

The Discovery of Logical Propositions in Numerical Data

Hiroshi Tsukimoto

Systems & Software Engineering Laboratory, Research & Development Center, Toshiba Corporation, 70, Yanagi-cho, Saiwai-ku, Kawasaki 210, Japan

Abstract: This paper presents a method to discover logical propositions in numerical data. The method is based on the space of multi-linear functions, which is made into a Euclidean space. A function obtained by multiple regression analysis in which data are normalized to $[0,1]$ belongs to this Euclidean space. Therefore, the function represents a non-classical logical proposition and it can be approximated by a Boolean function representing a classical logical proposition. We prove that this approximation method is a pseudo maximum likelihood method using the principle of indifference. We also experimentally confirm that correct logical propositions can be obtained by this method. This method will be applied to the discovery of logical propositions in numerical data.

1 Introduction

Statisticians deal with numerical data, but they do not discover logical propositions in the numerical data. In the AI field, several researchers have studied the discovery of scientific laws [Langley 81]. This paper presents a method to discover logical propositions in numerical data. The method is based on the space of multi-linear functions, which is made into a Euclidean space. A function obtained by multiple regression analysis in which data are normalized to $[0,1]$ belongs to this Euclidean space. Therefore, the function represents a non-classical logical proposition and it can be approximated by a Boolean function representing a classical logical proposition. We prove that this approximation method is a pseudo maximum likelihood method using the principle of indifference [Keynes 21]. The algorithm is as follows.

1. The numerical data are normalized to $[0,1]$.
2. Multiple regression analysis is performed.
3. The linear function obtained by multiple regression analysis is approximated by a Boolean function.
4. The Boolean function is reduced to the minimum one.

We experimentally confirm that correct logical propositions can be obtained by this method. This method will be applied to the discovery of logical propositions in numerical data. Section 2 explains the space of multi-linear functions as an extended model for logics. Section 3 explains the discovery of logical propositions in numerical data.

Some readers may think that the problem in this paper is a classification problem with continuous classes and the problem can be dealt with by some other algorithms such as C4.5. However, C4.5 does not work well for the classification problem with continuous classes [Quinlan 93].

2 Multi-linear functions—A model for logics

Due to the space limitation, we briefly explain an extended model for logics, where logical functions including non-classical logical functions are represented as points (vectors) in a Euclidean space. More detailed explanation including proofs can be found in [Tsukimoto 94a]. Hereinafter, let F, G, \dots stand for propositions, f, g, \dots stand for functions, X, Y, \dots stand for propositional variables and x, y, \dots stand for variables.

2.1 Intuitive explanation

We explain why a logical function is represented as a vector. It is worth noticing that classical logic has properties similar to a vector space. These properties are seen in Boolean algebra with atoms which is a model for classical logic. Atoms in Boolean algebra have the following properties:

1. $a_i \cdot a_i = a_i$ (unitarity).
2. $a_i \cdot a_j = 0$ ($i \neq j$) (orthogonality).
3. $\sum a_i = 1$ (completeness).

For example, for the proposition $\overline{X} \vee Y = XY \vee \overline{X}Y \vee \overline{X}\overline{Y}$, $\overline{X} \vee Y$ is represented as $(1, 0, 1, 1)$, where $XY = (1, 0, 0, 0)$, $X\overline{Y} = (0, 1, 0, 0)$, $\overline{X}Y = (0, 0, 1, 0)$, $\overline{X}\overline{Y} = (0, 0, 0, 1)$. Atoms in Boolean algebra correspond to unit vectors. In other words, atoms in Boolean algebra are similar to the orthonormal functions in a Euclidean space.

This paper shows that the space of logical functions actually becomes a Euclidean space. The process from Boolean algebra to Euclidean space is divided into three stages.

1. Represent Boolean algebra by elementary algebra. In other words, present an elementary algebra model for classical logic.
2. Expand the domain from $\{0,1\}$ to $[0,1]$ and the model to the space of multilinear functions.
3. Introduce an inner product to the above space and construct a Euclidean space where logical functions are represented as vectors.

2.2 An elementary algebra model for classical logic

2.2.1 Definitions

(1) Definition of τ

Let $f(x)$ be a real polynomial function. Consider the following formula:

$$f(x) = p(x)(x - x^2) + q(x),$$

where $f : \{0,1\} \rightarrow R$ and $q(x) = ax + b$, where a and b are real. τ_x is defined as follows:

$$\tau_x : f(x) \rightarrow q(x).$$

The above definition implies the following property:

$$\tau_x(x^n) = x.$$

In the case of n variables, τ is defined as follows:

$$\tau = \prod_{i=1}^n \tau_{x_i}.$$

For example, $\tau(x^2y^3 + y + 1) = xy + y + 1$.

(2) Definition of L

Let L be the set of all functions satisfying $\tau(f) = f$. Then $L = \{f : \tau(f) = f\}$.

In the case of two variables, $L = \{axy + bx + cy + d | a, b, c, d \in \mathbf{R}\}$.

(3) Definition of L_1

L_1 is inductively defined as follows:

1. Variables are in L_1 .
2. If f and g are in L_1 , then $\tau(f \cdot g)$, $\tau(f + g - f \cdot g)$ and $\tau(1 - f)$ are in L_1 . (We call these three calculations τ calculation.)
3. L_1 consists of all functions finitely generated by the (repeated) use of 1. and 2.

2.2.2 A new model for classical logic

Let the correspondence between Boolean algebra and τ calculation be as follows:

1. $F \wedge G \Leftrightarrow \tau(fg)$.
2. $F \vee G \Leftrightarrow \tau(f + g - fg)$.
3. $\bar{F} \Leftrightarrow \tau(1 - f)$.

Then $(L_1, \tau$ calculation) is a model for classical logic; that is, L_1 and τ calculation satisfy the axioms for Boolean algebra. The proof is omitted due to space limitations. We give a simple example. $(X \vee Y) \wedge (X \vee Y) = X \vee Y$ is calculated as follows:

$$\begin{aligned}\tau((x + y - xy)(x + y - xy)) &= \tau(x^2 + y^2 + x^2y^2 + 2xy - 2x^2y - 2xy^2) \\ &= x + y + xy + 2xy - 2xy - 2xy \\ &= x + y - xy.\end{aligned}$$

2.3 Extension of the model

First, the domain is extended to $[0,1]$. $f : [0,1]^n \rightarrow \mathbf{R}$. By this extension, we have continuously-valued logic functions, which satisfy all axioms of classical logic. The proof can be found in [Tsukimoto 94b]. Hereinafter, not only in the case of $\{0,1\}$, but also in the case of $[0,1]$, the functions satisfying all axioms of classical logic are called Boolean functions.

The model is extended from L_1 to L which was defined in 2.2.1. $f : \{0,1\}^n \rightarrow \mathbf{R}$ or $f : [0,1]^n \rightarrow \mathbf{R}$. L obviously includes Boolean functions and linear functions. L is called the space of multi-linear functions. In the case of $\{0,1\}$, the space is the same as in [Linial 93]. Hereinafter, L will be made into a Euclidean space (a finite dimensional inner product space).

2.4 Euclidean space

2.4.1 Definition of inner product and norm

In the case of $[0,1]$, an inner product is defined as follows:

$$\langle f, g \rangle = 2^n \int_0^1 \tau(fg) dx,$$

where f and g are in L , and the integral is generally a multiple integral.

In the case of $\{0,1\}$, an inner product is defined as

$$\langle f, g \rangle = \sum \tau(fg),$$

where this sum spans the whole domain. For example, in the case of two variables, let $f = f(x, y)$ and $g = g(x, y)$, then $\langle f, g \rangle = (\tau(fg))(1, 1) + (\tau(fg))(1, 0) + (\tau(fg))(0, 1) + (\tau(fg))(0, 0)$, where $(\tau(fg))(1, 1)$ is the value of $\tau(fg)$ at $x = 1$ and $y = 1$.

A norm is defined as

$$N(f) = \sqrt{\langle f, f \rangle}.$$

L becomes an inner product space with the above norm. The dimension of this space is finite, because L consists of the multi-linear functions of n variables, where n is finite. Therefore, L becomes a finite dimensional inner product space, namely a Euclidean space.

2.4.2 Orthonormal system

The orthonormal system is as follows:

$$\phi_i = \prod_{j=1}^n e(x_j) \quad (i = 1, \dots, 2^n, j = 1, \dots, n),$$

where $e(x_j) = 1 - x_j$ or x_j .

It is easily understood that these orthonormal functions are the expansion of atoms in Boolean algebra. In addition, it can easily be verified that the orthonormal system satisfies the following properties:

$$\langle \phi_i, \phi_j \rangle = \begin{cases} 0 & (i \neq j), \\ 1 & (i = j), \end{cases}$$

$$f = \sum_{i=1}^{2^n} \langle f, \phi_i \rangle \phi_i.$$

For example, the representation by orthonormal functions of $x + y - xy$ of two variables (dimension 4) is as follows:

$$f = 1 \cdot xy + 1 \cdot x(1 - y) + 1 \cdot (1 - x)y + 0 \cdot (1 - x)(1 - y).$$

and the vector representation is $(1, 1, 1, 0)$, where the bases are $xy = (1, 0, 0, 0)$, $x(1 - y) = (0, 1, 0, 0)$, $(1 - x)y = (0, 0, 1, 0)$ and $(1 - x)(1 - y) = (0, 0, 0, 1)$. The space of the functions of n variables is 2^n -dimensional. The vector representation of Boolean functions is the same as the representation expanded by atoms in Boolean algebra. This vector representation is called a logical vector.

3 Acquiring logical propositions from the functions obtained by multiple regression analysis

3.1 The outline of the method

The method is to approximate the (multi-)linear functions by Boolean functions. Approximating a linear function by a Boolean function means obtaining the nearest Boolean function with respect to the Euclidean norm. Let (a_i) be the logical vector which represents the linear function obtained by multiple regression analysis. Let $(b_i)(b_i = 0, 1)$ be the logical vector which represents the nearest Boolean function. The nearest Boolean function minimizes $\sum(a_i - b_i)^2$. Each term can be minimized independently and $b_i = 1$ or 0 . Therefore, the approximation method is as follows:

If $a_i \geq 0.5$, then $b_i = 1$, otherwise $b_i = 0$.

For example, let $z = 0.6x - 1.1y + 0.3$ be obtained by multiple regression analysis. The function is transformed to

$$z = -0.2xy + 0.9x(1 - y) - 0.8(1 - x)y + 0.3(1 - x)(1 - y),$$

that is, the logical vector is $(-0.2, 0.9, -0.8, 0.3)$. By the above method, the logical vector is approximated to $(0, 1, 0, 0)$, which represents $x(1 - y)$. Thus, $0.6x - 1.1y + 0.3$ is approximated to $x(1 - y)$, that is, $X \wedge \bar{Y}$.

3.2 Pseudo maximum likelihood method

The approximation can be regarded as a pseudo maximum likelihood method using the principle of indifference [Keynes 21]. Maximum likelihood method is explained in many books, for example, [Wilks 62]. The principle of indifference says that a probability distribution is uniform when we have no information.

3.2.1 A relation between logic and probability

The above approximation method is based on the norm of logical vectors, while maximum likelihood method is based on probability. The relation between the norm of logical vectors and probability must be studied. First, the amount of information of a logical function H is defined as follows:

$$H(f) = n - \log_2(N(f))^2,$$

where n is the number of variables. It can be verified that H is equal to I , which is the amount of information of probability using the principle of indifference [Keynes

21] in the case of classical logic, where $I = n - H_e$, ($H_e = -\sum_1^{2^n} p_i \log_2 p_i$) [Shannon 49]. The proof is omitted due to the space limitation. We also use this formula in non-classical logics. The above formula is transformed to the one below, which is used to prove that the approximation method is a pseudo maximum likelihood method.

$$N(f) = 2^{H_e/2}$$

The proof is as follows:

$$\begin{aligned} H(f) &= I \\ \rightarrow \log_2(N(f))^2 &= H_e \\ \rightarrow (N(f))^2 &= 2^{H_e} \\ \rightarrow N(f) &= 2^{H_e/2}. \end{aligned}$$

3.2.2 Pseudo maximum likelihood method

Assume that a logical vector \mathbf{f} be near to a Boolean vector \mathbf{g} , that is,

$$f_i \simeq g_i, (i = 1, \dots, n)$$

The approximation is

$$|\mathbf{f} - \mathbf{g}| \rightarrow \min,$$

since $f_i \simeq g_i, (i = 1, \dots, n)$, the above formula is transformed to

$$\|\mathbf{f}\| - \|\mathbf{g}\| \rightarrow \min.$$

Since $\|\mathbf{f}\|$ and $\|\mathbf{g}\|$ are $2^{H_e/2}$ and $2^{H'_e/2}$ respectively from the discussion in the preceding subsection, the formula is transformed to

$$|2^{H_e/2} - 2^{H'_e/2}| \rightarrow \min,$$

$$|H_e - H'_e| \rightarrow \min.$$

Assume $H'_e > H_e$, then the formula is

$$H'_e - H_e \rightarrow \min.$$

Since $H_e = -\sum_1^{2^n} (p_i \log_2 p_i)$ and $H'_e = -\sum_1^{2^n} (q_i \log_2 q_i)$, the formula is

$$\sum_1^{2^n} (p_i \log_2 p_i) - \sum_1^{2^n} (q_i \log_2 q_i) \rightarrow \min.$$

Since $f_i \simeq g_i, (i = 1, \dots, n)$, that is, $p_i \simeq q_i, (i = 1, \dots, n)$, and a logarithmic function is rapidly descending in the domain $[0,1]$, the formula is transformed to

$$\sum_1^{2^n} (p_i \log_2 p_i) - \sum_1^{2^n} (p_i \log_2 q_i) \rightarrow \min.$$

This value is K-L(Kullback-Leibler) information. Thus, the approximation method has been proved to be a pseudo maximum likelihood method.

3.3 Procedures

Hereinafter, multi-linear functions are limited to linear functions for simplification. The procedures are as follows.

1. Normalize data to $[0,1]$.

There are a few methods for normalization. In this paper, we use the following method:

Let a stand for the maximum data and let b stand for the minimum data.

$$y = (x - b)/(a - b),$$

where x is a given data and y is the normalized data.

2. Execute multiple regression analysis.

A linear function is obtained by the usual multiple regression analysis.

3. Orthonormal expansion

The linear function is expanded by the orthonormal functions.

4. Approximate it by a Boolean function.

A Boolean function is obtained by the approximation using the method in the preceding subsection.

5. Transform it to a logical proposition.

The Boolean function is transformed to a logical proposition.

6. Reduce it to the minimum one.

The proposition is reduced to the minimum one.

3.4 An example

Table 1 shows how a property of a metal depends on time and temperature.

1. Normalization

Table 1 is normalized to Table 2.

Table 1: A property of a metal

Sample number	Temperature(X)	Time(Y)	Property(Z)
1	1700	30	36
2	1800	25	39
3	1800	20	44
4	1850	30	44
5	1900	10	59
6	1930	10	51

Table 2: A property of a metal (normalized)

Sample number	Temperature(X)	Time(Y)	Property(Z)
1	0.000	1.00	0.000
2	0.435	0.75	0.130
3	0.435	0.50	0.348
4	0.652	1.00	0.348
5	0.87	0.00	1.000
6	1.000	0.00	0.652

2. Multiple regression analysis

$z = 0.46x - 0.41y + 0.38$ is obtained by multiple regression analysis.

3. Orthonormal expansion

The orthonormal expansion of the above function is

$$z = 0.46x - 0.41y + 0.38 = 0.43xy + 0.84x(1-y) - 0.03(1-x)y + 0.38(1-x)(1-y).$$

4. Approximation

The following Boolean function is obtained by the approximation.

$$0.0xy + 1.0x(1 - y) - 0.0(1 - x)y + 0.0(1 - x)(1 - y).$$

5. Logical proposition

A logical proposition $X \wedge \bar{Y}$ is obtained.

6. Reduction

Reduction is not necessary in this case.

As a result, a proposition $X \wedge \bar{Y}$ has been obtained. If the domain is $[0,1]$, this proposition can be interpreted as "If temperature gets higher and time gets shorter, the property gets higher." If the domain is $\{0, 1\}$, the proposition can be interpreted as "If temperature is high and time is short, the property is high."

3.5 An experiment

Neural networks and some electrical circuits consisting of nonlinear elements are the examples whose structure is logical or qualitative and only numerical data of which are observed. Consider the following function:

$$w(x, y, z) = P((f(x) \vee \bar{g}(y)) \wedge h(z)),$$

where $f(x)$, $g(y)$ and $h(z)$ are as follows:

$$f(x) = 1(x \geq 0.5), = 0(x < 0.5),$$

$$g(y) = 1(y \geq 0.6), = 0(y < 0.6),$$

$$h(z) = 1(z \geq 0.4), = 0(z < 0.4).$$

$P(t)$ is a probabilistic function such that if $t = 1$, then $P(t)$ is a value in $[0.5, 1]$ and if $t = 0$, then $P(t)$ is a value in $[0, 0.5]$, where $0 \leq x, y, z, w \leq 1$. Notice that the threshold values of f, g and h are different.

The purpose of the experiment is to check if we can obtain the correct logical proposition $W = (X \vee \bar{Y}) \wedge Z$ from some appropriate data. Let the data in the table below be given.

$w = 0.2x - 0.2y + 0.4z + 0.2$ is obtained by multiple regression analysis and the orthonormal expansion is

$$0.6xyz + 0.2xy(1 - z) + 0.8x(1 - y)z + 0.4x(1 - y)(1 - z) + 0.4(1 - x)yz + 0.0(1 - x)y(1 - z) + 0.6(1 - x)(1 - y)z + 0.2(1 - x)(1 - y)(1 - z).$$

The nearest Boolean function is

Table 3: Experiment

x	y	z	w
0.9	0.6	0.6	0.5
0.7	0.9	0.1	0.2
1.0	0.5	1.0	0.7
0.6	0.5	0.2	0.3

$1.0xyz + 0.0xy(1 - z) + 1.0x(1 - y)z + 0.0x(1 - y)(1 - z) + 0.0(1 - x)yz + 0.0(1 - x)y(1 - z) + 1.0(1 - x)(1 - y)z + 0.0(1 - x)(1 - y)(1 - z)$.

The following proposition is obtained.

$$XYZ \vee X\bar{Y}Z \vee \bar{X}\bar{Y}Z = (X \vee \bar{Y}) \wedge Z$$

Now we have the correct logical proposition, which cannot be seen in the given numerical data and the linear function obtained by multiple regression analysis. Whether the correct logical proposition can be obtained or not depends on the data and the logical proposition. That is, if the correct logical proposition is “complicated”, it cannot be obtained due to the roughness of multiple regression analysis using linear functions and if the given numerical data are “wrong”, the correct logical proposition cannot be obtained. If more correct results are desired, multi-linear functions of a higher order (For example, in the case of two variables, a second order multi-linear function can be represented as $axy + bx + cy + d$.) should be used for multiple regression analysis.

3.6 Error analysis

In the case of the domain $\{0, 1\}$, Linial showed the following result [Linial 93]:

Assume that the probability distribution is the uniform distribution. Let f be a Boolean function, there exists a k -multi-linear function g such that $\|f - g\| < \epsilon$, where k is at most $O(\log(n/\epsilon)^2)$, where n is the number of variables.

This theorem means that Boolean functions can be well approximated by a multi-linear function of low order. We can conjecture that this property also holds in the domain $[0,1]$, which will be included in future work.

4 Conclusions

This paper has presented a method to discover logical propositions in numerical data. The method is based on the space of multi-linear functions, which is made into a Euclidean space. A function obtained by multiple regression analysis in which data are normalized to $[0,1]$ belongs to this Euclidean space. Therefore, the function represents a non-classical logical proposition and it can be approximated by a Boolean function representing a classical logical proposition. We have proved that this approximation method is a pseudo maximum likelihood method using the principle of indifference. We also have experimentally confirmed that correct logical propositions can be obtained by this method. This method will be applied to the discovery of logical propositions in numerical data.

References

- [Keynes 21] J.M. Keynes: *A Treatise on Probability*, Macmillan, London, 1921.
- [Langley 81] P. Langley, G. Bradshaw and H.A. Simon: BACON.5: The discovery of conservation laws, *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pp.121-126, 1981.
- [Linial 93] N. Linial, Y. Mansour and N. Nisan: Constant depth circuits, Fourier Transform, and Learnability, *Journal of the ACM*, Vol.40, No.3, pp.607-620, 1993.
- [Quinlan 93] J.R. Quinlan: *Programs for Machine Learning*, Morgan Kaufmann, 1993.
- [Shannon 49] C.E. Shannon and W. Weaver: *The Mathematical Theory of Communication*, Univ. III. Press, 1949.
- [Tsukimoto 94a] H. Tsukimoto and C. Morita: The Discovery of Propositions in Noisy Data, *Machine Intelligence 13*, Oxford University Press, 1994.
- [Tsukimoto 94b] H. Tsukimoto: On Continuously Valued Logical Functions Satisfying All Axioms of Classical Logic, *Transactions of The Institute of Electronics, Information and Communication Engineers*, Vol.J77-D-I No.3, pp.247-252, 1994 (in Japanese).
- [Wilks 62] S.S. Wilks: *Mathematical Statistics*, John Wiley & Sons, 1962.