Information Modelling and Knowledge Bases XXXV
M. Tropmann-Frick et al. (Eds.)

doi:10.3233/FAIA231162

A Spatio-Temporal and Categorical Correlation Computing Method for Inductive and Deductive Data Analysis

Yasuhiro HAYASHI [a,1], Yasushi KIYOKI [a,2],
Yoshinori HARADA [b], Kazuko MAKINO [b] and Seigo KANEOYA [b]
[a] *Musashino University*
[b] *Credit Saison Co., Ltd*
ORCiD ID: Yasuhiro HAYASHI https://orcid.org/0000-0001-9581-635X

Abstract. This paper proposes a spatio-temporal and categorical correlation computing method for induction and deduction analysis. This method is a data analytics method to reveal spatial, temporal, and categorical relationships between two heterogeneous sets in past events by correlation calculation, thereby finding insights to build new connections between the sets in the future. The most significant feature of this method is that it allows inductive and deductive data analysis by applying context vectors to compute the relationship between the sets whose elements are time, space, and category. Inductive analysis corresponds to data mining, which composes a context vector as a hypothesis to extract meaningful relationships from trends and patterns of past events. Deductive analysis searches past events similar to a context vector's temporal, spatial, and categorical conditions. Spatio-temporal information about the events and information such as frequency, scale, and category are used as parameters for correlation computing. In this method, a multi-dimensional vector space that consists of time, space, and category dimensions is dynamically created, and the data of each set expressed as vectors is mapped onto the space. The similarity degree of the computing shows the strength of relationships between the two sets. This context vector is also mapped onto the space and is calculated distances between the context vector and other vectors of the sets. This paper shows the details of this method and implementation method and assumed applications in commerce activities.

Keywords. Spatio-Temporal & Categorical Correlation Computing, Induction and Deduction Analysis, Dynamic Multi-Dimensional Vector Space Creation, Vector Composition Operator, Context-Based Data Mining

1. Introduction

In real space and online commerce, identifying factors at the connection point between customers and stores can help to increase purchasing opportunities. When analyzing the factors of connection between two sets of data, data analysts try to derive meaningful data by induction, which finds trends or patterns from events, or deduction, which finds events by making hypotheses.

[1] Corresponding Author: Yasuhiro Hayashi, yhayashi@musashino-u.ac.jp.
[2] Corresponding Author: Yasushi Kiyoki, y-kiyoki@musashino-u.ac.jp.

This paper proposes a spatio-temporal and categorical correlation computation method for inductive and deductive analysis. This data analysis method uses correlation calculations to reveal spatial, temporal, and categorical relationships between two heterogeneous sets of past events and thereby find insights to construct new connections between sets in the future. In general data analysis, target data is associated with extrinsic data. For example, combining commercial transaction data and weather data as external data. In this method, we analyze the relationship inherent in the two sets as the data. Concretely, the spatio-temporal and categorical relationships inherent in the sets of customers and stores in a commercial transaction are used.

The most significant feature of this method is that it allows inductive and deductive data analysis by applying context vectors to compute the relationship between the sets whose elements are time, space, and category. Inductive analysis corresponds to data mining, which composes a context vector as a hypothesis to extract meaningful relationships from trends and patterns of past events. Deductive analysis searches past events similar to a context vector's temporal, spatial, and categorical conditions. Spatio-temporal information about the events and information such as frequency, scale, and category are used as parameters for correlation computing. In this method, a multi-dimensional vector space that consists of time, space, and category dimensions is dynamically created, and the data of each set expressed as vectors is mapped onto the space. The similarity degree of the computing shows the strength of relationships between the two sets. This context vector is also mapped onto the space and is calculated distances between the context vector and other vectors of the sets.

(1) Semantic Computing by the Mathematical Model of Meaning and Meta-level System

The Mathematical Model of Meaning and Meta-level System is the core method that inspired this research. The mathematical model of meaning proposed by Kiyoki et al. [1,2] is a method for computing semantic associations between data that change dynamically according to context or situation. An orthonormal space called the metadata space is created, and media data is mapped onto the space. By calculating distances in the metadata space, this method realizes retrieval of media data that are semantically similar to the query. Suppose the context is given along with the query at the time of retrieval. In that case, the dimensionality selection control of the space is dynamically executed, and the retrieval of semantically similar media data is executed according to the context.

Furthermore, the meta-level system proposed by Kiyoki et al. [2,3] is a method that enables the integration and linkage of heterogeneous local database systems by setting up a meta-database system in the upper layer of heterogeneous local database systems. The correlation between temporal, spatial, and semantic features obtained from each local database is weighed by realizing an integrated semantic space and a mechanism for semantic distance calculation in the meta-database system. With the proposed mathematical model of meaning and the meta-level system, Kiyoki et al. aim to realize a memory processing mechanism that interprets dynamically changing meanings and sensitivities depending on context or situation [3].

Our method analyzes data in which two attributes have some relationships. The data is inputted to our method by the table join process on the meta-level of heterogeneous relational databases. Plus, our method dynamically controls dimensions to calculate correlations between two sets. Meta-level system and The Mathematical Model of Meaning are fundamental calculation models of this method.

(2) Image-Query Creation Method

The image-query creation method proposed by Hayashi, Kiyoki, and Chen [4,5] creates image queries for content-based image retrieval by combining images. In this method, an image-query creation database and image-query creation operators are set up in the query part of the content-based image database system. The combination of the image database and the operators is used to operate the color and shape features. Based on the color and shape features of the images that the searcher wants to focus on, this method dynamically controls the dimensions of the image query and the image database to be searched.

In our method, vectors of two sets are created in the integrated space of past events in various fields. Creating the vectors in the image-query creation method and calculating contextual correlation quantities in the orthogonal space of color and shape features is one of the methods that inspired this research.

(3) Emotional MaaS (Mobility as a Service)

The Emotional MaaS, proposed by Kawashima, Hayashi, and Kiyoki [6], is an application of the Mathematical Model of Meaning and Meta-Level System that calculates travel routes and facilities based on the context of tourists. MaaS provides mobility and related services to tourists across the board by highly integrating real space and information space. In this method, the context of the tourist's speed of move, distance in real space, and purpose are set in advance, and transportation and related facilities that are highly correlated with that context are weighed. The intention and situation of each traveler are described as each vector, and the context is defined by composing these vectors.

The correlation calculation and the multi-dimensional vector space creation to describe various contexts are similar to our method in this research. In addition, our approach can be applied to the commerce activity field to clarify human behaviors. The research area overlaps with that of data utilization in mobility information services.

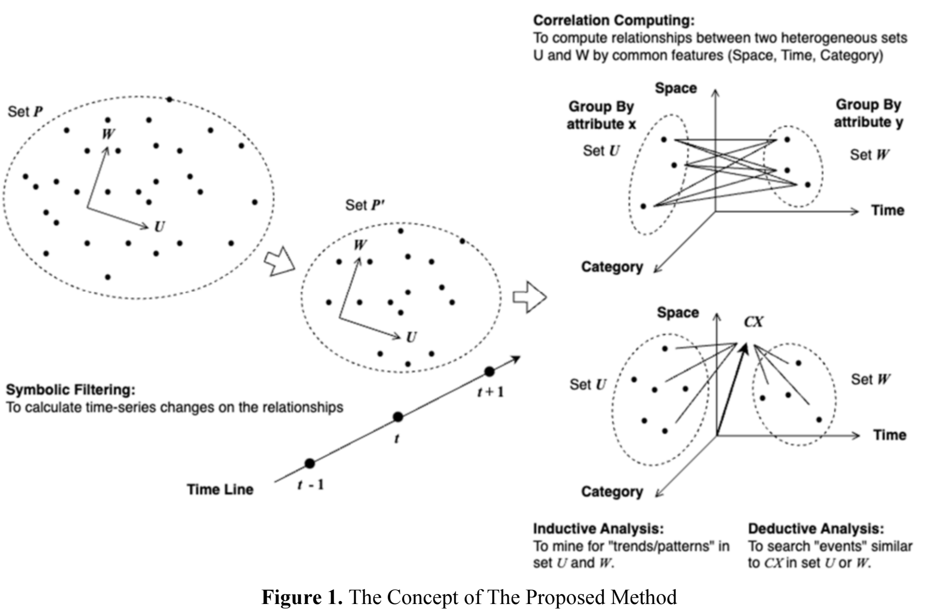

Figure 1. The Concept of The Proposed Method

2. A Spatio-Temporal & Categorical Correlation Computing Method for Induction and Deduction Analysis

2.1. Data Structure & Calculation Method

This method executes spatio-temporal and categorical correlation calculations for induction and deduction analysis. The concept of this model is shown in Figure 1. The concrete calculations are defined as follows:

The given set ***P*** is expressed as Formula 1. The elements of the set ***P*** are expressed as p_{ij}. Where $i := 1 \ldots, q$, q is the number of attributes of the set ***P***. Furthermore, $j := 1 \ldots, r$, r is the number of elements of the set ***P***. Also, the attributes of the set ***P*** are expressed as a_i.

$$P := \begin{array}{cccc} a_1 & a_2 & \cdots & a_q \\ \end{array} \\ P := \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{q1} \\ p_{12} & p_{22} & \cdots & p_{q2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1r} & p_{2r} & \cdots & p_{qr} \end{pmatrix} \tag{1}$$

$$p_{ij} \in P \quad where \quad \{\, p_{ij} \mid i = 1 \ldots q \,,\, j = 1 \ldots r \,\} \tag{2}$$

$$a_i \quad where \quad \{\, a_i \mid i = 1 \ldots q \,\} \tag{3}$$

Based on the given necessary conditions about time, space, and category, selection and projection are executed on the set ***P***. The set ***P*** reduced in number of elements and attributes by this symbolic filtering is expressed as the set ***P′***. The set ***P′*** is assumed to have two attributes, a_x and a_y, that have significant relationships as entities. Where **1 <= *x* < *q*, 1 <= *y* < *q*, *and x not equals y***.

$$a_x \quad where \quad \{\, a_x \mid 1 \leqq x < q \,,\, x \neq y \,\}$$

$$a_y \quad where \quad \{\, a_y \mid 1 \leqq y < q \,,\, x \neq y \,\} \tag{4}$$

Here, the set ***P′*** aggregated by the independent element ***P*[a_x]** in the attribute a_x is defined as a set ***U***.

$$U := P' \; groupby \; a_x = \begin{array}{c} \begin{array}{ccccc} a_x & a_1 & a_2 & \cdots & a_q \end{array} \\ \begin{pmatrix} u_{x1} & u_{11} & u_{21} & \cdots & u_{q1} \\ u_{x2} & u_{12} & u_{22} & \cdots & u_{q2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{xm} & u_{1m} & u_{2m} & \cdots & u_{qm} \end{pmatrix} \end{array} \tag{5}$$

Elements in the set ***U*** is expressed by u_{ij}. Where $i := 1 \ldots, q$, q is the number of attributes in the set ***U***. Furthermore, $j := 1 \ldots, m$, where m is the number of elements in the set ***U***.

$$u_{ij} \in U \quad where \quad \{\, u_{ij} \mid i = 1 \ldots q \,,\, j = 1 \ldots m \,\} \tag{6}$$

When temporal attributes in the set ***U*** are $\boldsymbol{a_t}$, the temporal elements are expressed as $\boldsymbol{u[a_t]}$. Where $\boldsymbol{\{\ a_t \mid t = 1\ \ldots,\ tt\ \}}$ and ***tt*** is an arbitrary number. When spatial attributes in the set ***U*** are $\boldsymbol{a_t}$, the temporal elements are expressed as $\boldsymbol{u[a_t]}$. Where $\boldsymbol{\{\ a_t \mid t = 1\ \ldots,\ ss\ \}}$ and ***ss*** is an arbitrary number. Furthermore, when categorical attributes in the set ***U*** are $\boldsymbol{a_t}$, the temporal elements are expressed as $\boldsymbol{u[a_t]}$. Where $\boldsymbol{\{\ a_t \mid t = 1\ \ldots,\ cc\ \}}$ and ***cc*** is an arbitrary number.

With the temporal feature extraction function ***tf***, the spatial feature extraction function ***sf***, and the categorical feature extraction function ***cf*** defined below, the temporal feature $\boldsymbol{u[a_{xj},\ t]}$, the spatial feature $\boldsymbol{u[a_{xj},\ s]}$, and the categorical feature $\boldsymbol{u[a_{xj},\ c]}$ is calculated as follows:

$$
\begin{aligned}
&u[a_{xj}, t] := tf\,(u[a_{xj}], u[a_t]) \\
&u[a_{xj}, s] := sf\,(u[a_{xj}], u[a_s]) \\
&u[a_{xj}, c] := cf\,(u[a_{xj}], u[a_c])
\end{aligned}
\tag{7}
$$

Note that $\boldsymbol{u[a_x] = p[a_x], j := 1\ldots,\ m}$, where ***m*** is the number of elements in the set ***U***. The spatio-temporal and categorical feature vector $\boldsymbol{v[a_{xj}]}$ of the set $\boldsymbol{u[a_{xj}]}$ is created by this process.

$$
v[a_{xj}] := (u[a_{xj}, t],\ \ u[a_{xj}, s],\ \ u[a_{xj}, c])
\tag{8}
$$

Thus, the set ***P′*** aggregated by the independent element $\boldsymbol{P[a_y]}$ in the attribute $\boldsymbol{a_y}$ is defined by a set ***W***.

$$
W := P'\ groupby\ a_y =
\begin{array}{c}
\begin{array}{cccccc} a_y & a_1 & a_2 & \cdots & a_q \end{array} \\
\begin{pmatrix}
w_{y1} & w_{11} & w_{21} & \cdots & w_{q1} \\
w_{y2} & w_{12} & w_{22} & \cdots & w_{q2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
w_{yn} & w_{1n} & w_{2n} & \cdots & w_{qn}
\end{pmatrix}
\end{array}
\tag{9}
$$

Elements in the set ***W*** are expressed as $\boldsymbol{w_{ij}}$. Where $\boldsymbol{i := 1\ \ldots,\ q}$, ***q*** is the number of attributes of the set ***W***. Plus, $\boldsymbol{j := 1\ \ldots,\ n}$, ***n*** is the number of elements in the set ***W***.

$$
w_{ij} \in W \quad where \quad \{\, w_{ij} \mid i = 1 \ldots\, q\, ,\ j = 1 \ldots\, n\, \}
\tag{10}
$$

When temporal attributes in the set ***W*** is $\boldsymbol{a_t}$, the temporal elements are expressed by $\boldsymbol{w[a_t]}$. Where $\boldsymbol{\{\ a_t \mid t = 1\ \ldots,\ tt\ \}}$ and ***tt*** is an arbitrary number. When spatial attributes in the set ***W*** are $\boldsymbol{a_t}$, the temporal elements are expressed as $\boldsymbol{w[a_t]}$. Where $\boldsymbol{\{\ a_t \mid t = 1\ \ldots,\ ss\ \}}$ and ***ss*** is an arbitrary number. Furthermore, when categorical attributes in the set ***W*** are $\boldsymbol{a_t}$, the temporal elements are expressed as $\boldsymbol{w[a_t]}$. Where $\boldsymbol{\{\ a_t \mid t = 1\ \ldots,\ cc\ \}}$ and ***cc*** is an arbitrary number.

With the temporal feature extraction function ***tf***, the spatial feature extraction function ***sf***, and the categorical feature extraction function ***cf*** defined below, the temporal feature $\boldsymbol{w[a_{yj},\ t]}$, the spatial feature $\boldsymbol{w[a_{yj},\ s]}$, and the categorical feature $\boldsymbol{w[a_{yj},\ c]}$ are calculated as follows:

$$w[a_{yj}, t] := tf\,(w[a_{yj}], w[a_t])$$
$$w[a_{yj}, s] := sf\,(w[a_{yj}], w[a_s])$$
$$w[a_{yj}, c] := cf\,(w[a_{yj}], w[a_c]) \quad (11)$$

Note that $\boldsymbol{w[a_y] = p[a_y]}$, $\boldsymbol{j := 1\ldots, n}$, where $\boldsymbol{n}$ is the number of elements in the set $\boldsymbol{W}$. The spatio-temporal and categorical feature vector $\boldsymbol{v[a_{yj}]}$ of the set $\boldsymbol{w[a_{yj}]}$ is created by this process.

$$v[a_{yj}] := (w[a_{yj}, t],\ \ w[a_{yj}, s],\ \ w[a_{yj}, c]) \quad (12)$$

The vectors $\boldsymbol{v[a_{xj}]}$ and $\boldsymbol{v[a_{yj}]}$ created by Equations 8 and 12 are mapped to a multi-dimensional vector space $\boldsymbol{V}$ with time, space, and category as dimensions. The distance between vectors $\boldsymbol{d_t}$, $\boldsymbol{d_s}$, ***and*** $\boldsymbol{d_c}$ is calculated for each temporal, spatial, and categorical feature by the temporal feature distance function ***td***, spatial feature distance function ***sd***, and categorical feature distance function ***cd*** defined by Formula 13. Plus, to calculate the total correlation ***score*** between mapped vectors $\boldsymbol{v[a_{xj}]}$ and $\boldsymbol{v[a_{yj}]}$, the similarities $\boldsymbol{d_t}$, $\boldsymbol{d_s}$, $\boldsymbol{d_c}$ calculated in different methods are normalized and expressed as $\boldsymbol{d_t'}$, $\boldsymbol{d_s'}$, $\boldsymbol{d_c'}$ (Formula 14).

$$d_t := td\,(v[a_{xj}], v[a_{yj}])$$
$$d_s := sd\,(v[a_{xj}], v[a_{yj}])$$
$$d_c := cd\,(v[a_{xj}], v[a_{yj}]) \quad (13)$$

$$d_t' := norm(d_t)$$
$$d_s' := norm(d_s)$$
$$d_c' := norm(d_c) \quad (14)$$

The result of each normalized distance calculation is multiplied by the weights $\boldsymbol{wt_t}$, $\boldsymbol{wt_s}$, $\boldsymbol{wt_c}$ and calculated as a sum value ***score***.

$$score := sim\,(v[a_{xj}], v[a_{yj}]) = wt_t \times d_t' + wt_s \times d_s' + wt_c \times d_c' \quad (15)$$

2.2. Context Vector for Inductive and Deductive Data Analysis

This method's originality is applying context vector ***CX*** to achieve temporal, spatial, categorical, inductive, and deductive data analysis. The context vector ***CX***, consisting of the temporal feature $\boldsymbol{cx_t}$, the spatial feature $\boldsymbol{cx_s}$, and the category feature $\boldsymbol{cx_c}$, is expressed as follows.

$$CX := (cx_t, cx_s, cx_c) \quad (16)$$

Using Formula 15, the similarity calculation between the elements $\boldsymbol{v[a_{xj}]}$ of the set $\boldsymbol{U}$ and the context vector ***CX,*** or the elements $\boldsymbol{v[a_{yj}]}$ of the set $\boldsymbol{W}$ and the context vector ***CX***, enables an inductive or deductive data analysis approach.

$$score := sim\left(v[a_{xj}], CX\right)$$
$$score := sim\left(v[a_{yj}], CX\right) \tag{17}$$

Inductive analysis in this method corresponds to data mining, in which hypotheses are derived from elements of a set ***U*** or set ***W*** based on trends or patterns. Hypotheses are expressed as context vectors ***CX*** whose elements are temporal, spatial, and categorical trends or patterns. Statistics or difference calculations set the value of the context vector ***CX***. Alternatively, when a set of context vectors ***CXs*** consisting of multiple context vectors ***CXs*** is set in advance, appropriate hypotheses are extracted by iteratively calculating the distance between the elements of the set ***U*** or ***W*** and the set of context vectors ***CXs***.

Deductive analysis in this method corresponds to searching a set ***U*** or ***W*** for elements similar to a particular model event's temporal, spatial, and categorical features, as indicated by the context vector ***CX***.

2.3. Spatio-Temporal and Categorical Feature Extraction Functions

2.3.1. Temporal Feature Extraction Function

This function extracts temporal average and variance. The date in the purchase history data has two facets: the usage date and the usage date interval. The extracted temporal feature vector V_T consists of the average ***ta*** of the usage dates in the ***n*** data in the purchase history, its variance ***tv***, the average ***tia*** of the usage date interval, and its variance ***tiv***. Note that the usage interval date is not constant.

$$V_T = (ta, tv, tia, tiv) \tag{18}$$

When each usage date data in the purchase history is expressed as p_i, the average ***ta*** of usage dates in ***n*** data is calculated as follows. Where, ***i := 1, ..., n***. The ***dtoi*** function converts a Gregorian date to a standard date integer. The ***itod*** function converts an integer value of a date in a standard format to a date in Gregorian format. The variance ***tv*** of the date of use in ***n*** data is calculated by Formula 20. Note that absolute values are used to simplify the calculation.

$$ta := itod\left(\frac{1}{n}\sum_{i=1}^{n} dtoi(p_i)\right) \tag{19}$$

$$tv := itod\left(\frac{1}{n}\sum_{i=1}^{n}\left|dtoi(p_i) - \frac{1}{n}\sum_{i=1}^{n} dtoi(p_i)\right|\right) \tag{20}$$

When each usage date data in the purchase history is expressed as p_i, the average ***tia*** of usage date intervals for ***n*** data is calculated as follows. Where ***i := 1, ..., n***. When ***n*>0**, the average ***tia*** and its variance ***tiv*** are obtained. When ***n*=0**, ***tia*** and ***tiv*** are zero. Additionally, the variance ***tiv*** of the usage date interval in ***n*** data is calculated as follows.

$$tia := itod\left(\frac{1}{n}\sum_{i=1}^{n}\left(dtoi(p_{i+1}) - dtoi(p_i)\right)\right) \tag{21}$$

$$tiv := itod\left(\frac{1}{n}\sum_{i=1}^{n}\left|dtoi(p_i) - \frac{1}{n}\sum_{i=1}^{n}(dtoi(p_{i+1}) - dtoi(p_i))\right|\right) \tag{22}$$

2.3.2. Spatial Feature Extraction Function

This function calculates spatial features. The spatial feature vector V_S consists of the center position ***sa*** and variance ***sv*** of latitude and longitude of stores included in the ***n*** data in the purchase history.

$$V_S = (sa, sv) \tag{23}$$

The latitude and longitude ***sa*** corresponding to the center of gravity of the latitude and longitude in the ***n*** data is calculated by Formula 24. When the latitude and longitude of two points ***p, q*** on the earth are given as ***p(latitude1, longitude1), q(latitude2, longitude2)***, the great circle distance ***sd*** between ***p*** and ***q*** is calculated by Formula 25. This formula is also used for spatial similarity degree.

$$\begin{aligned}
x &:= \frac{1}{n}\sum_{i=1}^{n}\cos\left(\frac{latitude[i]\ \pi}{180}\right) \times \cos\left(\frac{longitude[i]\ \pi}{180}\right)\\
y &:= \frac{1}{n}\sum_{i=1}^{n}\cos\left(\frac{latitude[i]\ \pi}{180}\right) \times \sin\left(\frac{longitude[i]\ \pi}{180}\right)\\
z &:= \frac{1}{n}\sum_{i=1}^{n}\sin\left(\frac{latitude[i]\ \pi}{180}\right)\\
sa &:= \left(\frac{180 \cdot atan2(z, \sqrt{x^2+y^2})}{\pi}, \frac{180 \cdot atan2(y, x)}{\pi}\right)
\end{aligned} \tag{24}$$

$$\begin{aligned}
&r := 6378.137\ [km]\\
&x1 := r\cos\left(\frac{latitude1\ \pi}{180}\right) * \cos\left(\frac{longitude1\ \pi}{180}\right) \qquad x2 := r\cos\left(\frac{latitude2\ \pi}{180}\right) * \cos\left(\frac{longitude2\ \pi}{180}\right)\\
&y1 := r\cos\left(\frac{latitude1\ \pi}{180}\right) * \sin\left(\frac{longitude1\ \pi}{180}\right) \qquad y2 := r\cos\left(\frac{latitude2\ \pi}{180}\right) * \sin\left(\frac{longitude2\ \pi}{180}\right)\\
&z1 := r\sin\left(\frac{latitude1\ \pi}{180}\right) \qquad z2 := r\sin\left(\frac{latitude2\ \pi}{180}\right)\\
&sd := 2r\ asin\left(\frac{\sqrt{(x1-x2)^2+(y1-y2)^2+(z1-z2)^2}}{2r}\right)
\end{aligned} \tag{25}$$

When each latitude and longitude data in the purchase history is expressed as p_i, the variance ***sv*** of latitude and longitude in ***n*** data is calculated as follows. Where ***i* := 1, ..., *n***. ***sd*** is a function to calculate the great distance between p_i and the center of gravity ***sa*** of latitude and longitude in ***n*** data.

$$sv := \frac{1}{n}\sum_{i=1}^{n}|sd(p_i, sa)| \tag{26}$$

2.3.3. Categorical Feature Extraction Function

The Category Feature Histogram V_C is the sum of n stores' category vector data C_i in the purchase history. Where ***i* := 1, ..., *k***. The category is expressed as tree data consisting of four levels: large, medium, small, and detailed. By converting the tree data format to vector data format, distance calculation in the vector space can be applied. When ***L*** major category, ***M*** medium category, ***S*** minor category, and ***D*** detailed category consist of ***l, m,***

s, and d elements, respectively, tree data $\boldsymbol{T}$ is converted to vector data $\boldsymbol{C_i}$ consisting of $\boldsymbol{k := l + m + s + d}$ elements.

$$k := l + m + s + d$$
$$C_i := (c_{i1}, c_{i2}, \ldots, c_{ik})$$
$$V_C := \sum_{i=1}^{k} C_i = \sum_{i=1}^{k} (c_{i1}, c_{i2}, \ldots, c_{ik}) \tag{27}$$

2.3.4. Vector Composition Function

Given $\boldsymbol{k}$ vectors $\boldsymbol{V_i := (v_{i1}, v_{i2}, \ldots, v_{in})}$ consisting of $\boldsymbol{n}$-dimensions, this function that is named Vector Creation Operator composes a new vector by computing the sum of the same elements of those vectors. Where $\boldsymbol{2 <= i <= k}$.

$$V_{composition} := composition\,(V_1, V_2, \cdots V_n) = \sum_{i=1}^{k} (v_{i1}, v_{i2}, \cdots v_{in}) \tag{28}$$

2.3.5. Distance Calculation Function

The inner product $\boldsymbol{ip}$ is used to calculate the semantic distance as similarity ***similarity*** between two context vectors $\boldsymbol{V_a, V_b}$ created by this method. Where $\boldsymbol{D := 2{+}4{+}k}$. Note that if semantical distance is required between elements of two vectors, Euclidean distance calculation and geographical distance calculation are also used.

$$V_a := (v_{a1}, v_{a2}, \cdots, v_{ad})$$
$$V_b := (v_{b1}, v_{b2}, \cdots, v_{bd})$$
$$similarity := ip\,(V_a, V_b) = \sum_{i=1}^{d} (v_{ai} \times v_{bi}) \tag{29}$$

3. Assumed Implementations and Applications

This method deductively and inductively analyzes spatio-temporal and categorical associations between entities related to each other in a given set. By setting up this method as a storied function in a commercial transaction database with many commercial transactions and as a related tool for data analysts, it is possible to reveal the spatiotemporal and categorical relationships between customers and stores. The related tool should have a user interface to accept the needs of the data analyst who repeatedly tests the hypothesis, queries the database as SQL, and visualizes the results on the screen. The data to be calculated are payment systems, such as credit cards, used in real space and online. Based on information obtained from the payment system, such as the date of purchase, customer information, name of the purchasing store, and type of business, a customer and a store set are formed. The following practical applications can be assumed from these sets to be formed.

1. To clarify the purchasing habits and tendencies of all customers or specific customers in a spatio-temporal and categorical context.

2. To clarify the spatiotemporal and categorical correlations between stores used by specific customers based on a particular purchasing context.
3. To clarify the spatiotemporal and categorical purchasing habits and tendencies of specific customers who use specific stores.

Based on the calculation of meaningful correlations between customers and stores obtained through a series of analytical processes, new business marketing is expected to enhance the Well-Being and happiness of each customer and each store.

4. Conclusion

This paper described a spatio-temporal and categorical correlation computing method for induction and deduction analysis. This method's originality applies context vector ***CX*** to achieve temporal, spatial, and categorical correlation calculation between two heterogeneous sets as inductive, and deductive data analysis. Inductive analysis corresponds to data mining, which composes a context vector as a hypothesis to extract meaningful relationships from trends and patterns of past events. Deductive analysis searches past events similar to a context vector's temporal, spatial, and categorical conditions. By repeating the analysis, insights will be found to build new connections between sets in the future. This paper also described the proposed method's assumed implementation and applications in commercial activities.

As the next step, we will develop a proto-type system that applies the proposed method and experiments to evaluate effectiveness and feasibility, and business deployment.

References

[1] Yasushi Kiyoki, Xing Chen, "A Semantic Associative Computation Method for Automatic Decorative-Multimedia Creation with "Kansei" Information" (Invited Paper), The Sixth Asia-Pacific Conferences on Conceptual Modelling (APCCM 2009), 9 pages, January 20-23, 2009.
[2] Yasushi Kiyoki and Saeko Ishihara: "A Semantic Search Space Integration Method for Meta-level Knowledge Acquisition from Heterogeneous Databases," Information Modeling and Knowledge Bases (IOS Press), Vol. 14, pp.86-103, May 2002.
[3] Kiyoki, Y., Chen, X., Veesommai, C., Wijitdechakul, J., Sasaki, S., Koopipat, C., & Chawakitchareon, P.: "A semantic-associative computing system with multi-dimensional world map for ocean-environment analysis", Information Modelling and Knowledge Bases XXX, pp. 147-168.
[4] Hayashi, Y., Kiyoki, Y., and Chen, X.: "An Image-Query Creation Method for Expressing User's Intentions by Combining Multiple Images", Information Modelling and Knowledge Bases, Vol.XXI, IOS Press, pp. 188-207, 2010.
[5] Hayashi, Y., Kiyoki, Y., and Chen, X.: "A Combined Image-Query Creation Method for Expressing User's Intentions with Shape and Color Features in Multiple Digital Images", Information Modelling and Knowledge Bases, Vol. XXII, IOS Press, pp. 258-277, 2011.
[6] Kawashima, K., Hayashi, Y., Kiyoki, Y., Mita., T.: "A Mobility and Activity Integration System Supporting Sensitivity to Contexts in Dynamic Routing - Emotional MaaS -", Information Modelling and Knowledge Bases, Vol. XXXIII, IOS Press, pp. 297-308, 2021.
[7] UN SDGs-3, Ensure healthy lives and promote well-being for all at all ages, https://sdgs.un.org/goals/goal3, 2023/01/28.
[8] Tal Ben-Shahar, "Even Happier: A Gratitude Journal for Daily Joy and Lasting Fulfillment," McGraw Hill, 2009.
[9] Kahneman, D, and A Deaton. 2010. "High income improves evaluation of life but not emotional well-being." Proceedings of the National Academy of Sciences 107 (38): 16489-16493.