DATA PROCESSING METHOD AND SYSTEM, AND STORAGE MEDIUM

TECHNICAL FIELD

[0001] This specification relates to the field of data processing technologies, and in particular, to a data processing method and system, and a storage medium.

BACKGROUND

[0002] Currently, in many application scenarios, a risk control system needs to be deployed to perform risk prevention and control. A transaction scenario is used as an example. To ensure security of a transaction, the risk control system is usually deployed on a transaction platform, and a risk control policy is integrated into the risk control system. When the transaction platform generates a transaction, the risk control system can identify, based on the risk control policy, whether the transaction is a risk transaction.

[0003] The risk control policy can be a series of risk control rules. The risk control rules are usually obtained when a risk control person obtains a risk mode through summarization based on an occurred risk transaction, and then sorts the summarized risk mode based on experience of the risk control person. The risk control policy can alternatively be a pre-trained risk control model. A training process of the risk control model is, for example, marking a risk transaction and a non-risk transaction from a large quantity of historical transactions, and performing model training based on the historical transaction and marked information of the historical transaction, so that the trained model has a capability of identifying the risk transaction.

[0004] It can be seen from the above-mentioned descriptions that the risk control policy is obtained through summarization/learning from a known risk transaction regardless of whether the risk control policy is a risk control rule or a risk control model. However, risk prevention and control experience obtained through summarization/learning from the known risk transaction has hysteresis, that is, is later than occurrence of the risk transaction. Consequently, the risk control policy only prevents and controls a known risk, and has no prevention and control capability for

an unknown risk. Therefore, there is a possibility that the risk control policy misses prevention and control of an unknown risk.

SUMMARY

[0005] This specification provides a data processing method and system, and a storage medium. A plurality of historical events are used as an initial population, and multi-generation breeding is performed on the initial population by simulating a population evolution process, to obtain a target event with an unknown risk. Therefore, the target event can be used to update a current risk control policy, so that an updated risk control policy has a capability of preventing and controlling the unknown risk, and a problem that risk control experience has hysteresis is resolved.

[0006] According to a first aspect, this specification provides a data processing method, including: obtaining a plurality of historical events, where at least some of the plurality of historical events are events with a known risk, and the known risk is a risk that can be identified based on a current risk control policy; using the plurality of historical events as an initial population, and performing N-generation breeding on the initial population by simulating a population evolution process, to obtain a target population, where in a process of the N-generation breeding, the known risk has evolved into an unknown risk that cannot be identified based on the current risk control policy, and N is an integer greater than or equal to 1; and determining at least one target event with the unknown risk in the target population, where the at least one target event is used for an update process of the current risk control policy.

[0007] According to a second aspect, this specification further provides a data processing system, including: at least one storage medium, storing at least one instruction set, to perform data processing; and at least one processor, communicatively connected to the at least one storage medium. When the data processing system runs, the at least one processor reads the at least one instruction set, and performs the following operations based on an indication of the at least one instruction set: obtaining a plurality of historical events, where at least some of the plurality of historical events are events with a known risk, and the known risk is a risk that can be identified based on a current risk control policy; using the plurality of historical events as an initial population, and performing N-generation breeding on the initial population by simulating a population evolution process, to obtain a target population, where in a process of the N-generation breeding,

the known risk has evolved into an unknown risk that cannot be identified based on the current risk control policy, and N is an integer greater than or equal to 1; and determining at least one target event with the unknown risk in the target population, where the at least one target event is used for an update process of the current risk control policy.

[0008] According to a third aspect, this specification further provides a computer-readable non-volatile storage medium. The computer-readable non-volatile storage medium stores at least one instruction set, and when the at least one instruction set is executed by at least one processor, the data processing method according to the first aspect is implemented.

[0009] Other functions of the data processing method provided in this specification are partially listed in the following descriptions. Creative aspects of the data processing method provided in this specification can be fully explained by practice or using the methods, apparatuses, and combinations described in the following detailed examples.

BRIEF DESCRIPTION OF DRAWINGS

[0010] To describe the technical solutions in the embodiments of this specification more clearly, the following briefly describes the accompanying drawings needed for describing the embodiments. Clearly, the accompanying drawings in the following descriptions show merely some embodiments of this specification, and a person of ordinary skill in the art can still derive other drawings from these accompanying drawings without creative efforts.

[0011] FIG. 1 is a schematic diagram of a risk control scenario according to an embodiment of this specification;

[0012] FIG. 2 is a diagram of a hardware structure of a computing system 200 according to an embodiment of this specification;

[0013] FIG. 3 is a flowchart of a data processing method according to an embodiment of this specification;

[0014] FIG. 4 is a flowchart of simulating a population evolution process according to an embodiment of this specification;

[0015] FIG. 5 is a schematic diagram of a process of ith-generation breeding according to an embodiment of this specification;

[0016] FIG. 6 is a schematic diagram of performing a crossover operation according to an

embodiment of this specification;

[0017] FIG. 7 is a schematic diagram of a mutation operation of adding an attribute according to an embodiment of this specification;

[0018] FIG. 8 is a schematic diagram of a mutation operation of deleting an attribute according to an embodiment of this specification; and

[0019] FIG. 9 is a schematic diagram of a mutation operation of changing an attribute according to an embodiment of this specification.

DESCRIPTION OF EMBODIMENTS

[0020] Specific application scenarios and requirements of this specification are provided in the following descriptions, so that a person skilled in the art can manufacture and use content of this specification. For the person skilled in the art, various local modifications to the disclosed embodiments are clear, and the general principles defined here can be applied to other embodiments and applications without departing from the spirit and scope of this specification. Therefore, this specification is not limited to the shown embodiments, but has a widest scope consistent with that of the claims.

[0021] The terms used here are only intended to describe a particular example embodiment, but do not impose a limitation. For example, unless otherwise specified in the context, the singular forms "one", "a", and "the" used here can also include a plural form. When used in this specification, the terms "include", "comprise", and/or "have" indicate/indicates existence of an associated integer, step, operation, element, and/or component, but does not exclude existence of one or more other features, integers, steps, operations, elements, components, and/or groups, or another feature, integer, step, operation, element, component, and/or group can be added to the system/method.

[0022] In consideration of the following descriptions, these and other features of this specification, operations and functions of related components of the structure, and economy of combination and manufacturing of components can be significantly improved. With reference to the accompanying drawings, all of these form a part of this specification. However, it should be clearly understood that the accompanying drawings are merely used for the purpose of illustration and description, and are not intended to limit the scope of this specification. It should be further

understood that the accompanying drawings are not drawn to scale.

[0023] The flowchart used in this specification illustrates operations implemented by a system in some embodiments of this specification. It should be clearly understood that operations in the flowchart can be implemented out of sequence. On the contrary, operations can be implemented reversely or simultaneously. In addition, one or more other operations can be added to the flowchart. One or more operations can be removed from the flowchart.

[0024] For ease of description, the terms that appear below in this specification are first explained.

[0025] Term 1: Genetic algorithm. The genetic algorithm is an algorithm based on a natural selection and genetic mechanism, and is one type of evolutionary computation. The genetic algorithm searches for an optimal solution by simulating mechanisms such as selection, crossover, and mutation in a biological evolution process. Usually, the genetic algorithm is widely applied to various complex optimization problems such as function optimization, path planning, and machine learning.

[0026] The following describes an application scenario of this specification.

[0027] The technical solutions provided in this specification are applicable to a scenario in which risk prevention and control need to be performed. A transaction scenario is used as an example. To ensure security of a transaction, the risk control system is usually deployed on a transaction platform, and a risk control policy is integrated into the risk control system. When the transaction platform generates a transaction, the risk control system can identify, based on the risk control policy, whether the transaction is a risk transaction.

[0028] The risk control policy is a key factor of whether the risk control system can accurately identify a risk transaction. Currently, the risk control policy is usually obtained through summarization and learning based on a known risk transaction, and can identify an event with a known risk, but does not have a prevention and control capability for an unknown risk.

[0029] This specification provides a data processing method, which can be executed by a data processing system. The data processing system uses a plurality of historical events as an initial population, and perform N-generation breeding on the initial population by simulating a population evolution process, to obtain a target population. At least some events in the plurality of historical events have a known risk (namely, a risk that can be identified based on a current risk control policy). In a process of the N-generation breeding, the known risk has evolved into an

unknown risk (namely, a risk that cannot be identified based on the current risk control policy). Further, the data processing system determines at least one target event with the unknown risk from the target population. The at least one target event can be used for an update process of the current risk control policy. In this way, an updated risk control policy has a prevention and control capability for the unknown risk.

[0030] It should be noted that the above-mentioned transaction scenario is only one of a plurality of use scenarios provided in this specification. The data processing method provided in this specification not only can be applied to the transaction scenario, but also can be applied to all scenarios in which risk prevention and control need to be performed, such as a virus prevention and control scenario or a firewall prevention and control scenario. A person skilled in the art should understand that, when the data processing method provided in this specification are applied to another use scenario, an implementation and a technical effect thereof are similar.

[0031] FIG. 1 is a schematic diagram of a risk control scenario according to an embodiment of this specification. As shown in FIG. 1, the scenario 100 can include a data processing system 11 and a risk control system 12. The scenario 100 involves a data processing phase and a risk control phase.

[0032] In the data processing phase, the data processing system 11 can generate at least one target event with an unknown risk based on a plurality of historical events by simulating a population evolution process. The at least one target event can be used to update a current risk control policy, so that an updated risk control policy has a prevention and control capability for the unknown risk. The updated risk control policy can be deployed in the risk control system 12. In the risk control phase, the risk control system 12 can perform risk prevention and control based on the updated risk control policy.

[0033] The data processing system 11 can be a computing system that has a specific computing capability. The data processing system 11 can include one or more computing devices. The data processing method provided in this specification can be performed by the data processing system 11. The data processing system 11 can store data or instructions for performing the data processing method described in this specification, and can execute or be configured to execute the data or the instructions. In some embodiments, the data processing system 11 can include a hardware device with a data information processing function and a necessary program needed for driving the hardware device to work. The data processing system 11 can use the plurality of historical events

as an initial population, and perform N-generation breeding on the initial population by simulating a population evolution process, to obtain a target population. At least some events in the plurality of historical events have a known risk (namely, a risk that can be identified based on a current risk control policy). In a process of the N-generation breeding, the known risk has evolved into an unknown risk (namely, a risk that cannot be identified based on the current risk control policy). Further, the data processing system 11 can determine, from the target population, at least one target event that may have the unknown risk.

[0034] The risk control system 12 can be a computing system that has a specific computing capability. As shown in FIG. 1, an updated risk control policy can be configured in the risk control system 12, and risk prevention and control are performed based on the updated risk control policy.

[0035] In some embodiments, a process of updating the current risk control policy based on the at least one target event can be performed by the data processing system 11. For example, after obtaining the at least one target event, the data processing system 11 can update the current risk control policy based on the at least one target event, and provide the updated risk control policy for the risk control system. In some embodiments, a process of updating the current risk control policy based on the at least one target event can be performed by the risk control system 12. For example, after obtaining the at least one target event, the data processing system 11 can pass the at least one target event to the risk control system 12. The risk control system 12 can update the current risk control policy in the risk control system 12 based on the at least one target event.

[0036] The data processing system 11 and the risk control system 12 can correspond to the same entity system, or can correspond to different entity systems. When the data processing system 11 and the risk control system 12 correspond to different entity systems, the risk control system 12 can be communicatively connected to the data processing system 11 via a network. When the data processing system 11 and the risk control system 12 correspond to the same entity system, the risk control system 12 can be communicatively connected to the data processing system 11 through a bus in the entity system.

[0037] It should be understood that a quantity of data processing systems 11 and a quantity of risk control systems 12 in FIG. 1 are merely examples. According to an implementation requirement, for example, a target event obtained by one data processing system 11 can be used to update risk control policies in a plurality of risk control systems 12. In this case, there can be any quantity of risk control systems 12. For example, when the data processing method is performed

by using a distributed device group, the data processing method can be executed in a distributed manner in a plurality of data processing systems 11. In this case, there can be any quantity of data processing systems 11.

[0038] FIG. 2 is a diagram of a hardware structure of a computing system 200 according to an embodiment of this specification. The computing system 200 can be used as the data processing system 11 in FIG. 1, to perform the data processing method described in this specification.

[0039] As shown in FIG. 2, the computing system 200 can include at least one storage medium 230 and at least one processor 220. In some embodiments, the computing system 200 can further include a communication port 250 and an internal communication bus 210. The computing system 200 can further include an I/O component 260.

[0040] The internal communication bus 210 can be connected to different system components. For example, the internal communication bus 210 can be connected to the storage medium 230, the processor 220, the communication port 250, and the I/O component 260.

[0041] The I/O component 260 supports an input/output between the computing system 200 and another component.

[0042] The communication port 250 is used for data communication between the computing system 200 and the outside. For example, the communication port 250 can be used for data communication between the computing system 200 and a network. The communication port 250 can be a wired communication port, or can be a wireless communication port.

[0043] The storage medium 230 can include a data storage apparatus. The data storage apparatus can be a non-transitory storage medium, or can be a transitory storage medium. For example, the data storage apparatus can include one or more of a disk 232, a read-only storage medium (ROM) 234, or a random access storage medium (RAM) 236. The storage medium 230 further includes at least one instruction set stored in the data storage apparatus. The instruction set can include computer program code. The computer program code can include a program, a routine, an object, a component, a data structure, a process, a module, etc.

[0044] The at least one processor 220 can be communicatively connected to the at least one storage medium 230. When the computing system 200 runs, the at least one processor 220 reads the at least one instruction set, and performs, based on an indication of the at least one instruction set, the data processing method provided in this specification. The processor 220 can perform steps included in the data processing method. The processor 220 can be in a form of one or more

processors. In some embodiments, the processor 220 can include one or more hardware processors, for example, a microcontroller, a microprocessor, a reduced instruction set computer (RISC), an application-specific integrated circuit (ASIC), an application-specific instruction set processor (ASIP), a central processing unit (CPU), a graphics processing unit (GPU), a physical processing unit (PPU), a microcontroller unit, a digital signal processor (DSP), a field programmable gate array (FPGA), an advanced RISC machine (ARM), a programmable logic device (PLD), any circuit or processor that can perform one or more functions, or any combination thereof.

[0045] To merely describe a problem, only one processor 220 is shown in the computing system 200 in the accompanying drawings. However, it should be noted that, the computing system 200 in this specification can further include a plurality of processors. Therefore, the operations and/or method steps disclosed in this specification can be performed by one processor, or can be jointly performed by a plurality of processors. For example, if the processor 220 in the computing system 200 described in this specification performs step A and step B, it should be understood as that step A and step B can be jointly or separately performed by two different processors 220 (for example, a first processor performs step A, and a second processor performs step B, or a first processor and a second processor jointly perform step A and step B).

[0046] FIG. 3 is a flowchart of a data processing method according to an embodiment of this specification. As described above, a data processing system 11 can perform the data processing method in this specification.

[0047] As shown in FIG. 3, the data processing method can include the following steps.

[0048] S310: Obtain a plurality of historical events, where at least some of the plurality of historical events are events with a known risk, and the known risk is a risk that can be identified based on a current risk control policy.

[0049] The historical event can be an event that occurs at a historical moment. Event information of a historical event can include a plurality of attribute-attribute value key-value pairs. It should be noted that, in different application scenarios, a reference of the historical event and an attribute in the historical event may be different. For example, in a transaction scenario, the historical event can be a transaction event that occurs at the historical moment. An attribute of a transaction event can include an account number, device information, environment information, transaction process information, etc. of two transaction participants. For ease of understanding, when an example is used in the following, the transaction scenario is used as an example for

description.

[0050] A risk control system can be deployed with a risk control policy, and perform risk control on an event based on the current risk control policy. For example, event information of an event (a plurality of attribute-attribute value key-value pairs) is input to the risk control system, and the risk control system identifies, based on the event information, whether the event has a risk, or identifies a specific risk of the event, etc.

[0051] In this specification, the current risk control policy is a risk control policy currently deployed in the risk control system. The current risk control policy is usually obtained through summarization/learning based on a known risk event. For a manner of obtaining the current risk control policy, references can be made to the above-mentioned related descriptions. Details are omitted here for simplicity The current risk control policy can include one or more of a risk control rule or a risk control model. In addition, in this specification, the risk that can be identified based on the risk control policy is referred to as a known risk, and the risk that cannot be identified based on the risk control policy is referred to as an unknown risk.

[0052] In S310, at least some events in the plurality of historical events obtained by the data processing system are events (referred to as black samples) with the known risk, and the remaining events are events without the known risk. For example, the plurality of historical events can include black samples that account for 1% to 10%.

[0053] S320: Use the plurality of historical events as an initial population, and perform N-generation breeding on the initial population by simulating a population evolution process, to obtain a target population, where in a process of the N-generation breeding, the known risk has evolved into an unknown risk that cannot be identified based on the current risk control policy, and N is an integer greater than or equal to 1.

[0054] In some embodiments, the data processing system can simulate the population evolution process based on a genetic algorithm or another similar algorithm. Subsequently, the genetic algorithm is used as an example for description.

[0055] FIG. 4 is a flowchart of simulating a population evolution process according to an embodiment of this specification. As shown in FIG. 4, in the genetic algorithm, the data processing system can use the plurality of historical events as the initial population, perform N-generation breeding on the initial population, and simulate the population evolution process, to finally obtain the target population. For example, the initial population is used as the 1st-generation population,

and the data processing system generates the 2nd-generation population based on the 1st-generation population by simulating the population evolution process. Further, the data processing system generates the 3rd-generation population based on the 2nd-generation population by simulating the population evolution process. By analogy, the data processing system generates the Nth-generation population based on the $(N-1)^{th}$-generation population by simulating the population evolution process. The Nth-generation population is the target population. Because the initial population includes the event with the known risk, in a process of simulating the population evolution process, the known risk has evolved into the unknown risk, so that the target population includes the event with the unknown risk.

[0056] The initial population includes the event with the known risk. Through N-generation breeding shown in FIG. 4, the data processing system can evolve the known risk into the unknown risk in a breeding process, so that the obtained target population includes a target event with the unknown risk. These target events with the unknown risk can be used to update the current risk control policy, so that the updated risk control policy can implement identification of the unknown risk, to reduce a probability of missing prevention and control of the unknown risk, and improve a risk control effect.

[0057] A person skilled in the art can understand that an implementation of the above-mentioned N-generation breeding is similar. The following specifically provides specific descriptions by using a process of the ith-generation breeding as an example.

[0058] A nature population evolution process usually involves steps of genetic inheritance, genetic mutation, and natural selection. To simulate the natural population evolution process, in some embodiments, the data processing system can perform operations such as crossover, mutation, and eviction when simulating the population evolution process. The crossover operation is an operation of generating an evolutionary event based on two different events in a current population. The crossover operation can simulate genetic inheritance in the population evolution process. The current population can be the initial population, or can be a specific-generation population in a breeding process. An evolutionary event of a next-generation population can be obtained through the crossover operation based on breeding of the current population. The mutation operation is randomly performing the mutation operation on an attribute of the evolutionary event, to simulate genetic mutation in the population evolution process. The eviction operation is evicting an evolutionary event that is in a population and that does not satisfy a preset

risk evolution condition, to simulate natural selection in the population evolution process.

[0059] FIG. 5 is a schematic diagram of a process of ith-generation breeding according to an embodiment of this specification. As shown in FIG. 5, the process of ith-generation breeding can include the following steps.

[0060] (1) Perform the crossover operation on different events in the current population, to generate a plurality of evolutionary events.

[0061] When $i = 1$, the current population is the initial population. When $1 < i \leq N$, the current population is a population obtained after $(i-1)^{th}$-generation breeding.

[0062] One event corresponds to a plurality of attributes. In different application scenarios, the event may correspond to different attributes. For example, the transaction scenario is used as an example. The plurality of attributes corresponding to one event can include one or more of the following: an account number, device information, environment information, and transaction process information of two transaction participants.

[0063] In some embodiments, the plurality of attributes corresponding to one event can further include a risk attribute. An attribute value corresponding to the risk attribute represents whether the event has a known risk.

[0064] In some embodiments, the attribute value corresponding to the risk attribute can be used as a risk label of the event. For example, in the plurality of historical events included in the initial population, if some events are events with the known risk, an attribute value corresponding to a risk attribute in the some events is 1 (indicating that there is the known risk), and an attribute value corresponding to a risk attribute in a remaining event is 0 (indicating that there is no known risk).

[0065] Alternatively, a specific attribute that is in the plurality of attributes and that can be used to represent an attribute that has a known risk can be used as a risk attribute. For example, when a specific transaction participant is in a blacklist, and a transaction in which the transaction participant participates is a transaction with a known risk, an account of the transaction participant can be used as the risk attribute.

[0066] The evolutionary event is obtained by performing the crossover operation on different events in the current population. Specifically, an attribute value of some attributes of an evolutionary event comes from a first event, and an attribute value of the other attributes comes from a second event. The first event and the second event are any two different events in the current

population.

[0067] In some embodiments, that the data processing system performs the crossover operation on the current population can include: The data processing system performs the following operations on the current population for M times: The data processing system selects the first event and the second event from the current population. The first event and the second event are different events in the current population. The data processing system obtains at least one evolutionary event based on the first event and the second event. An attribute value of some attributes of the evolutionary event is from the first event, and an attribute value of the other attributes of the evolutionary event is from the second event. M is an integer greater than or equal to 1.

[0068] In some embodiments, when the data processing system obtains each evolutionary event based on the first event and the second event, references can be made to the manner shown in FIG. 6. FIG. 6 is a schematic diagram of performing a crossover operation according to an embodiment of this specification.

[0069] As shown in FIG. 6, the first event includes a plurality of attributes, an attribute of the second event is the same as that of the first event, and evolutionary events generated based on the first event and the second event also have the same attribute. The data processing system can randomly obtain an attribute value of at least one attribute from the first event, obtain an attribute value of the remaining attribute from the second event, and generate an evolutionary event based on the obtained attribute value.

[0070] As shown in FIG. 6, the attributes of the first event include a risk attribute, an attribute 1, an attribute 2, an attribute 3, an attribute 4, and an attribute 5. Each attribute includes a corresponding attribute value. As shown in FIG. 5, an attribute value of the risk attribute is "attribute value A", an attribute value of the attribute 1 is "attribute 1a", an attribute value of the attribute 2 is "attribute 2a", an attribute value of the attribute 3 is "attribute 3a", an attribute value of the attribute 4 is "attribute 4a", and an attribute value of the attribute 5 is "attribute 5a".

[0071] As shown in FIG. 6, attributes of the second event are the same as those of the first event, but include different attribute values. For example, an attribute value of the risk attribute is "attribute value B", an attribute value of the attribute 1 is "attribute 1b", an attribute value of the attribute 2 is "attribute 2b", an attribute value of the attribute 3 is "attribute 3b", an attribute value of the attribute 4 is "attribute 4b", and an attribute value of the attribute 5 is "attribute 5b".

[0072] As shown in FIG. 6, in an example, the data processing system can randomly obtain the attribute value "attribute A" of the risk attribute of the first event, the attribute value "attribute 1a" of the attribute 1, and the attribute value "attribute 4a" of the attribute 4; and obtain the attribute value "attribute 2b" of the attribute 2, the attribute value "attribute 3b" of the attribute 3, and the attribute value "attribute 5b" of the attribute 5 from the second event. The data processing system can generate the evolutionary event based on the above obtained attribute values. To be specific, the attribute value of the risk attribute of the evolutionary event is "attribute A", the attribute value of the attribute 1 is "attribute 1a", the attribute value of the attribute 2 is "attribute 2b", the attribute value of the attribute 3 is "attribute 3b", the attribute value of the attribute 4 is "attribute 4a", and the attribute value of the attribute 5 is "attribute 5b".

[0073] In the above-mentioned solution, the crossover operation is performed on different events in the current population, so that a new evolutionary event can be generated. In a process of generating the evolutionary event, a risk in an event in the current population may evolve, helping evolve the known risk into the unknown risk.

[0074] It can be learned from the crossover operation shown in FIG. 6 that, it can be considered that the evolutionary event is obtained by replacing the attribute value of some attributes in the first event with the attribute value of the corresponding attribute in the second event. For example, with reference to FIG. 6, it can be considered that the evolutionary event is obtained by replacing attribute values of the attribute 2, the attribute 3, and the attribute 5 in the first event with attribute values of the attribute 2, the attribute 3, and the attribute 5 in the second event.

[0075] Alternatively, it can be considered that the evolutionary event is obtained by replacing the attribute value of some attributes in the second event with the attribute value of the corresponding attribute in the first event. For example, with reference to FIG. 6, it can be considered that the evolutionary event is obtained by replacing attribute values of the risk attribute, the attribute 1, and the attribute 4 in the second event with attribute values of the risk attribute, the attribute 1, and the attribute 4 in the first event.

[0076] In some embodiments, to improve evolution efficiency of the known risk, the data processing system can generate a plurality of evolutionary events by replacing attribute values of some attributes in the event with the known risk with attribute values of corresponding attributes in another event. To be specific, when replacing attribute values in different events, the data processing system can focus on using the event with the known risk as a replaced event (namely,

the first event shown in FIG. 6). Replaced some attributes include at least one attribute other than the risk attribute. For example, two events in FIG. 6 are still used as an example. In a process of generating the evolutionary event, the attribute value of the risk attribute in the first event is not replaced, and the attribute values of the attribute 2, the attribute 3, and the attribute 5 are replaced with the attribute values of the attribute 2, the attribute 3, and the attribute 5 in the second event.

[0077] In some embodiments, in the above-mentioned replacement process, a ratio of a quantity of replaced attributes in the first event to a quantity of all attributes in the first event is less than or equal to a preset threshold. For example, if the quantity of all the attributes in the first event is 10, the quantity of replaced attributes can be less than or equal to 5. In other words, when an attribute value replacement is performed on the first event with the known risk, a relatively small quantity of attribute values can be replaced, and a relatively large quantity of original attribute values are retained. Such an evolution process is more consistent with a natural evolution law. Therefore, the attribute value of the risk attribute in the first event can be accurately inherited into the evolutionary event.

[0078] In the above-mentioned solution, evolutionary efficiency of the known risk can be improved by replacing some attribute values in the first event with the known risk with an attribute value of a corresponding attribute in the second event.

[0079] (2) Evict some events from the plurality of evolutionary events, where the some events do not satisfy a preset risk evolution condition.

[0080] In some embodiments, the preset risk evolution condition is used to select the known risk and the unknown risk. For example, an event that satisfies the preset risk evolution condition is an event that may have an unknown risk, and an event that does not satisfy the preset risk evolution condition is an event that has a known risk.

[0081] In some embodiments, the risk evolution condition includes: the evolutionary event is identified as a risk-free event based on the current risk control policy.

[0082] In an example, when the current risk control policy is a risk control model, the data processing system can input event information of the evolutionary event into the risk control model, and receive an identification result fed back by the risk control model. If the risk control model identifies the evolutionary event as a risk event, it indicates that a risk in the evolutionary event can be identified based on the current risk control policy, and the evolutionary event is an event with a known risk. If the risk control model identifies the evolutionary event as a risk-free event,

it indicates that the evolutionary event is risk-free or has a risk that cannot be identified, that is, the evolutionary event may be an event with an unknown risk. The event with a known risk is evicted, to increase a proportion of events with an unknown risk in the remaining events, and further improve risk evolution efficiency.

[0083] In some embodiments, because too many evolutionary events may be generated based on the current population, in a case of continuous breeding, a final quantity of target populations may be too large. Therefore, the data processing system can limit a quantity of eviction events, so that the evolutionary event in the final target population is more likely to be an event with an unknown risk.

[0084] In some embodiments, a quantity of events evicted from the plurality of evolutionary events may satisfy the following requirement: an absolute value of a difference between a quantity of events remaining after eviction and a quantity of events in the initial population is less than a preset threshold. The preset threshold can be an integer greater than or equal to 0. For example, when the preset threshold is 0, the quantity of events remaining after eviction is equal to a quantity of events in the initial population. When the preset threshold is 100, the quantity of events remaining after eviction can be 100 more than the quantity of events in the initial population, or 100 less than the quantity of events in the initial population.

[0085] In some embodiments, before performing eviction, the data processing system can perform a mutation operation on at least some evolutionary events in the plurality of evolutionary events, to obtain an evolutionary event set including a plurality of mutated events. Then, the data processing system evicts some events from the evolutionary event set.

[0086] In some embodiments, a mutation operation corresponding to an evolutionary event includes at least one of the following: randomly adding an attribute value of at least one attribute of the evolutionary event; randomly deleting an attribute value of at least one attribute other than the risk attribute in the evolutionary event; and randomly changing an attribute value of at least one attribute other than the risk attribute in the evolutionary event.

[0087] FIG. 7 is a schematic diagram of a mutation operation of adding an attribute according to an embodiment of this specification.

[0088] In some embodiments, as shown in FIG. 7, attributes of the evolutionary event include a risk attribute, an attribute 1, an attribute 2, an attribute 3, an attribute 4, and an attribute 5. When the data processing system performs a mutation operation, the data processing system can

randomly add at least one attribute to the evolutionary event, and an attribute value of the attribute can be randomly generated, or can be obtained from another evolutionary event.

[0089] In an example, the data processing system randomly adds two attributes to the evolutionary event: an attribute 6 (a corresponding attribute value is an attribute value 6c) and an attribute 7 (a corresponding attribute value is an attribute value 7c). After two attributes are added, the original evolutionary event is mutated into a mutated event. The mutated event has a risk attribute, an attribute 1, an attribute 2, an attribute 3, an attribute 4, an attribute 5, an attribute 6, and an attribute 7.

[0090] FIG. 8 is a schematic diagram of a mutation operation of deleting an attribute according to an embodiment of this specification.

[0091] In some embodiments, as shown in FIG. 8, attributes of the evolutionary event include a risk attribute, an attribute 1, an attribute 2, an attribute 3, an attribute 4, and an attribute 5. When the data processing system performs the mutation operation, the data processing system can randomly delete at least one attribute other than the risk attribute for the evolutionary event.

[0092] In an example, the data processing system randomly deletes an attribute, for example, the attribute 5. After an attribute is deleted, the original evolutionary event is mutated into a mutated event, and the mutated event has a risk attribute, an attribute 1, an attribute 2, an attribute 3, and an attribute 4.

[0093] FIG. 9 is a schematic diagram of a mutation operation of changing an attribute according to an embodiment of this specification.

[0094] In some embodiments, as shown in FIG. 9, attributes of the evolutionary event include a risk attribute, an attribute 1, an attribute 2, an attribute 3, an attribute 4, and an attribute 5. When the data processing system performs a mutation operation, an attribute value of at least one attribute other than a risk attribute can be randomly changed for the evolutionary event.

[0095] In an example, the data processing system randomly deletes attribute values of two attributes. For example, as shown in FIG. 9, an attribute value of the attribute 1 can be changed from "attribute value 1b" to "attribute value 1c", and an attribute value of the attribute 4 can be changed from "attribute value 4b" to "attribute value 4c". After the attribute values of the two attributes are changed, the original evolutionary event is changed to a mutated event.

[0096] It should be noted that, for an evolutionary event, the data processing system can perform only one mutation operation in addition, deletion, or a change, or can perform any two of

the mutation operations, or perform all the three mutation operations once.

[0097] The mutation operation is performed on the evolutionary event, to improve a probability that the evolutionary event evolves into an unknown risk, and further improve efficiency of generating the unknown risk.

[0098] (3) Use a remaining event in the plurality of evolutionary events as a population obtained after i^{th}-generation breeding.

[0099] (1) to (3) in the above-mentioned descriptions describe a process of the ith-generation breeding. The data processing system performs 1st-generation breeding, 2nd-generation breeding, 3rd-generation breeding, ..., and N^{th}-generation breeding based on the breeding process. Therefore, a population obtained after N^{th}-generation breeding is the target population.

[0100] Through N-generation breeding, the data processing system can evolve the known risk in the initial population into the unknown risk in the breeding process, so that the obtained target population may include an event with an unknown risk. The event with an unknown risk is used for an update process of the current risk control policy, to identify the unknown risk, reduce a probability of missing prevention and control of the unknown risk, and improve a risk control effect.

[0101] S330: Determine at least one target event with the unknown risk in the target population, where the at least one target event is used for the update process of the current risk control policy.

[0102] In some embodiments, in the process of the N-generation breeding, attribute values of the risk attributes respectively corresponding to the plurality of historical events are inherited to an event in the target population generation by generation. The attribute value of the risk attribute is used to indicate whether the event has a known risk (or whether the event is evolved from the event with a known risk). The data processing system can select, based on the attribute value of the risk attribute, at least one target event that may have an unknown risk in the target population.

[0103] In some embodiments, the determining at least one target event with the unknown risk in the target population includes: for each event in the target population, if an attribute value of the risk attribute corresponding to the event represents that there is the known risk, determining the event as the target event.

[0104] In some embodiments, the attribute value of the risk attribute can specifically represent whether an account that initiates the historical event is a blacklist account. In an example, the attribute value of the risk attribute can include "belonging to a blacklist account" or "does not

belong to a blacklist account". In the target population, when an attribute value of an account attribute of an event is "belonging to a blacklist account", it is considered that the event may have an unknown risk. In this case, the data processing system can determine that an event that is in the target population and whose attribute value of the risk attribute is "belonging to a blacklist account" is the target event.

[0105] In the above-mentioned solution, the evolutionary event is generated by replacing the attribute value of some attributes in the event (the attribute value of the risk attribute represents that there is a known risk) with a known risk with an attribute value of a corresponding attribute in another event. In this way, evolutionary efficiency of the known risk can be improved. In addition, the risk attribute in the event with a known risk can be inherited to the target population generation by generation. Further, the data processing system can determine the target event in the target population based on the attribute value of the risk attribute, so that the target event with an unknown risk can be conveniently and quickly located in the target population.

[0106] In some embodiments, the data processing system can select the target event with an unknown risk from the target population in a manner such as a manual manner, a large model manner, or a machine learning manner.

[0107] The target event can be used for the update process of the current risk control policy, so that the updated risk control policy has a capability of identifying the unknown risk. For example, when the current risk control policy includes the risk control model, the data processing system or the risk control system can train the risk control model based on the at least one target event, to obtain a trained risk control model. The trained risk control model has a capability of identifying the unknown risk. For example, when the current risk control policy includes a risk control rule, the data processing system, the risk control system, or the risk control person can perform risk pattern analysis based on the at least one target event, to obtain, through summarization, a risk control rule that can identify an unknown risk.

[0108] In conclusion, in the data processing method and system provided in this specification, the historical event is used as the initial population, and the target population is obtained by performing multi-generation breeding on the initial population by simulating the population evolution process. In the breeding process, a known risk in a plurality of historical events is evolved into an unknown risk, and at least one target event that has an unknown risk is determined in the target population and used for the update process of the current risk control policy. The

updated current risk control policy has a capability of identifying the unknown risk. In this way, a problem that the risk control experience is later than the risk event and prevention of the unknown risk is missed can be resolved.

[0109] Another aspect of this specification provides a computer-readable non-transitory storage medium, storing at least one instruction set, to perform data processing. When the at least one instruction set is executed by a processor, the at least one instruction set instructs the processor to implement the steps of the data processing method in this specification. In some possible implementations, aspects of this specification can further be implemented in a form of a program product, including program code. When the program product runs on a computing system 200, the program code is used to enable the computing system 200 to perform the steps of the data processing method described in this specification. The program product used to implement the above-mentioned method can be a portable compact disk read-only memory (CD-ROM) including program code, and can run on the computing system 200. However, the program product in this specification is not limited thereto. In this specification, a readable storage medium can be any tangible medium that includes or stores a program, and the program can be used by or in combination with an instruction execution system. The program product can be any combination of one or more readable media. The readable medium can be a readable signal medium or a readable storage medium. The readable storage medium can be, for example, an electrical system, apparatus, or device, a magnetic system, apparatus, or device, an optical system, apparatus, or device, an electromagnetic system, apparatus, or device, an infrared system, apparatus, or device, or a semiconductor system, apparatus, or device, or any combination thereof. More specific examples of the readable storage medium include: an electrical connection including one or more conducting wires, a portable disk, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage component, a magnetic storage component, or any proper combination thereof. The computer-readable storage medium can include a data signal propagated in a baseband or as part of a carrier, and carries readable program code. The propagated data signal can be in a plurality of forms, and includes but is not limited to an electromagnetic signal, an optical signal, or any proper combination thereof. The readable storage medium can further be any readable medium other than the readable storage medium, and the readable medium can be used to send, propagate, or transmit a program used by

or in combination with an instruction execution system, apparatus, or component. The program code included in the readable storage medium can be transmitted in any proper medium, including but not limited to a wireless medium, a wired medium, optical cable, an RF medium, or any proper combination thereof. Program code for performing an operation of this specification can be written in any combination of one or more program design languages. The program design language includes an object-oriented program design language, for example, Java or C++, and further includes a conventional procedural program design language, for example, "C" language, or a similar program design language. The program code can be completely executed on the computing system 200, partially executed on the computing system 200, executed as an independent software package, partially executed on the computing system 200 and partially executed on a remote computing device, or completely executed on a remote computing device.

[0110] Specific embodiments of this specification are described above. Other embodiments fall within the scope of the appended claims. In some cases, the actions or steps described in the claims can be performed in an order different from that in the embodiments, and the desired results can still be achieved. In addition, the process depicted in the accompanying drawings does not necessarily need a particular order or consecutive order to achieve the desired results. In some implementations, multitasking and parallel processing are possible or can be advantageous.

[0111] In conclusion, after reading the detailed disclosure, a person skilled in the art can understand that the detailed disclosure can be presented only in an example manner, and can impose no limitation. Although it is not explicitly stated here, a person skilled in the art can understand that the requirements of this specification include various proper changes, improvements, and modifications to the embodiments. These changes, improvements, and modifications are intended to be presented in this specification, and fall within the spirit and scope of the example embodiments of this specification.

[0112] In addition, some terms in this specification are used to describe the embodiments of this specification. For example, "one embodiment", "an embodiment", and/or "some embodiments" mean that particular features, structures, or characteristics described with reference to the embodiments can be included in at least one embodiment of this specification. Therefore, it can be emphasized and understood that two or more of "an embodiment", "one embodiment", or "alternative embodiments" referred in parts of this specification do not necessarily indicate the same embodiment. In addition, particular features, structures, or characteristics can be properly

combined in one or more embodiments of this specification.

[0113] It should be understood that in the above-mentioned descriptions of the embodiments of this specification, to help understand a feature, for the purpose of simplifying this specification, various features are combined in a single embodiment, accompanying drawing, or description thereof in this specification. However, this does not mean that the combination of these features is necessary. It is entirely possible for a person skilled in the art to mark some devices and understand the devices as separate embodiments when reading this specification. In other words, the embodiments in this specification can also be understood as an integration of a plurality of secondary embodiments. Content of each secondary embodiment is also true when features are less than all features in one of the above-mentioned disclosed embodiments.

[0114] Each patent, patent application, publication of a patent application, and other materials, for example, articles, books, instructions, publications, documents, and products, that are cited in this specification and that are other than any document that is inconsistent or conflict with this document or that imposes limiting impact on the widest scope of the claims can be incorporated here by reference and used for all purposes now or later associated with this document. In addition, the term in this document shall prevail when there is any inconsistency or conflict between the description, definition and/or use of a related term in any material and the description, definition and/or use of a related term in this document.

[0115] Finally, it should be understood that the principles of the implementation solutions of this specification are described in the implementation solutions of the application disclosed in this specification. Other modified embodiments also fall within the scope of this specification. Therefore, the embodiments disclosed in this specification are merely examples and not limitations. A person skilled in the art can use an alternative configuration according to the embodiments of this specification to implement the application in this specification. Therefore, the embodiments of this specification are not limited to the embodiments accurately described in the application.

CLAIMS

What is claimed is:

1. A data processing method, comprising:

obtaining a plurality of historical events, wherein at least some of the plurality of historical events are events with a known risk, and the known risk is a risk that can be identified based on a current risk control policy;

using the plurality of historical events as an initial population, and performing N-generation breeding on the initial population by simulating a population evolution process, to obtain a target population, wherein in a process of the N-generation breeding, the known risk has evolved into an unknown risk that cannot be identified based on the current risk control policy, and N is an integer greater than or equal to 1; and

determining at least one target event with the unknown risk in the target population, wherein the at least one target event is used for an update process of the current risk control policy.

2. The method according to claim 1, wherein a process of an ith-generation breeding comprises:

performing a crossover operation on different events in a current population, to generate a plurality of evolutionary events;

evicting some events from the plurality of evolutionary events, wherein the some events do not satisfy a preset risk evolution condition; and

using a remaining event in the plurality of evolutionary events as a population obtained after i^{th}-generation breeding, wherein

when $i = 1$, the current population is the initial population; when $1 < i \leq N$, the current population is a population obtained after the $(i-1)^{\text{th}}$-generation breeding; and the target population is a population obtained after N^{th}-generation breeding.

3. The method according to claim 2, wherein one event corresponds to a plurality of attributes, an attribute value of some attributes of the evolutionary event is from a first event, an attribute value of the other attributes of the evolutionary event is from a second event, and the first event and the second event are different events in the current population.

4. The method according to claim 3, wherein the plurality of attributes comprises at least a risk attribute, and an attribute value of the risk attribute represents whether there is a known risk;

and

the evolutionary event is obtained by replacing the attribute value of the some attributes in the first event with an attribute value of a corresponding attribute in the second event, the attribute value of the risk attribute in the first event represents that there is a known risk, and the some attributes comprise at least one attribute other than the risk attribute in the plurality of attributes.

5. The method according to claim 2, wherein the risk evolution condition comprises: the evolutionary event is identified as a risk-free event based on the current risk control policy.

6. The method according to claim 2, wherein a quantity of events evicted from the plurality of evolutionary events satisfies the following requirement: an absolute value of a difference between a quantity of events remaining after eviction and a quantity of events in the initial population is less than a preset threshold.

7. The method according to claim 2, wherein the evicting some events from the plurality of evolutionary events comprises:

performing a mutation operation on at least some evolutionary events in the plurality of evolutionary events, to obtain an evolutionary event set comprising a plurality of mutated events; and

evicting the some events from the evolutionary event set.

8. The method according to claim 7, wherein one event corresponds to a plurality of attributes, and the plurality of attributes comprise at least a risk attribute; and

the mutation operation corresponding to one evolutionary event comprises at least one of the following:

randomly adding an attribute value of at least one attribute of the evolutionary event;

randomly deleting an attribute value of at least one attribute other than the risk attribute in the evolutionary event; and

randomly changing an attribute value of at least one attribute other than the risk attribute in the evolutionary event.

9. The method according to claim 1, wherein each historical event corresponds to a risk attribute, and an attribute value of the risk attribute represents whether the historical event has the known risk;

in the process of the N-generation breeding, attribute values of the risk attributes respectively corresponding to the plurality of historical events are inherited to an event in the target population

generation by generation; and

the determining at least one target event with the unknown risk in the target population comprises: for each event in the target population, if an attribute value of the risk attribute corresponding to the event represents that there is the known risk, determining the event as the target event.

10. The method according to claim 1, wherein the current risk control policy comprises a risk control model, and the method further comprises:

training the risk control model based on the at least one target event, to obtain a trained risk control model, wherein the trained risk control model has a capability of identifying the unknown risk.

11. A data processing system, comprising:

at least one storage medium, storing at least one instruction set, to perform data processing; and

at least one processor, communicatively connected to the at least one storage medium, wherein when the data processing system runs, the at least one processor reads the at least one instruction set, and performs the following operations based on an indication of the at least one instruction set:

obtaining a plurality of historical events, wherein at least some of the plurality of historical events are events with a known risk, and the known risk is a risk that can be identified based on a current risk control policy;

using the plurality of historical events as an initial population, and performing N-generation breeding on the initial population by simulating a population evolution process, to obtain a target population, wherein in a process of the N-generation breeding, the known risk has evolved into an unknown risk that cannot be identified based on the current risk control policy, and N is an integer greater than or equal to 1; and

determining at least one target event with the unknown risk in the target population, wherein the at least one target event is used for an update process of the current risk control policy.

12. The system according to claim 11, wherein for breeding, the at least one processor performs the following operations:

performing a crossover operation on different events in a current population, to generate a plurality of evolutionary events;

evicting some events from the plurality of evolutionary events, wherein the some events do not satisfy a preset risk evolution condition; and

using a remaining event in the plurality of evolutionary events as a population obtained after i^{th}-generation breeding, wherein

when $i = 1$, the current population is the initial population; when $1 < i \leq N$, the current population is a population obtained after the $(i-1)^{th}$-generation breeding; and the target population is a population obtained after N^{th}-generation breeding.

13. The system according to claim 12, wherein one event corresponds to a plurality of attributes, an attribute value of some attributes of the evolutionary event is from a first event, an attribute value of the other attributes of the evolutionary event is from a second event, and the first event and the second event are different events in the current population.

14. The system according to claim 13, wherein the plurality of attributes comprises at least a risk attribute, and an attribute value of the risk attribute represents whether there is a known risk; and

the evolutionary event is obtained by replacing the attribute value of the some attributes in the first event with an attribute value of a corresponding attribute in the second event, the attribute value of the risk attribute in the first event represents that there is a known risk, and the some attributes comprise at least one attribute other than the risk attribute in the plurality of attributes.

15. The system according to claim 12, wherein the risk evolution condition comprises: the evolutionary event is identified as a risk-free event based on the current risk control policy.

16. The system according to claim 12, wherein a quantity of events evicted from the plurality of evolutionary events satisfies the following requirement: an absolute value of a difference between a quantity of events remaining after eviction and a quantity of events in the initial population is less than a preset threshold.

17. The system according to claim 12, wherein to evict the some events from the plurality of evolutionary events, the at least one processor performs the following operation:

performing a mutation operation on at least some evolutionary events in the plurality of evolutionary events, to obtain an evolutionary event set comprising a plurality of mutated events; and

evicting the some events from the evolutionary event set.

18. The system according to claim 17, wherein one event corresponds to a plurality of

attributes, and the attributes comprise at least a risk attribute; and

the mutation operation corresponding to one evolutionary event comprises at least one of the following:

randomly adding an attribute value of at least one attribute of the evolutionary event;

randomly deleting an attribute value of at least one attribute other than the risk attribute in the evolutionary event; and

randomly changing an attribute value of at least one attribute other than the risk attribute in the evolutionary event.

19. The system according to claim 11, wherein each historical event corresponds to a risk attribute, and an attribute value of the risk attribute represents whether the historical event has the known risk;

in the process of the N-generation breeding, attribute values of the risk attributes respectively corresponding to the plurality of historical events are inherited to an event in the target population generation by generation; and

to determine the at least one target event with the unknown risk in the target population, the at least one processor performs the following operation:

for each event in the target population, if an attribute value of the risk attribute corresponding to the event represents that there is the known risk, determining the event as the target event.

20. A computer-readable non-volatile storage medium, wherein the computer-readable non-volatile storage medium stores at least one instruction set, and when the at least one instruction set is executed by at least one processor, the method according to any one of claims 1 to 10 is implemented.