SYSTEM AND METHOD FOR EARLY FAKE ACCOUNT DETECTION

TECHNICAL FIELD

[001] Various aspects of this disclosure relate to a system configured for early fake account detection. Various aspects of this disclosure relate to a method for early fake account detection. Various aspects of this disclosure relate to a non-transitory computer-readable medium storing computer executable code for early fake account detection. Various aspects of this disclosure relate to a computer executable code for early fake account detection.

BACKGROUND

[002] Fake accounts creation has been one of the most threatening attacks to e-commerce platforms and riding hailing service providers. By creating many fake accounts, fraudsters can conduct promo gaming or unfunded gaming (cashless payment), which are extremely harmful to the whole ecosystem. For example, a lot of promos go to those fake accounts instead of genuine users, which defeats the purpose of a promo campaign and results in significant losses to the company.

[003] Existing approaches for fake accounts detection look at signals such as hardware device sharing, fake bookings, and user click patterns. However, all of those models act independently which may result in late detection whereby fake accounts are only be detected after damage has already been observed on the platform. There may also be near miss detections whereby fake accounts operate just under the radar of each individual model, and none of those existing models are able to detect those fake accounts on its own.

SUMMARY

[004] Therefore, there may be a need to provide a system to accurately and/or quickly detect fake accounts early in time.

[005] Various embodiments may provide a system for early fake account detection. The system may include one or more processor(s) and a memory having instructions stored therein. The instructions when executed by the one or more processor(s), may cause the one or more processor(s) to: for each predetermined time step of a sequence of predetermined time steps, collect user registration signals when a user registers for an account; collect booking signals when the user books a ride hailing service using the

account; collect transaction signals when the user makes a transaction using the account; collect clickstream data when the user interacts with the account; and concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals; and determine whether the account is fake using a sequence of vectors of signals comprising the vector of signals for each predetermined time step. According to various embodiments, this is achieved by feeding the sequence of vectors of signals to a machine learning model trained for early account detection.

[006] Concatenating may mean that representations of the signals (user registration signals, the booking signals, the transaction signals and the clickstream data) as data words are written one after another to form a vector signal. This may then be used as machine model input.

[007] According to various embodiments, the vector of signals is obtained for each predetermined time step.

[008] According to various embodiments, the sequence of vectors of signals is obtained by concatenating the vectors of signals for each predetermined time step.

[009] According to various embodiments, the sequence of vectors of signals is used as input to a machine learning model, and the processor is configured to use the machine learning model to determine whether the account is fake.

[0010] According to various embodiments, the machine learning model is trained (i.e. optimized) for both prediction accuracy and detection time. This is achieved by training the machine learning model using training data for supervised learning which include both an indication whether an account is fake and a time step after which the account is to be determined as fake. The machine learning model is then trained (for each training data element comprising such a label) to predict whether an account is fake at the latest at the time step indicated in the label. Thus, the machine learning model is trained both for prediction (i.e. classification) accuracy as well as early detection (i.e. detection with a time limit as given by the time step indicated in the label). In deployment (inference), the machine learning model thus allows early fake account detection.

[0011] According to various embodiments, the processor is configured to determine whether the account is fake after a predetermined number T of predetermined time According to various embodiments, fake accounts are detected early in time.

[0012] According to various embodiments, the processor is configured to process the sequence of vectors of signals and embed the sequence of vectors of signals into a hidden

state to obtain an intermediate output, wherein the intermediate output is a hazard rate, and wherein the hazard rate indicates an instantaneous probability that an account ban is triggered. According to various embodiments, a final output is the survival probability of an account which takes into account the hazard rate before the current time step, wherein the survival probability is monotonically decreasing.

[0013] According to various embodiments, the processor is configured to add paddings to the sequence of vectors of signals if a current time step of the sequence of vectors of signals is less than the predetermined number T of predetermined time steps .

[0014] Various embodiments may provide a method for early fake account detection. The method may include using one or more processor(s) to: collect user registration signals when a user registers for an account; collect booking signals when the user books a ride hailing service using the account; collect transaction signals when the user makes a transaction using the account; collect clickstream data when the user interacts with the account; concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals; determine whether the account is fake using a sequence vectors of signals comprising the vector of signals for each predetermined time step.

[0015] Various embodiments may provide a non-transitory computer-readable medium storing computer executable code including instructions for early fake account detection according to the various embodiments disclosed herein.

[0016] Various embodiments may provide a computer executable code including instructions for early fake account detection according to the various embodiments disclosed herein.

[0017] It should be noted that embodiments described in context of the system are analogously valid for the method, the computer-readable medium and the computer executable code.

[0018] To the accomplishment of the foregoing and related ends, the one or more embodiments include the features hereinafter fully described and particularly pointed out in the claims. The following description and the associated drawings set forth in detail certain illustrative features of the one or more aspects. These features are indicative, however, of but a few of the various ways in which the principles of various aspects may be employed, and this description is intended to include all such aspects and their equivalents.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

[0020] FIG. 1 shows a flowchart of a method for early fake account detection according to various embodiments.

[0021] FIG. 2 shows a schematic diagram of a system configured for early fake account detection according to various embodiments.

[0022] FIG. 3 shows an exemplary diagram of a lifecycle of an account according to various embodiments.

[0023] FIG. 4A shows an exemplary diagram of a positive label according to various embodiments.

[0024] FIG. 4B shows an exemplary diagram of a negative label according to various embodiments.

[0025] FIG. 5 shows an exemplary diagram of a network architecture of a survival analysis ML model according to various embodiments.

[0026] FIG. 6 shows an exemplary boxplot of time taken to detect fake accounts according to various embodiments.

DETAILED DESCRIPTION

[0027] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the invention may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the invention. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0028] Embodiments described in the context of one of the systems or server or methods or computer program are analogously valid for the other systems or server or methods or computer program and vice-versa.

[0029] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments.

Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0030] The word "exemplary" is used herein to mean "serving as an example, instance, or illustration." Any embodiment or design described herein as "exemplary" is not necessarily to be construed as preferred or advantageous over other embodiments or designs

[0031] In the context of various embodiments, the articles "a", "an", and "the" as used with regard to a feature or element include a reference to one or more of the features or elements.

[0032] As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items.

[0033] The terms "at least one" and "one or more" may be understood to include a numerical quantity greater than or equal to one (e.g., one, two, three, four, […], etc.). The term "a plurality" may be understood to include a numerical quantity greater than or equal to two (e.g., two, three, four, five, […], etc.).

[0034] The words "plural" and "multiple" in the description and the claims expressly refer to a quantity greater than one. Accordingly, any phrases explicitly invoking the aforementioned words (e.g. "a plurality of [objects]", "multiple [objects]") referring to a quantity of objects expressly refers more than one of the said objects. The terms "group (of)", "set [of]", "collection (of)", "series (of)", "sequence (of)", "grouping (of)", etc., and the like in the description and in the claims, if any, refer to a quantity equal to or greater than one, i.e. one or more. The terms "proper subset", "reduced subset", and "lesser subset" refer to a subset of a set that is not equal to the set, i.e. a subset of a set that contains less elements than the set.

[0035] The term "data" as used herein may be understood to include information in any suitable analog or digital form, e.g., provided as a file, a portion of a file, a set of files, a signal or stream, a portion of a signal or stream, a set of signals or streams, and the like. Further, the term "data" may also be used to mean a reference to information, e.g., in form of a pointer. The term data, however, is not limited to the aforementioned examples and may take various forms and represent any information as understood in the art.

[0036] The term “processor” or “controller” as, for example, used herein may be understood as any kind of entity that allows handling data, signals, etc. The data, signals, etc. may be handled according to one or more specific functions executed by the processor or controller.

[0037] A processor or a controller may thus be or include an analog circuit, digital circuit, mixed-signal circuit, logic circuit, processor, microprocessor, Central Processing Unit (CPU), Graphics Processing Unit (GPU), Digital Signal Processor (DSP), Field Programmable Gate Array (FPGA), integrated circuit, Application Specific Integrated Circuit (ASIC), etc., or any combination thereof. Any other kind of implementation of the respective functions, which will be described below in further detail, may also be understood as a processor, controller, or logic circuit. It is understood that any two (or more) of the processors, controllers, or logic circuits detailed herein may be realized as a single entity with equivalent functionality or the like, and conversely that any single processor, controller, or logic circuit detailed herein may be realized as two (or more) separate entities with equivalent functionality or the like.

[0038] The term “system” (e.g., a drive system, a position detection system, etc.) detailed herein may be understood as a set of interacting elements, the elements may be, by way of example and not of limitation, one or more mechanical components, one or more electrical components, one or more instructions (e.g., encoded in storage media), one or more controllers, etc.

[0039] A “circuit” as user herein is understood as any kind of logic-implementing entity, which may include special-purpose hardware or a processor executing software. A circuit may thus be an analog circuit, digital circuit, mixed-signal circuit, logic circuit, processor, microprocessor, Central Processing Unit (“CPU”), Graphics Processing Unit (“GPU”), Digital Signal Processor (“DSP”), Field Programmable Gate Array (“FPGA”), integrated circuit, Application Specific Integrated Circuit (“ASIC”), etc., or any combination thereof. Any other kind of implementation of the respective functions which will be described below in further detail may also be understood as a “circuit.” It is understood that any two (or more) of the circuits detailed herein may be realized as a single circuit with substantially equivalent functionality, and conversely that any single circuit detailed herein may be realized as two (or more) separate circuits with substantially equivalent functionality. Additionally, references to a “circuit” may refer to two or more circuits that collectively form a single circuit.

[0040] As used herein, “memory” may be understood as a non-transitory computer-readable medium in which data or information can be stored for retrieval. References to “memory” included herein may thus be understood as referring to volatile or non-volatile memory, including random access memory (“RAM”), read-only memory (“ROM”), flash memory, solid-state storage, magnetic tape, hard disk drive, optical drive, etc., or any combination thereof. Furthermore, it is appreciated that registers, shift registers, processor registers, data buffers, etc., are also embraced herein by the term memory. It is appreciated that a single component referred to as “memory” or “a memory” may be composed of more than one different type of memory, and thus may refer to a collective component including one or more types of memory. It is readily understood that any single memory component may be separated into multiple collectively equivalent memory components, and vice versa. Furthermore, while memory may be depicted as separate from one or more other components (such as in the drawings), it is understood that memory may be integrated within another component, such as on a common integrated chip.

[0041] FIG. 1 shows a flowchart of a method for early fake account detection according to various embodiments.

[0042] In some embodiments, the method 100 may include a step 102 of using one or more processor(s) of a system to collect user registration signals when a user registers for an account. The method 100 may include a step 104 of using the one or more processor(s) to collect booking signals when the user books a ride hailing service using the account.

[0043] In some embodiments, the method 100 may include a step 106 of using the one or more processor(s) to collect transaction signals when the user makes a transaction using the account. The method 100 may include a step 108 of using the one or more processor(s) to collect clickstream data when the user interacts with the account.

[0044] In some embodiments, the method 100 may include a step 110 of using the one or more processor(s) to concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals. The method 100 may include performing the above for each predetermined time step of a sequence of predetermined time steps. The method 100 may include a step 112 of using the one or more processor(s) to determine whether the account is fake using a sequence of vectors of signals comprising the vector of signals for each predetermined time step.

[0045] Steps 102 to 112 are shown in a specific order, however other arrangements are possible. Steps may also be combined in some cases. Any suitable order of steps 102 to 112 may be used.

[0046] According to various embodiment, the one or more processor(s) determine whether an account is fake by supplying the sequence of vectors of signals to a machine learning model trained for that purpose. The method may include training the machine learning model by supervised learning using training data comprising training data elements, wherein each training data element includes a sequence of vectors of signals and a (ground truth) label comprising an indication whether the account is fake and a time step (i.e. time limit) until which it should be determined that the account is fake (if it is fake). The time step may correspond to the index of one of the vectors of signals, i.e. the vectors of signals may by indexed by indices corresponding to the time steps (or indices of the time steps), i.e. the sequence may have a vector of signals for each time step.

[0047] The one or more processor(s) can detect a fake account much earlier in time as compared to a baseline model that is not optimize for detection time.

[0048] FIG. 2 shows a schematic diagram of a system configured for early fake account detection according to various embodiments.

[0049] According to various embodiments, the communication system 200 may include a server 210, and/or a user device 220

[0050] In some embodiments, the server 210 and the user device 220 may be in communication with each other through communication network 230. Even though FIG. 2 shows lines connecting the server 210, and the one or more image acquisition apparatus 220, to the communication network 230, in some embodiments, the server 210, and the user device 220, may not be physically connected to each other, for example through a cable. Instead, the server 210, and the user device 220 may be able to communicate wirelessly through communication network 230 by internet communication protocols or through a mobile cellular communication network.

[0051] In various embodiments, the server 210 may be a single server as illustrated schematically in FIG. 2, or have the functionality performed by the server 210 distributed across multiple server components. The server 210 may include one or more server processor(s) 212. The various functions performed by the server 210 may be carried out by the one or more server processor(s) 212. In some embodiments, the various functions performed by the server 210 may be carried out across the one or more server processor(s).

In other embodiments, each specific function of the various functions performed by the server 210 may be carried out by specific server processor(s) of the one or more server processor(s).

[0052] In some embodiments, the server 210 may include a database 214. The server 210 may also include a memory 216. The database 214 may be in or may be the memory 216. The memory 216 and the database 214 may be one component or may be separate components. The memory 216 of the server may include computer executable code defining the functionality that the server 210 carries out under control of the one or more server processor 212. The memory 216 may include or may be a computer program product such as a non-transitory computer-readable medium.

[0053] In some embodiments, the memory 216 may be part of the one or more server processor(s) 212. In some embodiments, the one or more server processor(s) 212 may also include a neural network processor 215, a decision-making processor 217.

[0054] According to various embodiments, a computer program product may store the computer executable code including instructions for early fake account detection according to the various embodiments. The computer executable code may be a computer program. The computer program product may be a non-transitory computer-readable medium. The computer program product may be in the communication system 100 and/or the server 210.

[0055] In some embodiments, the server 210 may also include an input and/or output module allowing the server 210 to communicate over the communication network 230. The server 210 may also include a user interface for user control of the server 210. The user interface may include, for example, computing peripheral devices such as display monitors, user input devices, for example, touchscreen devices and computer keyboards.

[0056] In various embodiments, the user device 220 may include a user device memory 222 and user device processor 224. The one or more image acquisition apparatus memory 222 may include computer executable code defining the functionality the one or more image acquisition apparatus 220 carries out under control of the user device processor 224. The one or more image acquisition apparatus memory 222 may include or may be a computer program product such as a non-transitory computer-readable medium. The user device 220 may also include an input and/or output module allowing the user device 220 to communicate over the communication network 230. The user device 220 may also include

a user interface for the user to control the user device 220. The user interface may include a display monitor, and/or buttons.

[0057] In various embodiments, the server 210 may be configured for early fake account detection. The server 210 may be configured to: for each predetermined time step of a sequence of predetermined time steps collect user registration signals when a user registers for an account; collect booking signals when the user books a ride hailing service using the account; collect transaction signals when the user makes a transaction using the account; collect clickstream data when the user interacts with the account; concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals; and determine whether the account is fake using a sequence of vectors of signals comprising the vector of signals for each predetermined time step.

[0058] The method and system described above allows early fake account detection and thus makes the overall system more efficient since less processing is required. When a fake account is detected a service for that account may be stopped and/or the account may be deleted etc.

[0059] According to various embodiments, the vector of signals is obtained for each predetermined time step.

[0060] According to various embodiments, the sequence of vectors of signals is used as input to a machine learning model, and the processor is configured to use the machine learning model to determine whether the account is fake, e.g. as early as possible.

[0061] According to various embodiments, the processor is configured to determine whether the account is fake given a sequence of vectors of signals of length T. T indicates the number of steps where user behaviors are observed and various input signals are collected and T is a model hyper-parameter that can be predetermined and tuned. If the current time step is smaller than the predetermined sequence length T, the input sequence will be appended with paddings with default values

[0062] According to various embodiments, the processor is configured to process the sequence of vectors of signals and embed the sequence of vectors of signals into a hidden state to obtain an intermediate output, wherein the intermediate output is a hazard rate, and wherein the hazard rate indicates an instantaneous probability that an account ban is triggered.

[0063] According to various embodiments, the processor is configured to add paddings to the sequence of vectors of signals if a current time step of the sequence of vectors of signals is less than a predetermined time maximum time step T.

[0064] FIG. 3 shows an exemplary diagram 300 of a lifecycle of an account according to various embodiments.

[0065] In various embodiments, an input data may include different signals generated from different checkpoints in an account's life cycle. For example, when the account was registered on the platform, a signal 302 may be obtained on how likely the account is fake. Similarly signals from clickstream data 304 when the user interacts with the application and/or signals from metadata of a ride booking may be used to determine how likely the account is fake. Metadata may include booking signal 306 and/or transaction signal 308.

[0066] In various embodiments, by concatenating those signals together, a vector of signals at each time step, i.e., $[x^1, \ldots, x^k]$ whereby each element represents a type of signal and k represents the total number of signals.

[0067] The user account may be observed for T steps, which may lead to a sequence of signals of length T in the format of $x = [[x_1^1, \ldots, x_1^k], \ldots, [x_T^1, \ldots, x_T^k]]$, which may be used as an input to a machine learning model.

[0068] In an embodiment, using the input to the machine learning model, a ban signal 310 may be obtained if the system determines that the account should be banned.

[0069] FIG. 4A shows an exemplary diagram of a positive label according to various embodiments. FIG. 4B shows an exemplary diagram of a negative label according to various embodiments.

[0070] In an embodiment, labels from the historical data may be generated by checking various account activities starting from account registration, such as booking a ride, and/or making a transaction. As shown in FIG. 4A, if an account is banned due to suspicious activities, a positive label may be collected in the format of (1, t), where t indicates the number of steps from account registration to account ban. As shown in FIG. 4B, if an account survives longer than T steps, a negative label may be collected in the format of (0, T). T may be a hyperparameter that may be tunable in the machine learning model.

[0071] Unlike other supervised machine learning approaches, where the label only consists of 0 or 1. The label disclosed herein may have two components. The first component takes value from {0, 1}, and it indicates whether the account is good or fake. The second component takes value from {x | x = 1, 2, … T}, and it indicates how many

steps before the fake account was detected. With the additional time information, the machine learning model may achieve fake accounts early detection.

[0072] FIG. 5 shows an exemplary diagram of a network architecture of a survival analysis ML model according to various embodiments.

[0073] In an embodiment, the task may be modeled as a survival analysis model 500, i.e., a long-short term memory neural network.

[0074] In an embodiment, the input is a sequence of signals of length T, i.e., $x = [[x_1^1, \ldots, x_1^k], \ldots, [x_T^1, \ldots, x_T^k]]$, where k is the number of signals (booking signal, transaction signal, e.t.c), and T is the input sequence length in survival analysis.

[0075] The sequence of vectors of signals is being processed and embedded into a hidden state h_i . The immediate output is the hazard rate λ_i , which indicates the instantaneous probability that an account ban is triggered. The final survival probability s_T incorporates all the hazard rates before time step T, and it is monotonically decreasing, i.e., $s_T < s_{T-1}$

[0076] In order to get a survival probability score for an unknown account, a sequence of signals of length T may be generated for that account, a forward pass on the neural network may be done. However, T steps may be required to generate the input sequence of signals.

[0077] To achieve early detection, paddings may be added to the input sequence. For example, if current time step m < T, the input sequence will be of length m and paddings of length (T-m) may be added at the end of the sequence. With the paddings, the input sequence has the following format, where $[x_1, \ldots, x_m, \hat{x}_{m+1}, \ldots, \hat{x}_T]$, where $\{\hat{x}_i = [0, \ldots, 0] | i = m+1, \ldots, T\}$ is the paddings being added. The intuition of the paddings is that an account has an instantaneous risk score of 0 if no information has been provided at that time step.

[0078] FIG. 6 shows an exemplary boxplot 600 of time taken to detect fake accounts according to various embodiments. It is significantly less than a baseline model that does not optimize for detection time.

[0079] In an embodiment, the machine learning model has been deployed for early fake accounts detection. The model banned 16K fake accounts per month, with a low appeal rate of 4%. More importantly, it is able to detect fake accounts much earlier than existing machine models.

[0080] The boxplot 600 shows that the model is able to reduce the detection time from 4 days to 1.6 days on average, which significantly reduces the fraud losses because the fraudsters have much less time to conduct fraudulent activities on the application.

[0081] In the following, various examples of this disclosure are illustrated:

[0082] Example 1 is a system for early fake account detection, the system comprising: one or more processor(s); and a memory having instructions stored therein, the instructions, when executed by the one or more processor(s), cause the one or more processor(s) to: for each predetermined time step of a sequence of predetermined time steps, collect user registration signals when a user registers for an account; collect booking signals when the user books a ride hailing service using the account; collect transaction signals when the user makes a transaction using the account; collect clickstream data when the user interacts with the account; and concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals; and determine whether the account is fake using a sequence of vectors of signals comprising the vector of signals for each predetermined time step.

[0083] Example 2 is the system of Example 1, wherein the system is optimized for early detection time.

[0084] Example 3 is the system of Example 1 or 2, wherein the sequence of vectors of signals is obtained by concatenating the vectors of signals for each predetermined time step.

[0085] Example 4 is the system of any one of Examples 1 to 3, wherein the sequence of vectors of signals is used as input to a machine learning model, and the processor is configured to use the machine learning model to determine whether the account is fake.

[0086] Example 5 is the system of Example 4, wherein the machine learning model is trained for both prediction accuracy and detection time.

[0087] Example 6 is the system of any one of Examples 1 to 5, wherein the processor is configured to determine whether the account is fake given a sequence of vectors of signals of a predetermined length T.

[0088] Example 7 is the system of Example 4, wherein the processor is configured to process the sequence of vectors of signals and embed the sequence of vectors of signals into a hidden state to obtain an intermediate output, wherein the intermediate output is a hazard rate, and wherein the hazard rate indicates an instantaneous probability that an account ban is triggered.

[0089] Example 8 is the system of Example 6, wherein the processor is configured to add paddings to the sequence of vectors of signals if a current time step is less than the predetermined number T of predetermined time steps.

[0090] Example 9 is a method for early fake account detection, the method comprising using one or more processor(s) to: for each predetermined time step of a sequence of predetermined time steps, collect user registration signals when a user registers for an account; collect booking signals when the user books a ride hailing service using the account; collect transaction signals when the user makes a transaction using the account; and collect clickstream data when the user interacts with the account; concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals; and determine whether the account is fake using a sequence vectors of signals comprising the vector of signals for each predetermined time step.

[0091] Example 10 is the method of Example 9, wherein the system is optimized for early detection time.

[0092] Example 11 is the method of Example 9 or 10, wherein the vector of signals is obtained for each predetermined time step.

[0093] Example 12 is the method of any one of Examples 9 to 11, wherein the sequence of vectors of signals is used as input to a machine learning model, and the processor is configured to use the machine learning model to determine whether the account is fake.

[0094] Example 13 is the method of Example 12, comprising simultaneously optimizing the machine learning model is for prediction accuracy and detection time.

[0095] Example 14 is the method of Example 13, wherein the processor determines whether the account is fake given a sequence of vectors of signals of a predetermined length T.

[0096] Example 15 is the method of Example 12, wherein the processor processes the sequence of vectors of signals and embeds the sequence of vectors of signals into a hidden state to obtain an intermediate output, wherein the intermediate output is a hazard rate, and wherein the hazard rate indicates an instantaneous probability that an account ban is triggered.

[0097] Example 16 is the method of Example 14, wherein the processor adds paddings to the sequence of vectors of signals if a current time step is less than the predetermined number T of predetermined time steps.

[0098] Example 17 is a non-transitory computer-readable medium storing computer executable code comprising instructions for early fake account detection according to any one of Examples 9 to 16.

[0099] Example 18 is a computer executable code comprising instructions for early fake account detection according to any one of Examples 9 to 16.

[00100] While the invention has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A system for early fake account detection, the system comprising:

one or more processors; and

a memory having instructions stored therein, the instructions, when executed by the one or more processors, cause the one or more processors to:

for each predetermined time step of a sequence of predetermined time steps

collect user registration signals when a user registers for an account;

collect booking signals when the user books a ride hailing service using the account;

collect transaction signals when the user makes a transaction using the account;

collect clickstream data when the user interacts with the account; and

concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals; and

determine whether the account is fake using a sequence of vectors of signals comprising the vector of signals for each predetermined time step,

wherein a label is generated by checking activities of the account for training a machine learning model, and the label comprises a first component indicating whether the account is fake and a second component indicating a number of steps before the account is detected as fake.

2. The system of claim 1, wherein the system is optimized for early detection time.

3. The system of claim 1 or 2, wherein the sequence of vectors of signals is obtained by concatenating the vectors of signals for each predetermined time step.

4. The system of any one of claims 1 to 3, wherein the sequence of vectors of signals is used as input to the machine learning model, and the one or more processors are configured to use the machine learning model to determine whether the account is fake.

5. The system of claim 4, wherein the machine learning model is trained for both prediction accuracy and detection time.

6. The system of any one of claims 1 to 5, wherein the one or more processors are configured to determine whether the account is fake given the sequence of vectors of signals of a predetermined length T.

7. The system of claim 4, wherein the one or more processors are configured to process the sequence of vectors of signals and embed the sequence of vectors of signals into a hidden state to obtain an intermediate output, wherein the intermediate output is a hazard rate, and wherein the hazard rate indicates an instantaneous probability that an account ban is triggered.

8. The system of claim 6, wherein the one or more processors are configured to add paddings to the sequence of vectors of signals if a current time step is less than the predetermined length T of the predetermined time steps.

9. A method for early fake account detection, the method comprising using one or more processors to:

for each predetermined time step of a sequence of predetermined time steps

collect user registration signals when a user registers for an account;

collect booking signals when the user books a ride hailing service using the account;

collect transaction signals when the user makes a transaction using the account; and

collect clickstream data when the user interacts with the account;

concatenate the user registration signals, the booking signals, the transaction signals and the clickstream data to obtain a vector of signals; and

determine whether the account is fake using a sequence vectors of signals comprising the vector of signals for each predetermined time step,

wherein a label is generated by checking activities of the account for training a machine learning model, and the label comprises a first component indicating whether the account is fake and a second component indicating a number of steps before the account is detected as fake.

10. The method of claim 9, wherein the method is optimized for early detection time.

11. The method of claim 9 or 10, wherein the vector of signals is obtained for each predetermined time step.

12. The method of any one of claims 9 to 11, wherein the sequence of vectors of signals is used as an input to the machine learning model, and the method further comprises using the one or more processors to use the machine learning model to determine whether the account is fake.

13. The method of claim 12, comprising simultaneously optimizing the machine learning model for prediction accuracy and detection time.

14. The method of claim 13, wherein the method further comprises using the one or more processors to determine whether the account is fake given the sequence of vectors of signals of a predetermined length T.

15. The method of claim 12, wherein the method further comprises using the one or more processors to process the sequence of vectors of signals and embed the sequence of vectors of signals into a hidden state to obtain an intermediate output, wherein the intermediate output is a hazard rate, and wherein the hazard rate indicates an instantaneous probability that an account ban is triggered.

16. The method of claim 14, wherein the method further comprises using the one or more processors to add paddings to the sequence of vectors of signals if a current time step is less than the predetermined length T of the predetermined time steps.

17. A non-transitory computer-readable medium storing computer executable code comprising instructions for early fake account detection according to any one of claims 9 to 16.

18. A computer executable code comprising instructions for early fake account detection according to any one of claims 9 to 16.