159159

D E S C R I P T I O N

EVALUATION TARGET OF INTEREST EXTRACTION APPARATUS AND PROGRAM

G00002

Technical Field

Embodiments described herein relate generally to an evaluation target of interest extraction apparatus and program.

Background Art

In a computer and network environment, numeric time series data can be collected for a specific target, and text time series data can be acquired for the specific target. In such environment, a method of discovering a pattern which explains a variation of the numeric time series data by the text time series data, and predicting an evaluation target of interest in the next term based on the numeric time series data and text time series data has been studied.

For example, in the stock market, by considering stock prices as numeric time series data, and news articles related to brands as text time series data, a method of discovering a pattern which is extracted from news articles to explain a variation of stock prices, notifying a user of a brand of interest in the next term, and supporting decision-making related to sales by brand of the user has been studied.

As a method of predicting an evaluation target, for example, two methods have been proposed.

The first method characterizes an evaluation target in advance by explicitly given attribute values, generates time series data based on frequencies of events related to occurrence of the evaluation target, and calculates a degree of importance of the evaluation target or those of the attribute values of the evaluation target, so as to extract an important evaluation target in a specific problem domain.

On the other hand, the second method visually recognizably displays the relationship between an evaluation target and specific words to the user by associating the evaluation target and time-dependent changes in frequencies of occurrence of specific words with each other.

Citation List

Patent Literature

Patent Literature 1: Jpn. Pat. Appln. KOKAI Publication No. 2002-207755

Patent Literature 2: Jpn. Pat. Appln. KOKAI Publication No. 2001-216311

However, the aforementioned two methods normally do not pose any problem, but they suffer the following disadvantages according to the examinations of the present inventor.

For example, the first method cannot handle an

evaluation target which cannot be characterized in advance since an evaluation target is characterized in advance by explicit attribute values. Also, the first method limits time series data related to an evaluation target to those of events related to occurrence of the evaluation target.

On the other hand, the second method allows the user to visually recognize the relationship between an evaluation target and specific words, but it cannot automatically discover a pattern which can explain the case of occurrence of a specific relationship.

A solution to such problem of the present invention is to provide an evaluation target of interest extraction apparatus and program, which can handle an evaluation target which cannot be characterized in advance, does not limit time series data to those related to occurrence of the evaluation target, and can automatically discover a pattern.

Summary of Invention

An evaluation target of interest extraction apparatus according to one embodiment is an apparatus which can discover a pattern before an evaluation target of interest is extracted based on the pattern from a plurality of pieces of text information which are given along a time series and are related to a plurality of evaluation targets.

The evaluation target of interest extraction apparatus comprises text collection means, numeric collection means,

target expression storage means, item allocation means, class distinguishing means, transaction generation means, pattern discovery means, and pattern storage means.

The text collection means collects the plurality of pieces of text information.

The numeric collection means collects numeric information individually related to each evaluation target.

The target expression storage means stores an evaluation target expression which individually expresses each evaluation target, and a related expression which expresses a target individually related to each evaluation target in association with each other.

The item allocation means extracts a plurality of items for each piece of collected text information, and when the items include an item that matches the evaluation target expression or the related expression, the item allocation means allocates the evaluation target expression or an evaluation target expression associated with the related expression in the target expression storage means to the text information as an evaluation target item.

The class distinguishing means distinguishes a class related to a change in numeric information based on the numeric information related to an evaluation target expressed by the evaluation target expression as the evaluation target item, and allocates the class to text information to which the evaluation target item is allocated.

The transaction generation means generates, for each piece of collected text information, a class-attached transaction including the allocated evaluation target item, the extracted items, and the allocated class.

The pattern discovery means discovers the pattern indicating a combination of characteristic items from a set of the transactions.

The pattern storage means stores the discovered pattern.

Brief Description of Drawings

FIG. 1 is a block diagram showing the arrangement of an evaluation target of interest extraction apparatus according to the first embodiment.

FIG. 2 is a flowchart for explaining the operation according to the first embodiment.

FIG. 3 is a table showing an example of numeric time series data according to the first embodiment.

FIG. 4 is a table showing an example of text time series data according to the first embodiment.

FIG. 5 is a table showing an example of item sets according to the first embodiment.

FIG. 6 is a table showing an example of evaluation target knowledge according to the first embodiment.

FIG. 7 is a table showing an example of numeric time series data according to the first embodiment.

FIG. 8 is a table showing an example of numeric time series data according to the first embodiment.

FIG. 9 is a table showing an example of class allocation results according to the first embodiment.

FIG. 10 is a table showing an example of class-attached transactions according to the first embodiment.

FIG. 11 is a table showing an example of patterns according to the first embodiment.

FIG. 12 is a table showing an example of patterns according to the first embodiment.

FIG. 13 is a block diagram showing the arrangement of an evaluation target of interest extraction apparatus according to the second embodiment.

FIG. 14 is a flowchart for explaining the operation according to the second embodiment.

FIG. 15 is a table showing an example of text time series data according to the second embodiment.

FIG. 16 is a table showing an example of item sets according to the second embodiment.

FIG. 17 is a table showing an example of evaluation results of text time series data according to the second embodiment.

Description of Embodiments

Embodiments will be described hereinafter with reference to the drawings. Note that an evaluation target of interest extraction apparatus to be described below can be implemented by either a hardware arrangement or a combined arrangement of hardware resources and software.

As the software of the combined arrangement, a program, which is installed from a network or non-transitory computer-readable storage medium M in a computer in advance, and is executed by a processor of the computer to control the computer to implement functions of the evaluation target of interest extraction apparatus, as shown in FIGS. 1 and 13, is used.

<First Embodiment>

FIG. 1 is a block diagram showing the arrangement of an evaluation target of interest extraction apparatus according to the first embodiment. This evaluation target of interest extraction apparatus can discover a pattern before it extracts an evaluation target of interest based on a pattern from a plurality of pieces of text information which are given along a time series and are related to a plurality of evaluation targets.

More specifically, the evaluation target of interest extraction apparatus includes a text time series collection unit B1, item extraction unit B2, evaluation target knowledge storage unit B3, numeric time series collection unit B4, class distinguishing unit B5, transaction generation unit B6, pattern discovery unit B7, pattern storage unit B8, and evaluation target extraction unit B9. That is, the evaluation target of interest extraction apparatus can discover a pattern by the respective units B1 to B8 before it extracts an evaluation target of interest by the evaluation target extraction unit B9.

Note that the text time series collection unit (text collection means) B1 has a function of collecting a plurality of pieces of text information which are given along a time series and are related to a plurality of evaluation targets. Note that "text information" may also be referred to as "text time series data".

The item extraction unit (item allocation means) B2 has a function of extracting a plurality of items for each piece of text information collected by the text time series collection unit B1, and allocating, when the items include an item which matches an evaluation target expression or related expression in the evaluation target knowledge storage unit B3, that evaluation target expression or an evaluation target expression which is associated with the related expression in the evaluation target knowledge storage unit B3, to the text information as an evaluation target item.

The evaluation target knowledge storage unit (target expression storage means) B3 is a storage device which is read/write accessible from the respective units B1, B2, B4 to B7, and B9, and stores an evaluation target expression which individually expresses each evaluation target and a related expression which expresses a target individually related to each evaluation target in association with each other. For example, in the evaluation target knowledge storage unit B3, a control unit (not shown) writes, in advance, evaluation target knowledge which describes an

evaluation target expression and related expression in association with each other. Note that in the evaluation target knowledge storage unit B3, evaluation target knowledge described in advance in a part of a program of the evaluation target of interest extraction apparatus may be written from the control unit at an activation timing, or described evaluation target knowledge may be written according to a user operation.

The numeric time series collection unit (numeric collection means) B4 has a function of collecting numeric information individually related to each evaluation target. Note that "numeric information" may also be referred to as "numeric information which is given along a time series and is individually related to each evaluation target" or "numeric time series data". Also, time series data including "text time series data" and "numeric time series data" may be referred to as "composite event time series data".

The class distinguishing unit B5 has a function of distinguishing a class based on numeric information related to an evaluation target expressed by an evaluation target expression as an evaluation target item allocated by the item extraction unit B2, and allocating the class to text information to which the evaluation target item is allocated.

The transaction generation unit B6 has a function of generating, for each text information collected by the text

time series collection unit B1, a transaction including an evaluation target item allocated by the item extraction unit B2, items extracted by the item extraction unit B2, and a class allocated by the class distinguishing unit B5.

The pattern discovery unit B7 has a function of discovering a pattern indicating a combination of characteristic items from a set of transactions generated by the transaction generation unit B6. Note that the "combination of characteristic items" may also be referred to as a "combination of items having high frequencies of occurrence" or "combination of items of frequent occurrence". A discovered pattern is written in the pattern storage unit B8 by the pattern discovery unit B7.

The pattern storage unit B8 is a storage device which is read/write accessible from the respective units B1, B2, B4 to B7, and B9, and stores a pattern discovered by the pattern discovery unit B7. Note that different storage areas in a single storage device or different storage devices may be used as the target evaluation knowledge storage unit B3 and pattern storage unit B8.

The evaluation target extraction unit B9 has a function of extracting an evaluation target of interest by applying the pattern in the pattern storage unit B8 to the transaction generated from newly collected text information.

The operation of the evaluation target of interest extraction apparatus with the aforementioned arrangement

will be described with reference to the flowchart shown in FIG. 2. In the following description, news headings will be discussed as text time series data (text information), and stock prices will be discussed as numeric time series data (numeric information). Also, company names (stock price brand names) will be discussed as evaluation targets.

In step Sa1, the numeric time series collection unit B4 collects numeric time series data indicating stock price information for each evaluation target from a Web site, which delivers stock price information, via the Internet or the like. As the numeric time series data, for example, data including an issue code, date, and opening price, as shown in FIG. 3, are used. However, numeric time series data include only data of business days of the stock market, and do not include any data related to non-bushiness days of the stock market.

In step Sa2, the text time series collection unit B1 collects text time series data including news headings from a Web site, which delivers news headings, via the Internet or the like. As text time series data, data each including a news delivery date and news heading, as shown in FIG. 4, are used.

In step Sa3, the text time series collection unit B1 picks up one text data from the collected text time series data. At this time, if text data to be picked up is available, the process advances to step Sa4; otherwise, the process advances to step Sa11. Therefore, when text time

series data shown in FIG. 4 are collected, the process advances to step Sa11 when this step is executed in processing immediately after time series data Tn is picked up.

In step Sa4, the item extraction unit B2 extracts respective expressions which represent the picked-up text data as items. For example, the item extraction unit B2 applies morphological analysis to the text data to specify parts of speech of respective expressions, and extracts, as items, expressions given with nouns as a part of speech. Therefore, when morphological analysis is applied to text data of a text No. T1 to extract noun expressions, expressions "a 社" (Company A), "会計" (accounting), "予約" (pre-order), "ソフト" (software), "先行" (pre-), "販売" (sale), and "開始" (starts) are extracted as items. By applying the same processing to respective text data shown in FIG. 4, item sets shown in FIG. 5 can be generated from the respective text data.

In step Sa5, one item is picked up from a set of items extracted by the item extraction unit B2. If an item to be picked up is available, the process advances to step Sa6; otherwise, the process advances to step Sa7. Therefore, in the case of data T1, the process advances to step Sa7 when this step is executed in processing immediately after an item "開始" (starts) is picked up.

In step Sa6, the item extraction unit B2 refers to evaluation target knowledge stored in the evaluation target

knowledge storage unit B3 to determine whether or not the item is related to an evaluation target.

If the item extraction unit B2 determines that the item is related to an evaluation target, it allocates the evaluation target item to the text data. For example, assume that the evaluation target knowledge storage unit B3 stores evaluation target knowledge including an evaluation target expression and related expression, as shown in FIG. 6. At this time, an item "a 社" (Company A) is picked up from the text data of the text No. T1. Note that "text data of a text No. TO" will also be referred to as "text data TO" (suffix O expresses an arbitrary numeral). Likewise, an "item set of an item set No. IO" will also be referred to as an "item set IO". The item extraction unit B2 refers to the related expression of the evaluation target knowledge to determine whether or not the item is registered as the related expression. Thus, the item extraction unit B2 allocates "A 社" as an evaluation target expression corresponding to the related expression to the evaluation target item with respect to the text data T1.

Next, assume that an item "鶏肉" (chicken) is picked up from text data T7. The item extraction unit B2 refers to related expressions of the evaluation target knowledge to identify that the item is registered as related expressions in a plurality of locations. Thus, the item extraction unit B2 allocates "D1 社" (Company D1), "D2 社" (Company D2), and "D3 社" (Company D3) as evaluation target

expressions corresponding to respective related expressions to the evaluation target item with respect to the text data of the text No. T7.

On the other hand, since no corresponding related expression is available for an item "会社" (company) of text data T1 and an item "原産地" (place of origin) of text data T7, the process returns to step Sa5 without allocating any new evaluation target item to respective text data.

In step Sa7, the class distinguishing unit B5 picks up one evaluation target item allocated to text data. At this time, if an evaluation target item to be picked up is available, the process advances to step Sa8. On the other hand, if an evaluation target item to be picked up is not available, the process advances to step Sa9. For example, if an evaluation target item "A社" (Company A) is allocated to an item set I1, the process advances to step Sa9 in processing immediately after "A社" (Company A) is picked up.

In step Sa8, the class distinguishing unit B5 refers to numeric time series data corresponding to the picked-up evaluation target item to calculate a volatility as a criterion of class determination. In this case, assume that the class distinguishing unit B5 calculates a volatility corresponding to the evaluation target item based on, for example:

$$f(it,d)=\frac{value(it,dy)-value(it,d)}{value(it,d)} \quad \cdots (1)$$

where it is an evaluation target item, d is a delivery date of text data including an evaluation target, dy is a business day y days after the delivery date d, and value() is a function which returns an opening price of a stock price at a designated date of the evaluation target item. However, if d is a non-business day, a business day one day before is used.

A volatility when y = 2 is given will be calculated. Assume that "A社" (Company A) is picked up from an item set I1, and numeric time series data corresponding to "A社" are given, as shown in FIG. 3. At this time, "2010/4/26" can be obtained as a delivery date of text data T1 corresponding to the item set I1, and "2010/4/28" can be obtained as dy. By referring to FIG. 3, since value(A社, 2010/4/26) and value(A社, 2010/4/28) are respectively given as "902" and "886", f(A社, 2010/4/26) = (886 - 902)/902 = -0.01774 can be obtained. Next, a case will be examined below wherein "K社" (Company K) is picked up from an item set In, and numeric time series data corresponding to "K社" are given, as shown in FIG. 7. At this time, "2010/05/27" is obtained as a delivery date of text data Tn corresponding to the item set In, and "2010/5/31" can be obtained as a business day dy y days after. By referring to FIG. 7, since value(K社, 2010/5/27) and value(K社, 2010/5/31) are respectively

given as "802" and "816", f(K社, 2010/5/27) = (816 - 802)/802 = 0.05923 can be obtained. Likewise, when "L社" (Company L) is picked up from the item set In, and numeric time series data corresponding to "L社" are given, as shown in FIG. 8, f(L社, 2010/5/27) = (926 - 914)/914 = 0.01313 is given.

In step Sa9, the class distinguishing unit B5 calculates an evaluation value which combines volatilities calculated for respective items corresponding to text data, and distinguishes a class according to that evaluation value. In this case, the class distinguishing unit B5 calculates an average volatility as an average value of volatilities based on, for example:

$$\frac{1}{|I|}\cdot\sum_{it\in I} f(it,d) \qquad \cdots(2)$$

where I is a set of items corresponding to text data, || is a calculation for calculating the number of items included in an item set. However, assume that when an item set is an empty set, the average volatility cannot be calculated. Then, the class distinguishing unit B5 sets the average volatility as an evaluation value corresponding to text data.

If the average value is not more than a designated class distinguishing threshold Th (> 0) × -1 - Th, "down" is determined; if the average value is not less than Th, "up" is determined; and if the average value is larger

than -Th and smaller than Th, "average" is determined. However, when the average volatility cannot be calculated, a class is not allocated.

For example, in the case of text data T1, since only one item "A社" (Company A) is allocated, an average volatility = -0.01774 is given. In the case of text data Tn, since two items "K社" (Company K) and "L社" (Company L) are allocated, an average value (0.05923 + 0.01313)/2 = 0.03618 of the two items is given as an average volatility. When a class distinguishing threshold = 0.015 is given, classes "down" and "up" are respectively allocated to the text data T1 and Tn. On the other hand, since text data T2 and T8 do not include any item, and an average volatility cannot be calculated, no class is allocated.

As described above, the class distinguishing unit B5 distinguishes a class for each text data, and allocates the class, thus obtaining class allocation results to respective text data as data which represents a correspondence relationship between text data and classes, as shown in FIG. 9. However, when no class is allocated, a slash "/" is described.

In step Sa10, the transaction generation unit B6 generates a transaction by combining, from each class-attached text data, items extracted from that text data, an evaluation target item, and an allocated class. Therefore, class-attached transactions shown in FIG. 10 are generated from text data shown in FIG. 4.

In step Sa11, the pattern discovery unit B7 picks up one class from classes allocated to text data. At this time, if a class to be picked up is available, the process advances to step Sa12; otherwise, this algorithm ends. Also, assume that in this embodiment, pattern discovery is attempted from only the classes "up" and "down".

In step Sa12, the pattern discovery unit B7 extracts transactions including the designated class. Also, the pattern discovery unit B7 discovers all patterns as combinations of items of frequent occurrence having a confidence not less than a minimum confidence as a threshold related to the frequency of occurrence of a pattern using an algorithm described in a reference: R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", 20th VLDB Conference (1994), and stores the discovered patterns in the pattern storage unit B8 as those corresponding to the selected class.

The pattern discovery unit B7 generates a combination of items like "I社" (Company I) and "サーバー" (server) as a pattern candidate with respect to a class "up" in the transaction shown in FIG. 10, and leaves, as patterns, pattern candidates having a frequency of occurrence higher than a reference value of the generated pattern candidates. Thus, the pattern discovery unit B7 can discover patterns indicating combinations of characteristic items, as shown in FIG. 11, and can store them in the pattern storage unit B8. Also, as for a class "down", the pattern discovery

unit B7 can similarly discover patterns shown in FIG. 12 from frequencies of occurrence of pattern candidates, and can store them in the pattern storage unit B8.

The evaluation target extraction unit B9 extracts an evaluation target of interest by applying the patterns in the pattern storage unit B8 to transactions generated from newly collected text time series data.

As described above, according to this embodiment, a plurality of items are extracted from each piece of collected text information, and when an item which matches an evaluation target expression or related expression is found, the evaluation target expression or an evaluation target expression associated with the related expression is allocated as an evaluation target item to the text information, a class related to a change in numeric information related to the evaluation target item is distinguished, a class-attached transaction including an evaluation target item, items, and class is generated for each text information, and patterns indicating a combination of characteristic items are discovered from a set of transactions. With this arrangement, even when a clear correspondence relationship is not described between text information and numeric information, the text information and numeric information can be appropriately associated with each other, and a pattern which can explain a change in numeric information can be extracted from text information.

Further, according to this embodiment, an evaluation target which cannot be characterized in advance can be handled, time series data are not limited to those related to occurrence of an evaluation target, and a pattern can be automatically discovered.

In the conventional second method, an evaluation target of interest, which causes a specific relationship, cannot be predicted based on a discovered pattern. By contrast, according to this embodiment, the evaluation target extraction unit B9 evaluates new text information based on the discovered pattern, and can predict (extract) an evaluation target as an evaluation target which is more likely to cause a change in numeric information.

<Second Embodiment>

FIG. 13 is a block diagram showing the arrangement of an evaluation target of interest extraction apparatus according to the second embodiment. A detailed description of the same units as those in FIG. 1 will not be repeated, and only different units will be mainly described below.

That is, the second embodiment is a practical example of the first embodiment, and an evaluation target extraction unit B9 includes a pattern evaluation unit B10 and evaluation target of interest extraction unit B11. Thus, a transaction generation unit B6 has a function of generating a classless transaction, which includes evaluation target items allocated by an item extraction unit B2 and extracted items, but does not include any class

described above for each text information newly collected by a text time series collection unit B1 after patterns are stored in a pattern storage unit B8. Note that a "classless transaction" may use an arbitrary name as long as it can be distinguished from a class-attached transaction (transaction for pattern discovery) used in pattern discovery. For example, the "classless transaction" may be simply referred to as a "transaction" or may be referred to as a "transaction without any class", "transaction for evaluation target of interest extraction", "new transaction", or the like.

Note that the pattern evaluation unit (pattern evaluation means) B10 has a function of evaluating whether the classless transaction includes a pattern by applying the pattern in the pattern storage unit B8 to the classless transaction. Note that "evaluation" may also be referred to as "determination".

The evaluation target of interest extraction unit (evaluation target of interest extraction means) B11 has a function of incrementing, when the pattern is included as a result of evaluation, frequencies of occurrence associated with evaluation target items in the classless transaction, and extracting evaluation target items associated with the frequencies of occurrence as evaluation targets of interest based on the increment results of the frequencies of occurrence.

The operation of the evaluation target of interest

extraction apparatus with the aforementioned arrangement will be described below with reference to the flowchart shown in FIG. 14.

In step Sb1, the text time series collection unit B1 collects text time series data including news headings from a Web site, which delivers news headings, via the Internet or the like. As the text time series data, for example, as shown in FIG. 15, data including delivery dates of news and news headings are used. However, assume that a delivery date of text time series data of this embodiment is that after a collection date of text time series data used in pattern discovery by a pattern discovery unit B7.

In step Sb2, one text data is picked up from those collected by the text time series collection unit B1. At this time, if an item to be picked up is available, the process advances to step Sb3; otherwise, the process advances to step Sb9. For example, when text data shown in FIG. 15 are collected, the process advances to step Sb9 in processing immediately after text data Em is picked up.

In step Sb3, the item extraction unit B2 applies the same processing as in step Sa4 of FIG. 2 to the picked-up text data, thereby extracting items which represent the text data. Therefore, from text data E1, items "i 社" (Company i), "ミドル" (middle), "エンド" (end), "サーバー" (server), "音声" (speech), "認識" (recognition), "機能" (function), and "搭載" (releases) are extracted.

In step Sb4, the item extraction unit B2 picks up one

item in turn from the extracted items. At this time, if an item to be picked up is available, the process advances to step Sb5; otherwise, the process advances to step Sb6. For example, in the case of the text data E1, the process advances to step Sb6 in processing immediately after an item "搭載" is extracted.

In step Sb5, the item extraction unit B2 applies the same processing as in step Sa6 of FIG. 2 to the picked-up item, thereby evaluating whether or not an evaluation target item corresponding to the picked-up item is available. If such evaluation target item is available, the item extraction unit B2 allocates that evaluation target item to the text data.

In step Sb6, the transaction generation unit B6 generates, for each text data, a transaction including items extracted from the text data and evaluation target text data corresponding to the text data. This processing is similar to that of step Sa10 of FIG. 2, but dissimilar to the processing of step Sa10 of FIG. 2 in that no class is allocated to each transaction. Therefore, classless transactions shown in FIG. 16 can be generated from text data shown in FIG. 15.

In step Sb7, the pattern evaluation unit B10 picks up one pattern in turn from those stored in the pattern storage unit B8. At this time, if a pattern to be picked up is available, the process advances to step Sa8; otherwise, the process returns to step Sb2.

In step Sb8, the pattern evaluation unit B10 compares the picked-up classless transaction and pattern to evaluate whether or not items included in the classless transaction include those included in the pattern. At this time, if the pattern is included, a class corresponding to the pattern is referred to, and a frequency of occurrence of an evaluation target item for each class corresponding to the transaction is incremented by "1".

For example, when a transaction F1 and pattern up4 are selected, items "I社" (Company I), "サーバー" (server), and "搭載" (releases) included in the pattern up4 are included in those included in the transaction F1. Also, since "up" is given to a class of up4, values of an evaluation target item "I社" and class "up" are incremented by "1".

On the other hand, when a transaction F3 and pattern dw5 are selected, since "原産地" (place of origin) and "偽装" (disguised) included in the pattern dw4 are included in items included in the transaction F3, frequencies of occurrence of an evaluation target item "D2社" (company D2) and class "down" are incremented by "1".

Likewise, when a transaction Fm and pattern up8 are selected, frequencies of occurrence of an evaluation target item "D3社" (company D3), class "up", evaluation target item "J社" (company J), and class "up" are incremented by "1".

By making such pattern evaluation for all transactions and patterns, a relationship between evaluation target

items and frequencies of occurrence can be obtained, as shown in FIG. 17.

In step Sb9, the evaluation target of interest extraction unit B11 decides and outputs an evaluation target of interest based on frequencies of occurrence calculated by evaluating transactions using respective patterns.

For example, the number of targets of interest to be extracted is set, and as many evaluation target items with higher frequencies of occurrence as the set number of targets can be output in turn. In this case, when the top three are extracted as targets of interest from the relationship between the evaluation target items and frequencies of occurrence shown in FIG. 17, evaluation target items "D3 社" (company D3), "D2 社" (company D2), and "I 社" (company I) can be output as evaluation targets of interest by referring to the frequencies of occurrence for respective classes. Note that the evaluation target of interest extraction unit B11 may extract items having frequencies of occurrence higher than a reference value in place of some top items.

As described above, an evaluation target of interest can be decided on based on a pattern discovery and evaluation target of interest extraction apparatus from composite event time series data of claim 2.

The evaluation target of interest extraction apparatus of the second embodiment is not limited to the second

embodiment, and can be used with the following modifications described in [1] to [6] below.

[1] For example, in steps Sa11 and Sa12 of FIG. 2, a class is selected, and patterns are discovered based on a partial set of transactions divided based on the class. However, the present invention is not limited to this, and steps Sa11 and Sa12 of FIG. 2 execute processing based on an FP-tree and FP-growth to discover patterns from class-attached transactions without dividing the transactions using an algorithm described in a reference: Shigeaki Sakurai, "Discovery of Frequent Patterns from Item Sets with their Classes", The 24th Annual Conference of the Japanese Society for Artificial Intelligence, 2010, 2A3-4, <https://kaigi.org/jsai/webprogram/2010/pdf/10.pdf>.

Further, the pattern discovery unit B7 generates an FP-tree which stores item sets included in class-attached transactions in a tree structure format by this algorithm. Note that in the FP-tree for class-attached item sets, an area for storing pattern classes corresponding to patterns, a flag used to identify whether a pattern is a feature pattern or possibility pattern, and a flag indicating whether or not a degree of pattern feature support is equal to or more than a minimum degree of support are added to a header of the FP-tree. After generation of the FP-tree, the pattern discovery unit B7 executes the FP-growth of this algorithm to generate an item partial set conditioned by a specific item from the FP-tree, and further generates

an FP-tree from the item partial set. The pattern discovery unit B7 recursively repeats generation of an FP-tree and that of an item partial set, thereby discovering patterns.

[2] In step Sa9 of FIG. 2, three divided classes are obtained using one class distinguishing threshold. Using a plurality of class distinguishing thresholds, four or more divided classes can be obtained.

[3] In step Sb8 of FIG. 14, frequencies of occurrence of all evaluation target items corresponding to a transaction including a pattern are incremented by "1". However, when a plurality of evaluation target items are extracted from one transaction, a value normalized using the number of evaluation target items as a weight can be used as an increment.

[4] For evaluation target items which appear in both of a pattern and transaction, frequencies of occurrence may be incremented using a larger weight than those which appear in only a transaction.

[5] In step Sb9 of FIG. 14, evaluation targets of interest are decided based on frequencies of occurrence for respective classes. However, since a total evaluation value is calculated by incrementing frequencies of occurrence of individual classes, evaluation targets of interest may be decided using the evaluation value.

[6] In FIG. 14, after all the transactions are processed, evaluation targets of interest are output.

Alternatively, pattern evaluation may be made for each transaction, and evaluation target items corresponding to a transaction including a pattern may be output as targets of interest.

As described above, according to this embodiment, a pattern is applied to a classless transaction to evaluate whether or not the classless transaction includes the pattern. As a result of evaluation, if the transaction includes the pattern, frequencies of occurrence associated with evaluation target items in the classless transaction are incremented, and evaluation target items associated with the frequencies of occurrence are extracted as evaluation targets of interest based on the increment results of the frequencies of occurrence, thus obtaining the same effects as in the first embodiment.

Unlike a transaction for pattern discovery, since a classless transaction which does not include any class is generated, a transaction can be generated at high speed.

According to at least one embodiment described above, a plurality of items are extracted for each piece of collected text information, and when an item, which matches an evaluation target expression or related expression, of the items is found, the evaluation target expression or that associated with the related expression is allocated to the text information as an evaluation target item, a class related to a change in numeric information related to the evaluation target item is distinguished, a class-attached

transaction including an evaluation target item, respective items, and a class is generated for each text information, and a pattern indicating a combination of characteristic items is discovered from a set of transactions. With this arrangement, an evaluation target which cannot be characterized in advance can be handled, time series data are not limited to those related to occurrence of an evaluation target, and a pattern can be automatically discovered.

The method described in each embodiment can also be stored in a storage medium such as a magnetic disk (floppy™ disk, hard disk, or the like), an optical disk (CD-ROM, DVD, or the like), a magneto-optical disk (MO), or a semiconductor memory as a program which can be executed by a computer and distributed.

As the storage medium, any configuration which is a computer-readable storage medium in which a program can be stored may be used regardless of a storage format.

An OS (operating system) which operates on a computer on the basis of an instruction of a program installed from the storage medium in the computer, database management software, and MW (middleware) such as network software may execute a part of the processes to realize the embodiment.

Furthermore, the storage medium according to each embodiment includes not only a medium independent of a computer but also a storage medium in which a program transmitted through a LAN, the Internet, or the like is

downloaded and stored or temporarily stored.

The number of storage media is not limited to one. A case in which the process in each embodiment is executed from a plurality of media is included in the storage medium according to the present invention. Any medium configuration may be used.

A computer according to each embodiment is to execute the processes in each embodiment on the basis of the program stored in a storage medium. The computer may have any configuration such as one apparatus constituted by a personal computer or a system in which a plurality of apparatuses are connected by a network.

A computer in each embodiment includes not only a personal computer but also an arithmetic processing apparatus, a microcomputer, or the like included in an information processing apparatus. The computer is a generic name of an apparatus and a device which can realize the functions of the present invention by a program.

While certain embodiments have been described, these embodiments have been presented by way of example only, and are not intended to limit the scope of the inventions. Indeed, the novel embodiments described herein may be embodied in a variety of other forms; furthermore, various omissions, substitutions and changes in the form of the embodiments described herein may be made without departing from the spirit of the inventions. The accompanying claims and their equivalents are intended to cover such forms or

modifications as would fall within the scope and spirit of the inventions.

C L A I M S

1. An evaluation target of interest extraction apparatus configured to discover a pattern before an evaluation target of interest is extracted based on the pattern from a plurality of pieces of text information which are given along a time series and are related to a plurality of evaluation targets, characterized by comprising:

text collection means (B1) for collecting the plurality of pieces of text information;

numeric collection means (B4) for collecting numeric information individually related to each evaluation target;

target expression storage means (B3) for storing an evaluation target expression which individually expresses each evaluation target, and a related expression which expresses a target individually related to each evaluation target in association with each other;

item allocation means (B2) for extracting a plurality of items for each piece of collected text information, and for, when the items include an item that matches the evaluation target expression or the related expression, allocating the evaluation target expression or an evaluation target expression associated with the related expression in the target expression storage means to the text information as an evaluation target item;

class distinguishing means (B5) for distinguishing a class related to a change in numeric information based on

the numeric information related to an evaluation target expressed by the evaluation target expression as the evaluation target item, and allocating the class to text information to which the evaluation target item is allocated;

transaction generation means (B6) for generating, for each piece of collected text information, a class-attached transaction including the allocated evaluation target item, the extracted items, and the allocated class;

pattern discovery means (B7) for discovering the pattern indicating a combination of characteristic items from a set of the transactions; and

pattern storage means (B8) for storing the discovered pattern.

2. The evaluation target of interest extraction apparatus of claim 1, characterized by further comprising:

means (B6) for generating a classless transaction, which includes the evaluation target item allocated by the item allocation means and the extracted items, and does not include the class for each text information newly collected by the text collection means after the pattern is stored in the pattern storage means;

pattern evaluation means (B10) for applying the pattern in the pattern storage means to the classless transaction to evaluate whether or not the classless transaction includes the pattern; and

evaluation target of interest extraction means (B11)

for, when the classless transaction includes the pattern as a result of the evaluation, incrementing a frequency of occurrence associated with the evaluation target item in the classless transaction, and extracting an evaluation target item associated with the frequency of occurrence as the evaluation target of interest based on the increment result of the frequency of occurrence.

3. A program which is stored in a non-transitory computer-readable storage medium (M) and is executed by a processor of an evaluation target of interest extraction apparatus, which comprises target expression storage means (B3) and pattern storage means (B8) and is configured to discover a pattern before an evaluation target of interest is extracted based on the pattern from a plurality of pieces of text information which are given along a time series and are related to a plurality of evaluation targets, the program characterized by comprising:

a first program code for controlling the processor to execute processing (B1) for collecting the plurality of pieces of text information;

a second program code for controlling the processor to execute processing (B4) for collecting numeric information individually related to each evaluation target;

a third program code for controlling the processor to execute processing for storing an evaluation target expression which individually expresses each evaluation target, and a related expression which expresses a target

individually related to each evaluation target in association with each other in the target expression storage means (B3);

a fourth program code for controlling the processor to execute processing (B2) for extracting a plurality of items for each piece of collected text information, and for, when the items include an item that matches the evaluation target expression or the related expression, allocating the evaluation target expression or an evaluation target expression associated with the related expression in the target expression storage means to the text information as an evaluation target item;

a fifth program code for controlling the processor to execute processing (B5) for distinguishing a class related to a change in numeric information based on the numeric information related to an evaluation target expressed by the evaluation target expression as the evaluation target item, and allocating the class to text information to which the evaluation target item is allocated;

a sixth program code for controlling the processor to execute processing (B6) for generating, for each piece of collected text information, a class-attached transaction including the allocated evaluation target item, the extracted items, and the allocated class; and

a seventh program code for controlling the processor to execute processing (B7) for discovering the pattern indicating a combination of characteristic items from a set

of the transactions, and storing the discovered pattern in the pattern storage means.