ORDER CLUSTERING AND MALICIOUS INFORMATION COMBATING METHOD AND APPARATUS

[0001] The present application claims priority to Chinese Patent Application No. 201510795161.9, filed on November 18, 2015 and entitled "ORDER CLUSTERING AND MALICIOUS INFORMATION COMBATING METHOD AND APPARATUS", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present application relates to the field of network communications technologies, and in particular, to order clustering and malicious information combating method and apparatus.

BACKGROUND

[0003] In existing e-commerce technologies, each transaction is considered as an "order". For example, initiation of the "Collect" service in the Alipay application software is considered as an order, and each remittance transaction is also considered as an order. Malicious information can exist in an order. For example, a malicious user of the "Collect" service in the Alipay application software can send a "Collect" request to a large number of Alipay users, and changes the consumption record note corresponding to each order to a phishing website address. Most phishing websites include slogans such as "Seven Anniversary of Taobao". Because the websites are linked to the official website of Alipay, the users may trust and visit the websites. It can cause large capital losses for the users, and affect the reputation of Alipay.

[0004] In the existing technologies, keywords are usually used to filter and identify malicious information, but there are certain disadvantages. For example, keyword filtering can function only when the keywords filtered are known as junk keywords. When the malicious user makes adjustments by modifying the keywords included in the malicious information, the keyword filtering solution may fail and cannot play a proactive role in identifying the malicious information. Filtering

malicious information generated in various ways and in various forms can be improved only through manually monitoring and adding keywords. This passive countermeasure can lead to high costs in identifying malicious information.

SUMMARY

[0005] To resolve the existing problem of inaccurate and inflexible order classifications in electronic transactions, order clustering and malicious information combating method and apparatus can be provided. In the technical solutions of the present application, after an order is preliminarily processed, a clustering operation can be performed to more accurately classify the order and reduce human involvement. In addition, orders in a malicious information cluster can be quickly processed through the corresponding malicious information combating method and apparatus.

[0006] An implementation of the present application provides a method for order clustering in an electronic transaction, including: generating a unique order identifier (ID) and order content based on information about an order; converting the order content to an order vector; and performing a clustering operation based on the unique order ID and the order vector to obtain an order clustering result.

[0007] An implementation of the present application further provides a method for combating malicious information based on order clustering, including: generating a unique order ID and order content based on information about an order; converting the order content to an order vector; performing a clustering operation based on the unique order ID and the order vector to obtain an order clustering result; analyzing content of a representative order in each cluster; and if the content of the representative order includes malicious information, performing corresponding processing of orders in the cluster.

[0008] An implementation of the present application further provides an apparatus for order clustering in an electronic transaction, including: a generation unit, configured to generate a unique order ID and order content based on information about an order; a vector conversion unit, configured to convert the order content to an order vector; and a clustering operation unit, configured to perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result.

[0009] An implementation of the present application further provides an apparatus for combating malicious information based on order clustering, including: a generation unit, configured to generate a unique order ID and order content based on information about an order; a vector conversion unit, configured to convert the order content to an order vector; a clustering operation unit, configured to perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result; and a processing unit, configured to: analyze content of a representative order in each cluster, and if the content of the representative order includes malicious information, perform corresponding processing of orders in the cluster.

[0010] It can be seen from the technical solutions provided in the implementations of the present application, a clustering algorithm is used to automatically classify orders in electronic transactions; a new cluster can be obtained through analysis based on real-time content of the order; and an obtained cluster including malicious information can be correspondingly processed. As such, human involvement can be reduced, and a new cluster can be automatically processed.

[0011] Certainly, products or methods that implement the present application do not necessarily need to achieve the earlier mentioned advantages.

BRIEF DESCRIPTION OF DRAWINGS

[0012] To describe the technical solutions in the implementations of the present application or in the prior art more clearly, the following briefly describes the accompanying drawings required for describing the implementations or the prior art. Apparently, the accompanying drawings in the following description merely show some implementations of the present application, and a person of ordinary skill in the art can still derive other drawings from these accompanying drawings without creative efforts.

[0013] FIG. 1 is a flowchart illustrating a method for order clustering in an electronic transaction, according to an implementation of the present application;

[0014] FIG. 2 is a schematic structural diagram illustrating an apparatus for order clustering in an electronic transaction, according to an implementation of the present application;

[0015] FIG. 3 is a flowchart illustrating a method for combating malicious

information based on order clustering, according to an implementation of the present application;

[0016] FIG. 4 is a schematic structural diagram illustrating an apparatus for combating malicious information based on order clustering, according to an implementation of the present application;

[0017] FIG. 5 is a flowchart illustrating a method for combating malicious information based on order clustering in an electronic transaction, according to an implementation of the present application; and

[0018] FIG. 6 is a flowchart illustrating a method for quickly classifying an order after clustering processing, according to an implementation of the present application.

DESCRIPTION OF EMBODIMENTS

[0019] The implementations of the present application provide order clustering and malicious information combating method and apparatus.

[0020] To make a person skilled in the art understand the technical solutions in the present application better, the following clearly and completely describes the technical solutions in the implementations of the present application with reference to the accompanying drawings in the implementations of the present application. Apparently, the described implementations are merely some but not all of the implementations of the present application. Other implementations obtained by a person of ordinary skill in the art based on the implementations of the present application without creative efforts shall fall within the protection scope of the present application.

[0021] FIG. 1 is a flowchart illustrating a method for order clustering in an electronic transaction, according to an implementation of the present application. In this implementation, some information transferred is normal, and some information transferred is malicious. Because information transferred by a user is forwarded by a server in an electronic transaction system, the method in this implementation can be implemented by the server of the electronic transaction system. A clustering algorithm such as the mean-shift clustering algorithm can be used to automatically classify the information transferred, and a new cluster can be automatically created. A developer does not need to predetermine the number of clusters. This classification method can be more flexible, improves self-identification accuracy of content of the orders, and classification accuracy. In addition, it can reduce human involvement, improve work

efficiency, and improve processing efficiency of subsequent malicious information combating.

[0022] FIG. 1 includes the following steps: Step 101: Generate a unique order ID and corresponding order content based on information about an order.

[0023] Step 102: Convert the order content to an order vector.

[0024] Step 103: Perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result.

[0025] In this implementation, the mean-shift clustering algorithm can be used to perform the clustering operation on the order to obtain the order clustering result. A mean-shift clustering operation is a non-parametric multi-model segmentation method. Its basic computation module uses a conventional pattern recognition program. Segmentation is achieved based on analyzing an image eigenspace and a clustering method. In the mean-shift clustering operation, a local maximum value of an eigenspace probability density function is directly estimated to obtain a density pattern of an unknown cluster and determine a location of the pattern. Then the order is added to a cluster related to the pattern. In actual implementations, a proper clustering algorithm can be selected or used as needed.

[0026] According to an implementation of the present application, the method further includes a step of mapping content of an order in the cluster to a characteristic string. When a mapped string of content of an order to be analyzed is the same as the characteristic string, the order to be analyzed is directly added to the cluster.

[0027] According to an implementation of the present application, before generating the unique order ID and order content based on information about an order, the method further includes: importing the order to a Hadoop distributed file system (HDFS). Subsequent steps in the present application can be performed in the HDFS to improve processing efficiency. Certainly, it is optional to perform other steps of the present disclosure in the HDFS. Those steps can be performed in other distributed systems, and the efficiency may also be improved. Details are omitted here for simplicity.

[0028] According to an implementation of the present application, the generating a unique order ID and order content based on information about an order further includes: generating the unique order ID based on a user ID, an order ID, and a service type in the information about the order; and generating the order content based on an order title or an order remark in the information about the order.

[0029] In addition to generating the unique order ID based on a combination of the three fields: the user ID, the order ID, and the service type in this implementation, the unique order ID can be generated based on any combinations of the user ID, the order ID, and the service type. Alternatively, content of another field (such as an external order number) can also be included to generate the unique order ID.

[0030] The order title can include a product description and an order description. For example, an order title of a collection service can include a service name, a payor, or a message. The order remark can include promotional information of a user who initiates the order, such as a uniform resource locator (URL) address, a promotional text, or a promotional picture.

[0031] According to an implementation of the present application, converting the order content to an order vector includes: converting the order content to a term frequency–inverse document frequency (TF-IDF) vector.

[0032] In this implementation of the present application, the mean-shift clustering algorithm can be used to automatically classify the order in the electronic transaction, and a new cluster can be obtained through analysis based on real-time content of the order. When the order is processed, it can be transcoded. It can then be determined whether the transcoded order is the same as a characteristic string of an order in a certain cluster, in order to further improve order classification efficiency. Orders can be imported to a distributed file system such as the HDFS to further improve order classification efficiency, by taking advantage of the characteristics of the distributed system.

[0033] FIG. 2 is a schematic structural diagram illustrating an apparatus for order clustering in an electronic transaction, according to an implementation of the present application. The apparatus in this implementation can be installed on a server of an electronic transaction system. Or, a high-performance server can be separately used to implement the apparatus in this implementation. Each functional unit or module of the apparatus in this implementation can be implemented by a dedicated central processing unit (CPU), a single-chip microcomputer, or a field programmable gate array (FPGA), or can be implemented by a general-purpose processor executing control logic of each functional module. A functional module of the apparatus in this implementation can perform a mean-shift clustering operation to effectively classify an order, and can obtain a new cluster based on order content. A developer does not need to predetermine the number of clusters. As such, self-identification of combating

malicious information based on order clustering, classification accuracy, and work efficiency can be improved, and human involvement can be reduced.

[0034] The apparatus in this implementation includes: a generation unit 201, configured to generate a unique order ID and order content based on information about an order; a vector conversion unit 202, configured to convert the order content to an order vector; and a clustering operation unit 203, configured to perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result.

[0035] According to an implementation of the present application, the apparatus further includes a conversion unit 204 that connects the generation unit and the clustering operation unit, and configured to: map content of an order in the cluster to a characteristic string. When a mapped string of content of an order is the same as the characteristic string, add the order to the cluster.

[0036] According to an implementation of the present application, the apparatus further includes an importing unit 205, configured to import the order to a Hadoop distributed file system (HDFS).

[0037] According to an implementation of the present application, the generation unit is further configured to: generate the unique order ID based on a user ID, an order ID, and a service type included in the order information; and generate the order content based on an order title or an order description included in the order information.

[0038] According to an implementation of the present application, the vector conversion unit is further configured to convert the order content to a TF-IDF vector.

[0039] In the apparatus in this implementation of the present application, the mean-shift clustering algorithm can be used to automatically classify the order in the electronic transaction, and a new cluster can be obtained through analysis based on real-time content of the order. When the order is processed, it can be transcoded. It can then be determined whether the transcoded order is the same as a characteristic string of an order in a certain cluster, in order to further improve order classification efficiency. The order can be imported to a distributed file system such as HDFS to further improve order classification efficiency, by taking advantage of the characteristics of the distributed system.

[0040] FIG. 3 is a flowchart illustrating a method for combating malicious information based on order clustering, according to an implementation of the present

application. In this implementation, a generated cluster is analyzed to determine which cluster includes malicious information (such as a junk order). An electronic transaction risk can be avoided by performing corresponding processing of orders in the cluster.

[0041] In the figure, the following steps are included: Step 301: Generate a unique order ID and order content based on information about an order.

[0042] Step 302: Convert the order content to an order vector.

[0043] Step 303: Perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result.

[0044] Step 304: Analyze content of a representative order in each cluster, and if the content of the representative order includes malicious information, perform corresponding processing of orders in the cluster.

[0045] An existing method can be used to determine whether the content of the order in the cluster is malicious information. For example, whether content of a certain order in the cluster is malicious information is determined based on a keyword, or the order content is manually screened. The corresponding processing can include recording a user ID in the cluster, sending the user ID to an administrator for banning the user ID, sending alarm information to the user ID, etc.

[0046] According to an implementation of the present application, the method further includes a step of mapping content of an order in the cluster to a characteristic string. When a mapped string of order content of an order is the same as the characteristic string, the order is directly added to the cluster.

[0047] According to an implementation of the present invention, when the mapped string of the order is the same as the characteristic string, the method further includes: when the cluster has been marked as a junk class, directly perform automatic malicious information processing such as freezing the transaction or blocking the user account associated with the order.

[0048] The order content is of an order randomly selected from the cluster.

[0049] Based on the earlier described steps, the order to be analyzed can be classified based on the clusters obtained after the clustering operation. For a certain user, most information transferred (e.g. an order) can be the same. Therefore, the cluster that includes the information transferred has been obtained through the earlier described clustering algorithm. A characteristic sample (the most representative information transferred) in the cluster is transcoded to an MD5 code. When

information transferred is received again, whether the information transferred is included in the cluster can be determined by converting the information transferred to an MD5 code and comparing the two MD5 codes. Based on the earlier described technical solution, a person skilled in the art can infer that other types of transcoding can be feasible. Details are not repeated here for simplicity.

[0050] In the method in this implementation of the present application, the mean-shift clustering algorithm can be used to automatically classify the order in the electronic transaction (in actual implementations, a proper clustering algorithm can be selected or used as needed), and a new cluster can be obtained through analysis based on real-time content of the order. When the order is processed, it can be transcoded. It can then be determined whether the transcoded order is the same as a characteristic string of an order in a certain cluster, in order to further improve order classification efficiency. The order can be imported to a distributed file system such as an HDFS to further improve order classification efficiency, by taking advantage of the characteristics of the distributed system. Because only similar orders are clustered in one cluster, content of one representative order in a cluster can be analyzed to determine whether the entire cluster is a malicious information cluster, which simplifies the process of comparing and determining each order.

[0051] FIG. 4 is a schematic structural diagram illustrating an apparatus for combating malicious information based on order clustering, according to an implementation of the present application. The apparatus in this implementation can be installed on a server of an electronic transaction system. Or, a high-performance server can be separately used to implement the apparatus in this implementation. Each functional unit or module of the apparatus in this implementation can be implemented by a dedicated CPU, a single-chip microcomputer, or an FPGA, or can be implemented by a general-purpose processor executing control logic of each functional module. A functional module of the apparatus in this implementation can perform a mean-shift clustering operation to effectively classify an order, and can obtain a new cluster based on order content. Developers do not need to predetermine the number of clusters. As such, self-identification of combating malicious information based on order clustering, classification accuracy, and work efficiency can be improved, and human involvement can be reduced.

The implementation illustrated in the figure includes: a generation unit 401, configured to generate a unique order ID and order content based on information

about an order; a vector conversion unit 402, configured to convert the order content to an order vector; a clustering operation unit 403, configured to perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result; and a processing unit 404, configured to analyze content of a representative order in each cluster, and if the content of the representative order includes malicious information, perform corresponding processing of orders in the cluster.

[0052] According to an implementation of the present application, the apparatus further includes a conversion unit 405 that connects the generation unit and the clustering operation unit, and configured to map content of an order in the cluster to a characteristic string. When a mapped string of order content of an order is the same as the characteristic string, add the order to be analyzed to the cluster.

[0053] According to an implementation of the present application, the conversion unit is further configured to directly send the order to be analyzed to the processing unit for automatic malicious information processing (such as transaction freezing or account blocking), when the cluster has been marked as a junk class.

[0054] In the apparatus in this implementation of the present application, the mean-shift clustering algorithm can be used to automatically classify the order in the electronic transaction (in actual implementations, a proper clustering algorithm can be selected or used as needed), and a new cluster can be obtained through analysis based on real-time content of the order. When the order is processed, it can be transcoded. It can then be determined whether the transcoded order is the same as a characteristic string of an order in a certain cluster, in order to further improve order classification efficiency. The order can be imported to a distributed file system such as an HDFS to further improve order classification efficiency, by taking advantage of the characteristics of the distributed system. Because only similar orders are clustered in one cluster, content of one representative order in a cluster can be analyzed to determine whether the entire cluster is a malicious information cluster, which simplifies the process of comparing and determining each order.

[0055] FIG. 5 is a flowchart illustrating a method for combating malicious information based on order clustering in an electronic transaction, according to an implementation of the present application. In this implementation, the entire method is performed in a Hadoop distributed file system to improve overall work efficiency. In a clustering process, the mean-shift clustering method is used as an example for

description. For example, for a collection service, an order initiated by the collection service includes a user ID, an order ID, a service type, a payor ID, a date, an amount, etc. The order further includes remark information. The remark information includes a uniform resource locator (URL) of a phishing website. A malicious user sends the information transferred to many users.

[0056] The following steps are included: Step 501: Import an order obtained from a database to a Hadoop distributed file system (HDFS).

[0057] Step 502: Use a combination of a user ID, an order ID, and a service type in the order as a unique ID of the order.

[0058] The combination in the present step may include combining the user ID, the order ID, and the service type to form the unique ID of the order or combining the last two characters of the user ID, the last two characters of the order ID, and the service type to form the unique ID of the order. The unique ID can also be generated in another way to uniquely identify the order. For example, the order can be uniquely identified by only the order ID.

[0059] Step 503: Use remark information in the order as content of the order.

[0060] In this implementation, the remark information of the order describes the content of the order. For example, the remark information includes a URL address of a phishing website. In another implementation, if the order further includes content such as a transaction record, this part of content can also be used as the content of the order. The transaction record can include a product description in the order. For a payment collection application, the transaction record can include a service title, a payor, an amount, etc.

[0061] In this case, the unique ID of each order corresponds to the content of the order.

[0062] Step 504: Convert the unique ID of the order and the content of the order to a sequencefile format.

[0063] In the present step, the sequencefile format obtained after the conversion is a data format used for the Hadoop distributed file system (HDFS). The processing method based on the HDFS can be used only after the unique ID of the order and the content of the order are converted to this format, so as to improve processing efficiency.

[0064] Step 505: Convert the content of the order to a TF-IDF vector.

[0065] The TF-IDF vector obtained after the conversion has a mapping

relationship <key, vector> with the unique ID of the order. The key is the unique ID of the order. The vector is formed by the content of the order, and is a mathematical expression form of the content of the order.

[0066] Step 506: Perform a mean-shift clustering operation on the vector to obtain an order clustering result.

[0067] In the present step, the content of the order is iteratively compared with a cluster. A cluster or a new cluster that includes the content of the order is output. Because there is a key-value (the unique ID of the order) that has a mapping relationship with the vector, a cluster that includes a particular order can be determined. An iteration parameter, a convergence parameter, etc. in the mean-shift clustering operation can be set by a developer based on actual implementations. An output result is in the sequencefile format.

[0068] The order classification process based on the mean-shift clustering operation ends at this point.

[0069] Step 507: Convert the order clustering result to a text format.

[0070] The order clustering result is in the sequencefile format in the earlier described step. It may not be convenient to subsequently determine and process the order clustering result. Therefore, the order clustering result needs to be converted to a text format.

[0071] Step 508: Determine whether content of an order in the cluster includes malicious information. If the content includes malicious information, proceed to step 509. If the content does not include malicious information, proceed to step 510.

[0072] Technologies such as a keyword-based method or a manual determining method can be used to determine whether the content of the order includes malicious information. One cluster can include many orders. In some cases, just one order in the cluster needs to be selected for the determination.

[0073] In another implementation, since a large amount of information can be transferred, and especially a larger amount of malicious information can be transferred, each cluster includes a plurality of orders after the clustering operation ends. Several clusters that include larger numbers of orders are selected (for example, select the top ten clusters) and sent to an administrator for monitoring, to achieve good monitoring result. Which clusters are more likely to be classified as malicious information can be determined based on the content of the selected top ten clusters.

[0074] Step 509: Mark the user who initiates the order to provide prompt

information for an administrator.

[0075] In addition to the processing method described in the present step, there are other malicious information processing methods such as transaction freezing or account blocking. Those methods are not described here in detail for simplicity.

[0076] Step 510: If the content does not include malicious information, no further processing needs to be performed.

[0077] In the earlier described implementation, the mean-shift clustering operation can be used to classify an order in the electronic transaction system, and a new cluster can be automatically generated to reduce human involvement and adapt to classification of the changing malicious information in real time.

[0078] FIG. 6 is a flowchart illustrating a method for quickly classifying an order after clustering processing, according to an implementation of the present application. This method is based on the implementation illustrated in FIG. 5. Each cluster includes many orders. A system receives and processes a new order to be analyzed.

[0079] FIG. 6 includes the following steps: Step 601: Extract content of an order in each cluster.

[0080] In step 601, the content of the order in each cluster can be randomly extracted. Because content of orders in each cluster is similar, any extracted content of an order can be used as a characteristic of the cluster.

[0081] Step 602: Perform an MD5 transcoding operation on the extracted content of the order to obtain a characteristic string.

[0082] In the present step, a characteristic string of each cluster can be obtained. In an optional implementation, the MD5 transcoding operation is one of many transcoding operations, other transcoding methods can also be used. Or, a developer can design a transcoding method as needed. The transcoding method used is not limited here.

[0083] Step 603: Obtain an order to be analyzed.

[0084] In the present step, the order to be analyzed can be obtained based on, for example, step 503 in the implementation illustrated in FIG. 5. Because the implementation illustrated in FIG. 5 is usually used to analyze an existing order, the implementation illustrated in FIG. 6 can be performed based on the implementation illustrated in FIG. 5. A new order can be analyzed after a plurality of clusters are established by using the earlier described method. It can achieve better real-time performance and avoid relatively complicated clustering operations.

[0085] Or, in the present step, information transferred (the order) can be compared and analyzed in real time to determine a cluster that includes the order.

[0086] Step 604: Perform an MD5 transcoding operation on content of the order to obtain a mapped string.

[0087] Step 605: Compare the characteristic string with the mapped string. If the two strings are consistent, proceed to step 606. If the two strings are inconsistent, proceed to step 607.

[0088] Step 606: Add the order to be analyzed to a cluster that includes the characteristic string.

[0089] Afterward, return to step 508 to continue cluster processing. If the content includes malicious information, the information transferred can be blocked from being forwarded.

[0090] In an optional step, when the mapped string of the order to be analyzed is the same as a characteristic string of certain cluster that has been marked as a junk class, directly proceed to step 509 without performing step 606. And automatic preventive measures such as transaction freezing or account blocking can be performed on the order.

[0091] Step 607: Return to step 504 to continue to perform clustering processing.

[0092] In the methods and the apparatuses in the implementations of the present application, the mean-shift clustering algorithm can be used to automatically classify the order in the electronic transaction, and a new cluster can be obtained through analysis based on real-time content of the order. When the order is processed, it can be transcoded. It can then be determined whether the transcoded order is the same as a characteristic string of an order in a certain cluster, in order to further improve order classification efficiency. The order can be imported to a distributed file system such as the HDFS to further improve order classification efficiency, by taking advantage of the characteristics of the distributed system.

[0093] For technical improvements, one can distinguish between a hardware improvement (for example, an improvement in circuit structures of a diode, a transistor, a switch, etc.) and a software improvement (an improvement in a method process). However, as technologies develop, improvements in many current method processes can be considered as a direct improvement in a hardware circuit structure. Almost all designers obtain corresponding hardware circuit structures by programming improved method processes to hardware circuits. Therefore, it cannot

be said that the improvement in a method process cannot be implemented by using a hardware entity module. For example, a programmable logic device (PLD), such as a field programmable gate array (FPGA), is such an integrated circuit whose logic function is determined by programming a device by a user. A designer performs programming to "integrate" a digital system to a single PLD, without requiring a chip manufacturer to design and manufacture a dedicated integrated circuit chip 2. In addition, at present, instead of manually manufacturing an integrated chip, this type of programming is mostly implemented by using "logic compiler" software. The programming is similar to a software compiler used to develop and write a program. Original code needs to be written in a particular programming language for compilation. The language is referred to as a hardware description language (HDL). There are many HDLs, such as the Advanced Boolean Expression Language (ABEL), the Altera Hardware Description Language (AHDL), Confluence, the Cornell University Programming Language (CUPL), HDCal, the Java Hardware Description Language (JHDL), Lava, Lola, MyHDL, PALASM, and the Ruby Hardware Description Language (RHDL). The very-high-speed integrated circuit hardware description language (VHDL) and Verilog2 are most commonly used. It should also be clear to a person skilled in the art that a hardware circuit of a logic method process can be easily obtained by performing logic programming on the method process through several earlier described hardware description languages and programming the method process to an integrated circuit.

[0094] A controller can be implemented in any appropriate way. For example, the controller can take the form of, for example, a microprocessor or a processor and a computer readable medium storing computer readable program code (such as software or firmware) that can be executed by the microprocessor or the processor, a logic gate, a switch, an application specific integrated circuit (ASIC), a programmable logic controller, and an embedded microcontroller. Examples of the controller include, but are not limited to, the following microcontrollers: ARC 625D, Atmel AT91SAM, Microchip PIC18F26K20, and Silicone Labs C8051F320. A memory controller can also be implemented as a part of control logic of a memory.

[0095] A person skilled in the art also know that, in addition to implementing the controller by using the computer readable program code, logic programming can be performed on method steps to allow the controller to implement the same function in forms of the logic gate, the switch, the programmable logic controller, the

programmable logic controller, and the embedded microcontroller. Therefore, such a controller can be considered as a hardware component. An apparatus included in the controller and configured to implement various functions can be considered as a structure in the hardware component. Or, the apparatus configured to implement various functions can even be considered as both a software module implementing the method and a structure in the hardware component.

[0096] The system, apparatus, module or unit illustrated in the earlier described implementations can be implemented by a computer chip or an entity, or can be implemented by a product having a certain function.

[0097] For ease of description, the apparatus is described by dividing the functions into various units. Certainly, when the present application is implemented, the functions of all units can be implemented in one or more pieces of software and/or hardware.

[0098] It can be learned from description of the implementations that, a person skilled in the art can clearly understand that the present application can be implemented by using software in addition to a necessary universal hardware platform. Based on such an understanding, the technical solutions in the present application essentially or the part contributing to the prior art can be implemented in a form of a software product. The software product may be stored in a storage medium, such as a ROM/RAM, a magnetic disk, or an optical disc, and includes several instructions for instructing a computer device (a personal computer, a server, or a network device) to perform the methods described in the implementations or in some parts of the implementations of the present application.

[0100] The implementations in the present specification are described in a progressive way. For same or similar parts in the implementations, reference may be made to each other. Each implementation focuses on a difference from other implementations. Particularly, a system implementation is similar to a method implementation, and therefore, is described briefly. For related parts, refer to partial descriptions in the method implementation.

[0101] The present application can be used in many general-purpose or dedicated computer system environments or configurations, for example, a personal computer, a server computer, a handheld device or a portable device, a flat panel device, a multi-processor system, a microprocessor-based system, a set-top box, a programmable consumer digital device, a network PC, a minicomputer, a mainframe

computer, and a distributed computing environment including any one of the earlier described systems or devices.

[0102] The present application can be described in the general context of an executable computer instruction executed by a computer, for example, a program module. Generally, the program module includes a routine, a program, an object, a component, a data structure, etc. for executing a particular task or implementing a particular abstract data type. The present application can also be practiced in distributed computing environments. In the distributed computing environments, tasks are performed by remote processing devices that are connected through a communications network. In a distributed computing environment, the program module can be located in both local and remote computer storage media including storage devices.

[0103] The present application is described through implementations. However, a person of ordinary skill in the art knows that many modifications and variations of the present application may be made without departing from the spirit of the present application. It is intended that the claims include these modifications and variations without departing from the spirit of the present application.

CLAIMS

What is claimed is:

1. A method for order clustering in an electronic transaction, comprising:

generating a unique order ID and order content based on information about an order;

converting the order content to an order vector; and

performing a clustering operation based on the unique order ID and the order vector to obtain an order clustering result.

2. The method according to claim 1, wherein content of an order in the cluster is mapped to a characteristic string, and when a mapped string of content of an order to be analyzed is the same as the characteristic string, the order to be analyzed is directly added to the cluster.

3. The method according to claim 1, wherein the generating a unique order ID and order content based on information about an order further comprises: generating the unique order ID based on a user ID, an order ID, and a service type in the information about the order; and generating the order content based on an order title or an order remark in the information about the order.

4. A method for combating malicious information based on order clustering, comprising:

generating a unique order ID and order content based on information about an order;

converting the order content to an order vector;

performing a clustering operation based on the unique order ID and the order vector to obtain an order clustering result; and

analyzing content of a representative order in each cluster, and if the content of the representative order includes malicious information, performing corresponding processing of orders in the cluster.

5. The method according to claim 4, wherein content of an order in the cluster is mapped to a characteristic string, and when a mapped string of content of an order to be analyzed is the same as the characteristic string, the order to be analyzed is directly added to the cluster.

6. The method according to claim 5, wherein when the mapped string of the

order content of the order to be analyzed is the same as the characteristic string, the method further comprises: when the cluster has been marked as a junk class, directly performing malicious information processing on the order to be analyzed.

7. An apparatus for order clustering in an electronic transaction, comprising:

a generation unit, configured to generate a unique order ID and order content based on information about an order;

a vector conversion unit, configured to convert the order content to an order vector; and

a clustering operation unit, configured to perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result.

8. The apparatus according to claim 7, further comprising: a conversion unit, connected between the generation unit and the clustering operation unit, and configured to: map content of an order in the cluster to a characteristic string, and when a mapped string of content of an order to be analyzed is the same as the characteristic string, directly add the order to be analyzed to the cluster.

9. The apparatus according to claim 7, wherein the generation unit is further configured to: generate the unique order ID based on a user ID, an order ID, and a service type in the information about the order; and generate the order content based on an order title or an order remark in the information about the order.

10. An apparatus for combating malicious information based on order clustering, comprising:

a generation unit, configured to generate a unique order ID and order content based on information about an order;

a vector conversion unit, configured to convert the order content to an order vector;

a clustering operation unit, configured to perform a clustering operation based on the unique order ID and the order vector to obtain an order clustering result; and

a processing unit, configured to: analyze content of a representative order in each cluster, and if the content of the representative order includes malicious information, perform corresponding processing of orders in the cluster.

11. The apparatus according to claim 10, further comprising: a conversion unit, connected between the generation unit and the clustering operation unit, and configured to: map content of an order in the cluster to a characteristic string, and when a mapped string of content of an order to be analyzed is the same as the

characteristic string, directly add the order to be analyzed to the cluster.

12. The apparatus according to claim 11, wherein the conversion unit is further configured to: when the cluster has been marked as a junk class, directly send the order to be analyzed to the processing unit for malicious information processing.