
- 120: developing a graph mining model associated with the identity network;
- 125: receiving a new identity and adding the new identity to the identity network;
- 130: detecting one or more fraud identities within the identity network including the new identity using the graph mining model; and
- 140: updating the history data by including the new identity.

[0028] At step 110, an identity network is built with history data of identities. The history data of identities may contain information on all the identities as well as features of each identity. The identities in the history data may be labelled as a valid identity or a fraud identity. Each identity in the history data may have multiple features, which may include some or all of the following: name, face, identity document (ID) number, address, mobile number, user identifier (UID) for

login, device(s) for login, or the like. The identity network may be a graph which contains nodes that correspond to identities and features of each identity. The graph may also contain edges when there is relatedness between nodes. A method of building an identity work according to an embodiment is described later with reference to Figure 3 and Figure 4A.

[0029] At step 120, a graph mining model associated with the identity network built at step 110 is developed. Graph mining is a series of techniques to extract patterns (i.e. sub-graphs) of interest from graphs that describe the underlying data. The extracted patterns or sub-graphs can have further applications such as for classification and clustering. In the context of an identity network, the extracted patterns can be used to find strongly connected groups in the network and/or central nodes in the groups that correspond to an identity or a feature. Advantageously, graph mining techniques can be used to efficiently mine possible fraud identities from the graph, by developing graph mining models that are either rule-based or trained from history data of identities.

[0030] In some embodiments, graph mining techniques may include detecting fraud based on some patterns according to certain rules, and the graph mining model can be developed based on one or more of the rules. For example, each identity should have a unique ID number. A person cannot have different ID numbers at the same time, and two people cannot have the same ID number at the same time. In some implementations, such patterns can be used to detect fraud identities using the rule-based graph mining model.

[0031] In some embodiments, some patterns may not be as simple, and a graph mining model can be developed by training from the history data to detect fraud identities. In some embodiments, the trained graph mining model can be a classifier model which uses every identity node and the features of each identity as well as neighbor identity nodes and their features as input, and then predicts if the identity node is a fraud identity. After the model is trained, the trained model can advantageously contain some patterns that are not immediately visible from the graph, but are useful to detect fraud identities with a more complex relationship.

[0032] At step 125, when a new identity is received, the new identity is added to the identity network. The new identity and features of the new identity can be

added as new nodes of the identity network, and the new nodes can be connected to existing nodes if they are related nodes. In an embodiment, related nodes can be defined as nodes having the same type of feature and a similarity of values for the feature. In an embodiment, the features may be converted into vectors as representation, and the related nodes can be defined as nodes having an equal type of feature and a similarity between the vectors. The relatedness can be represented as an edge between the nodes.

[0033] At step 130, one may detect one or more fraud identities within the identity network after the new identity has been added in, using the graph mining model developed at step 120. Some embodiments of the types of fraud will be described in more details with reference to Figures 2A and 2B. It can be appreciated that, comparing with the traditional methods that use simple rules (e.g. counting by times of attempt), graph mining techniques is more powerful and efficient in detecting fraudulent identity by considering the complex relationships between different features.

[0034] At step 140, the new identity can be added to the history data and the history data can be updated. Advantageously, the history data can include the new identity and its related features and information, which may enhance the future training models.

[0035] Figure 2A is a schematic 200 of a fraud type showing a fraudster using one face with different information. In an embodiment, the fraudster 210 may use his face 220 on many identity cards 202a-202f that contain different information (e.g., name, address) to falsify different identities.

[0036] Figure 2B is a schematic 250 of a fraud type showing a fraudster using an existing user's information in a different identity. In an embodiment, the fraudster may create an identity card 252a that contain information 260 (e.g., ID number) that is the same as an existing user. As the information 260 is already in the history data, the identity card 252a may be detected as a fraud identity.

[0037] Figure 3 is a flow chart 110 illustrating a method for building an identity network for fraud detection as described above with reference to Figure 1. The method includes the following steps:

- 112: receiving at least one biometric feature and at least one document feature for each identity in the history data;

- 114: training a biometric recognition model and a document embedding model from the history data;
- 116: generating, by the trained biometric recognition model and the trained document embedding model, vectors representing the at least one biometric feature and the at least one document feature received;
- 118: creating nodes corresponding to each identity and features of each identity, and creating edges connecting related nodes.

[0038] At step 112, at least one biometric feature and at least one document feature are received for each identity forms the history data of identities. For each identity, the features may include some or all of the following: name, face, ID type, ID number, address, mobile number, UID for login, device(s) for login, or the like, depending on which features are of interest. In some embodiments, the features at least include biometric information of an identity (e.g. face, finger print, iris, or other types of consistent biometric information) and document information of an identity (e.g. name, address, ID number, ID type, and mobile number).

[0039] At step 114, a biometric recognition model and a document embedding model are trained from the history data of identities. The biometric recognition model can be a face recognition model. Face recognition is a biometric identification technique by scanning a person's face and matching it against a library of known faces, thereby identifying and verifying the person. Example face recognition models include DeepID3, FaceNet, or the like. Document embedding is a technique to convert a document into a vector as representation. Example document embedding models include sent2vec, doc2vec, or the like.

[0040] At step 116, vectors are generated by the biometric recognition model and the document embedding model for the history data. In some embodiments, face vectors are generated for a face using the face recognition model, and document vectors are generated for documents using the document embedding model. These vectors generated for the history data will be used to build the identity network.

[0041] At step 118, identities and features are created as nodes, and edges are created for related nodes. Figures 4A-4B are schematics 400, 450 of an identity network including nodes and edges. Referring to Figure 4A, identity nodes 40a, 40b, 40c are created for Identity A, Identity B and Identity C. Each identity may

include multiple features, such as face, name, address, and ID number. Accordingly, feature nodes are created for the features of the identities, for example, feature nodes 42a, 44a, 46a and 48a for face, name, ID number, and address for Identity A, respectively. Edges are then created to connect related feature nodes. In some embodiments, related nodes refer to nodes with an equal feature and a similarity between the vectors. For example, feature nodes 48a and 48c are nodes for “address”. If Identity A’s address and Identity C’s address, after converting to document vectors at step 116, are shown to be the same, edge 408 is created. Alternatively, feature nodes 42a and 42b are nodes for “face” for Identity A and Identity B. If the face vectors are found to be similar, edge 402 is created. In this way, information of history data is converted to a graph (i.e. the identity network) formed by such nodes and edges. Referring to Figure 4B, the identity network may further include edges 410, 420, 430 between identity nodes 40a, 40b, 40c. Edge 410 is created between identity nodes 40a and 40b because the two identities have similar faces and the same name. Edge 420 is created between identity nodes 40a and 40c because the two identities have the same name and the same address. Edge 430 is created between identity nodes 40b and 40c because the two identities have the same name. Advantageously, such additional patterns among the identity nodes may also provide useful information in detecting fraud identities.

[0042] Referring back to Figure 1, in an embodiment, when a new identity is received at step 125, vectors can be generated by the biometric recognition model and the document embedding model for the information of the new identity, in a manner similar to step 116. The new identity and its related features can be created as new nodes, in a manner similar to step 118. Upon performing calculations with respect to the vectors for the new identity and the vectors for the history data, one can determine the relatedness (e.g. equality or similarity) between new nodes and nodes created based on the history data. In a manner similar to step 118, for equality between the features and similarity between the vectors, edges can be created to connect the new nodes to their related nodes. Using the graph mining model developed from the history data, one can efficiently authenticate the new identity and determine if the new identity is a fraudulent identity.

[0043] From the above embodiments, one may appreciate that the identity network built in accordance with the method 110 can contain comprehensive

information for the identities in the history data. Further, it allows for selection of features of interest and allows for consideration of the complex relationships between features.

[0044] The system, apparatus, module, or unit illustrated in the previous embodiments can be implemented by using a computer chip or an entity, or can be implemented by using a product having a certain function. A typical embodiment device is a computer (and the computer can be a personal computer), a laptop computer, a cellular phone, a camera phone, a smartphone, a personal digital assistant, a media player, a navigation device, an email receiving and sending device, a game console, a tablet computer, a wearable device, or any combination of these devices. The modules described as separate parts may or may not be physically separate, and parts displayed as modules may or may not be physical modules, may be located in one position, or may be distributed on a number of network modules. Some or all of the modules can be selected based on actual demands to achieve the objectives of the solutions of the specification. A person of ordinary skill in the art can understand and implement the embodiments of the present application without creative efforts.

[0045] Figure 5 is a block diagram of an example computer system 500 suitable for executing at least some steps of the example methods shown in Figures 1 and 3. The following description of the computer system / computing device 500 is provided by way of example only and is not intended to be limiting.

[0046] As shown in Fig. 5, the example computing device 500 includes a processor 502 for executing software routines. Although a single processor is shown for the sake of clarity, the computing device 500 may also include a multi-processor system. The processor 502 is connected to a communication infrastructure 506 for communication with other components of the computing device 500. The communication infrastructure 506 may include, for example, a communications bus, cross-bar, or network.

[0047] The computing device 500 further includes a main memory 504, such as a random access memory (RAM), and a secondary memory 510. The secondary memory 510 may include, for example, a storage drive 512, which may be a hard disk drive, a solid state drive or a hybrid drive and/or a removable storage drive 514, which may include a magnetic tape drive, an optical disk drive, a solid state storage drive (such as a USB flash drive, a flash memory device, a solid state drive or a

memory card), or the like. The removable storage drive 514 reads from and/or writes to a removable storage medium 518 in a well-known manner. The removable storage medium 518 may include magnetic tape, optical disk, non-volatile memory storage medium, or the like, which is read by and written to by removable storage drive 514. As will be appreciated by persons skilled in the relevant art(s), the removable storage medium 518 includes a computer readable storage medium having stored therein computer executable program code instructions and/or data.

[0048] In an alternative implementation, the secondary memory 510 may additionally or alternatively include other similar means for allowing computer programs or other instructions to be loaded into the computing device 500. Such means can include, for example, a removable storage unit 522 and an interface 520. Examples of a removable storage unit 522 and interface 520 include a program cartridge and cartridge interface (such as that found in video game console devices), a removable memory chip (such as an EPROM or PROM) and associated socket, a removable solid state storage drive (such as a USB flash drive, a flash memory device, a solid state drive or a memory card), and other removable storage units 522 and interfaces 520 which allow software and data to be transferred from the removable storage unit 522 to the computer system 500.

[0049] The computing device 500 also includes at least one communication interface 524. The communication interface 524 allows software and data to be transferred between computing device 500 and external devices via a communication path 526. In various embodiments of the disclosure, the communication interface 524 permits data to be transferred between the computing device 500 and a data communication network, such as a public data or private data communication network. The communication interface 524 may be used to exchange data between different computing devices 500 which such computing devices 500 form part an interconnected computer network. Examples of a communication interface 524 can include a modem, a network interface (such as an Ethernet card), a communication port (such as a serial, parallel, printer, GPIB, IEEE 1394, RJ45, USB), an antenna with associated circuitry and the like. The communication interface 524 may be wired or may be wireless. Software and data transferred via the communication interface 524 are in the form of signals which can be electronic, electromagnetic, optical or other signals capable of being received by communication interface 524. These signals are provided to the communication interface via the communication path 526.

[0050] As shown in Fig. 5, the computing device 500 further includes a display interface 528 which performs operations for rendering images to an associated display 530 and an audio interface 532 for performing operations for playing audio content via associated speaker(s) 534.

[0051] As used herein, the term "computer program product" may refer, in part, to removable storage medium 518, removable storage unit 522, a hard disk installed in storage drive 512, or a carrier wave carrying software over communication path 526 (wireless link or cable) to communication interface 524. Computer readable storage media refers to any non-transitory, non-volatile tangible storage medium that provides recorded instructions and/or data to the computing device 500 for execution and/or processing. Examples of such storage media include magnetic tape, CD-ROM, DVD, Blu-ray™ Disc, a hard disk drive, a ROM or integrated circuit, a solid state storage drive (such as a USB flash drive, a flash memory device, a solid state drive or a memory card), a hybrid drive, a magneto-optical disk, or a computer readable card such as a PCMCIA card and the like, whether or not such devices are internal or external of the computing device 500. Examples of transitory or non-tangible computer readable transmission media that may also participate in the provision of software, application programs, instructions and/or data to the computing device 500 include radio or infra-red transmission channels as well as a network connection to another computer or networked device, and the Internet or Intranets including e-mail transmissions and information recorded on Websites and the like.

[0052] The computer programs (also called computer program code) are stored in main memory 504 and/or secondary memory 510. Computer programs can also be received via the communication interface 524. Such computer programs, when executed, enable the computing device 500 to perform one or more features of embodiments discussed herein. In various embodiments, the computer programs, when executed, enable the processor 607 to perform features of the above-described embodiments. Accordingly, such computer programs represent controllers of the computer system 500.

[0053] Software may be stored in a computer program product and loaded into the computing device 500 using the removable storage drive 514, the storage drive 512, or the interface 520. The computer program product may be a non-transitory computer readable medium. Alternatively, the computer program product may be downloaded to the computer system 500 over the communication path 526. The

software, when executed by the processor 502, causes the computing device 500 to perform the necessary operations to execute the method as shown in Figures 1 and 3.

[0054] It is to be understood that the embodiment of Fig. 5 is presented merely by way of example to explain the operation and structure of the system 500. Therefore, in some embodiments one or more features of the computing device 500 may be omitted. Also, in some embodiments, one or more features of the computing device 500 may be combined together. Additionally, in some embodiments, one or more features of the computing device 500 may be split into one or more component parts.

[0055] It will be appreciated that the elements illustrated in Fig. 5 function to provide means for performing the various functions and operations of the system as described in the above embodiments.

[0056] Figure 6 is a schematic diagram of an example system 600 including modules for identity authentication. The system 600 at least includes a processor module 602 and a memory module 604. The processor module 602 and the memory module 604 are interconnected. The memory module 604 stores history data 620 and computer program code (not shown in Figure 6). The memory module 604 and the computer program code are configured to, with the processor module 602, cause the system 600 to perform the steps for identity authentication as described in the present disclosure. The system 600 can include a receiver module 606, a training module 608, a generator module 610, a calculation module 612, an output module 616 and an update module 618. With reference to Figure 1 and Figure 3, the receiver module 606 can be configured to receive identities and features of the identities. The training module 608 can be configured to train the graph mining model, the biometric recognition model and the document embedding model. The generator module 610 can be configured to generate vectors by the biometric recognition model and the document embedding model for the features of the identities. The calculation module 612 can be configured to calculate relatedness, such as equality between the features and similarity between the vectors. The output module 616 can be configured to output the result if a new identity is determined to be a fraudulent identity. The update module 618 can be configured to update the history data 620 by including the new identity received by the receiver module 606. One or

more or any combination of these modules can be part of an apparatus for identity authentication.

[0057] It will be appreciated by a person skilled in the art that numerous variations and/or modifications may be made to the present invention as shown in the specific embodiments without departing from the scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.

CLAIMS

What is claimed is:

1. A method for identity authentication, the method comprising:
building an identity network based on history data of identities, wherein the identity network comprises nodes corresponding to each of the identities and features of each identity, and edges connecting related nodes;
developing a graph mining model associated with the identity network;
receiving a new identity and adding the new identity to the identity network;
and
detecting one or more fraud identities within the identity network including the new identity using the graph mining model.
2. The method as claimed in claim 1, wherein the graph mining model is based on one or more rules.
3. The method as claimed in claim 2, wherein the one or more rules comprise a rule of each of the identities having a unique identity document (ID) number.
4. The method as claimed in claim 1, wherein the graph mining model is trained from the identity network based on the history data of identities.
5. The method as claimed in claim 4, wherein the trained graph mining model comprises a classifier model.
6. The method as claimed in claim 1, wherein adding the new identity to the identity network comprises:
creating new nodes corresponding to the new identity and features of the new identity;
determining if the new nodes are related nodes with the nodes in the identity network; and
creating new edges connecting related nodes.
7. The method as claimed in any one of the preceding claims, wherein the related nodes comprise two nodes having at least one equal or similar feature.
8. The method as claimed in claim 1, wherein building the identity network based on the history data of identities comprises:

receiving at least one biometric feature and at least one document feature for each identity in the history data;

training a biometric recognition model and a document embedding model from the history data;

generating, by the trained biometric recognition model and the trained document embedding model, vectors representing the at least one biometric feature and the at least one document feature received; and

creating nodes corresponding to each identity and features of each identity, and creating edges connecting related nodes.

9. The method as claimed in claim 8, wherein the related nodes comprise two nodes having a same type of feature and similar values of vectors.

10. The method as claimed in claim 8, wherein the biometric feature comprises a face feature and the biometric recognition model comprises a face recognition model.

11. The method as claimed in claim 8, wherein the document feature is selected from the group comprising name, identity document (ID) type, ID number, device, mobile number and address.

12. The method as claimed in any one of the preceding claims, wherein the method further comprising:

updating the history data by including the new identity.

13. A system for identity authentication, the system comprising:

one or more processors; and

one or more computer-readable memories coupled to the one or more processors and having instructions stored thereon that are executable by the one or more processors to perform the method of any one of claims 1 to 12.

14. An apparatus for identity authentication, the apparatus comprising a plurality of modules for performing the method of any one of claims 1 to 12.