人　工　知　能 75－1
(1991. 3. 6)

確率論理の幾何的モデルと帰納推論への応用

月本　洋
(株) 東芝　システムソフトウェア技術研究所

本論文は確率論理の幾何的モデルについて述べる。この幾何的モデルは論理関数を空間上の一点として扱うと言うアイデアに基づいている。従って、本幾何的モデルは、関数解析の論理版と言える。ヒルベルト空間と同様の手法を採用することによって、論理関数のユークリッド空間が得られるが、この空間上に論理関数の情報量が定義されている。この情報量等を用いる事によって、可能世界上の確率分布から命題を導出すると言う帰納推論が可能である。

A Topological Model for Probabilistic Logic and Its Application to Inductive Inference

Hiroshi Tsukimoto
Systems & Software Engineering Laboratory
Toshiba Corporation

This paper presents a topological model for probabilistic logic. This topological model is based on an idea that logical functions are regarded as points in space. As an idea like this is used in functional analysis, this topological model can be regarded as a functional analysis in logic. By a method similar to that for Hilbert space, Euclidean space will be obtained on which the information in a proposition will be defined. The merits of this theory will be confirmed in an application to an inductive inference, by which a proposition is induced from a probability distribution over possible worlds.

1.Introduction

In artificial intelligence, reasoning with uncertain information is important and will be more important in the future[Bra90]. Therefore there are many studies about probabilistic reasoning(for example[Nil86,FH88]). However few researchers have introduced topology into logic. In [Tsu90a], the author presented a topological model for propositional logics. In this paper, the model will be developed for probabilistic logic and will be applied to inductive inference.

There are some distances among propositions concerning uncertain information in natural languages (= probabilistic propositions). Now we consider the following propositions.

1. She is an old woman.
2. She must be an old woman.
3. She might be an old woman.
4. She is not an old woman.

It is possible to argue the distances among propositions 1., 2., 3. and 4.. For example, proposition 3. is nearer to proposition 1. than proposition 4., and is further than proposition 2.. In addition, it is also possible to argue the information in a proposition. For example, the information in proposition 1. is greater than that in proposition 2.. As the above examples show, there are some distances (topologies) among propositions expressing uncertain information, and the propositions have certain information. Nevertheless, up to now, these distances among propositions and the information in a proposition have been ignored or thought little of. This paper proposes a topological model for probabilistic logic, including classical logic, on which the information in a proposition is defined. This paper treats the case of propositional logic. Euclidean distance is the most natural, therefore the goal is to introduce Euclidean distance in the set of probabilistic logical functions, which can be equated with probabilistic propositions.

There is the theory of functional analysis, which introduces distances in the sets of functions to treat functions as points in space topologically. Especially, Hilbert space is an extension of Euclidean space. Therefore, it is natural to try to construct the Euclidean space of probabilistic logical functions by a method similar to that for Hilbert space.

It is worth noticing that classical logic has properties similar to vector space. The properties are seen in Boolean algebra with atoms which is a model for classical logic. The atoms (a_i) in Boolean algebra have the following properties. $a_i \cdot a_i = a_i$ (unitarity), $a_i \cdot a_j = 0$ $(i \neq j)$ (orthogonality), $\Sigma a_i = 1$ (completeness). This shows that the atoms in Boolean algebra have properties similar to unit vectors. In other words, the atoms in Boolean algebra are similar to the orthonormal functions in Hilbert space.

As Boolean algebra has properties similar to Hilbert space, constructing Euclidean space starts with Boolean algebra. The process from Boolean algebra to Euclidean space is as follows:

1. Represent Boolean algebra by elementary algebra. In other words, present an elementary algebra model for classical logic. (Section 2)

2. Extend the truth value. From this, a probabilistic logic is obtained. (Section 3)

3. Expand the above model to the space where nonclassical logical functions, including probabilistic logical functions, are represented. In other words, eliminate the idempotent law partially and expand the model to the space where the idempotent law doesn't hold. (Section 3)

4. Introduce an inner product to the above space and construct Euclidean space where probabilistic logical functions are represented as vectors. (Section 3)

On this Euclidean space, the information of probabilistic logical functions will be defined. For example, the information of x is 1 bit and the information of xy is 2 bits. A principle will be presented about the correspondence between probabilistic logical functions and probability distributions over possible worlds. This principle is explained with a primitive case. The vectorial representation of $A \vee \overline{A}$ is (1,1), where atoms are A and $\overline{A}$. $A \vee \overline{A}$ is tautology, therefore the information in $A \vee \overline{A}$ is 0 bit. Let w_1 and w_2 be possible worlds corresponding to A and $\overline{A}$, respectively. The probability distribution over w_1 and w_2 which corresponds to $A \vee \overline{A}$ should have no information, namely the entropy of the probability distribution corresponding to tautology should be 0 bit. The probability distribution whose entropy is 0 bit is (1/2, 1/2). Therefore the proposition (1, 1) corresponds to the probability distribution (1/2, 1/2). This correspondence principle will be discussed in detail in Section 4.

According to the above principle, the proposition (1,1,1,0) which means $A \vee B$ corresponds to the probability distribution over possible worlds (1/3, 1/3, 1/3, 0). The transformation formula will be obtained from a probability distribution over possible worlds to the vectorial representation of a probabilistic logical function. In the above example, (1, 1, 1, 0) will be obtained from (1/3, 1/3, 1/3, 0) by the transformation formula. With this transformation formula, probabilistic propositions, such as 'rain falls with probability 0.7' will be represented in the space of probabilistic logical functions. Details are discussed in Section 5. The linear algebraic method in this paper is similar to [Nil86,FH88,Shv90] and so on. However, [Nil86]etc. treat probability distributions directly, while this paper treats (probabilistic) logical functions which are natural expansions of classical logical functions. Namely [Nil86]etc. treat the vectors of probability distributions, while this paper treats the vectorial representation of (probabilistic) logical functions. Therefore, the merits of the method presented in this paper are that probabilistic logic is treated topologically in addition to linear algebraically and that the probabilistic logic is a natural expansion of classical logic. These merits are confirmed in an application .

This theory will be applied to an inductive inference from a probability distribution over possible worlds. A primitive example is that the proposition 'whenever it rains, it is cloudy' will be induced from the probability distribution over possible worlds (0.2, 0.01, 0.29, 0.5), where the first figure is the probability that it rains and it is cloudy, and so on. An approximation method in addition to the above transformation formula is used in this inductive inference, which is realized by the fact that topology and information are introduced into logic. (Section 5)

2. A model for classical logic

We review the model introduced in [Tsu90a]. More detailed explanations can be found in [Tsu90b]. Hereinafter, let $X, Y, \ldots$ stand for propositional variables and let $x, y, \ldots$ stand for variables. Let $F, G, \ldots$ stand for propositions and let $f, g, \ldots$ stand for logical functions.

2.1 Some preliminaries

(1) Definition of τ_x

Consider $f(x) = q(x)(x - x^2) + r(x)$, where $f(x), q(x)$ and $r(x)$ are real polynomial functions, and $x \in \{0, 1\}$. τ_x is defined as follows:

$\tau_x(f(x)) = r(x)$.

(2) Definition of τ

When f is of n variables, τ is defined as follows:

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

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

(3) Definition of L

Let L be the set of all functions satisfying $\tau(f) = f$. Then $L = \{f | \tau(f) = f\}$. L is the set of linear real polynomial functions. In case of 1 variable, $L = \{ax + b | a, b \in \mathbf{R}\}$.

(4) Definition of L_1

Let $x, y, \ldots$ be variables, where $x, y, \ldots \in \{0, 1\}$. L_1 is inductively defined as follows:

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

(5) If f is in L_1, then f satisfies $\tau(f^2) = f$. That is, $f \in L_1 \Rightarrow \tau(f^2) = f$. This can be easily checked. Obviously, if $f \in L_1$, then $f \in L$.

2.2 A new model for classical logic

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

$F \wedge G \Leftrightarrow \tau(fg)$,

$F \vee G \Leftrightarrow \tau(f + g - fg)$,

$\overline{F} \quad \Leftrightarrow \tau(1 - f)$.

where $\wedge, \vee$ and $\overline{a}$ stand for conjunction, disjunction and negation, respectively. Then (L_1, τ calculation) is a model for classical propositional logic, namely L_1 and τ calculation satisfy the axioms for Boolean algebra. Proofs can be found in [Tsu90b].

2.3 On $\tau(f^2) = f$

It can be verified that if f satisfies $\tau(f^2) = f$, then f is in L_1. Then, $(f \in L) \wedge (\tau(f^2) = f) \Leftrightarrow f \in L_1$ follows from $(f \in L) \wedge (\tau(f^2) = f) \Rightarrow f \in L_1$ and $f \in L_1 \Rightarrow (\tau(f^2) = f) \wedge (f \in L)$ in 2.1 (4) . Thus, the subset of L which satisfies $\tau(f^2) = f$ is equal to L_1 (the set of classical propositional logical functions).

3. A topological model for probabilistic logic

3.1 Extension of truth value

The truth value is extended from $\{0, 1\}$ to $[0,1]$. By this extension, functions become continuous functions ($f : [0, 1]^n \rightarrow [0, 1]$).

This truth value is interpreted as probability. This extension makes classical logic a probabilistic logic. The formulas for this probabilistic logic are the same as those for classical logic.

3.2 A probabilistic logic

Let p_1, p_2 probabilities of X, Y. Then the calculations are as follows:

1. $X \wedge Y \Leftrightarrow p_1 p_2$,
2. $X \vee Y \Leftrightarrow p_1 + p_2 - p_1 p_2$,
3. $\overline{X} \Leftrightarrow 1 - p_1$.

Obviously, this calculation is valid only if propositional variables are independent, therefore some modifications(for example [Bun90]) or additional definitions of calculations like some weak logics(for example [Gir87]) will be necessary. However, let us go ahead with the problem unsolved as a future work. Here $X \wedge X$ is not calculated as $p_1 p_1 = p_1^2$. Because $X \wedge X = X$, namely $\tau(x^2) = x$, the accurate calculation of $X \wedge X$ is p_1.

3.3 Elimination of $\tau(f^2) = f$

In 3.1, truth value was extended, which represents uncertainty of variables. In this section, we will represent uncertainty of propositions. This means that probabilistic propositions are represented by probabilistic logical functions. Therefore, it is necessary to expand the space of classical logical functions (L_1). As 2.3 shows, $\tau(f^2) = f$ means the space of classical logical functions. Here, $\tau(f^2) = f$ is eliminated in order to expand this space, then the space of linear polynomial functions(L) is obtained. Hereinafter, this space(L) will be made into Euclidean space(= a finite dimensional inner product space). Elimination of $\tau(f^2) = f$ means the elimination of the idempotent law for formulas except variables (for example x, y,...).Therefore,this space is a model for a weak logic where the contraction rule and the weakening rule hold only for variables. Therefore this space must have much to do with Girard's linear logic, where the contraction rule and the weakening rule don't hold. The study of the relation between this model and linear logic will be a future work.

3.4 Euclidean space

(1) Definition of inner product

Inner product is defined as follows:

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

where f and g are in L, and the integral is generally a multiple integral. This has the properties of an inner product.

(2) Definition of norm

Norm is defined as follows:

$$\| f \| = \sqrt{< f,f >}.$$

This has the properties of a norm. This norm is denoted by $N_r(f)$.N_r stands for relative norm, which depends on the dimension of space (the number of variables). Therefore, L becomes an inner product space with the above norm. The dimension of this space is finite, because L consists of the linear polynomial functions of n variables, where n is finite. Therefore, L becomes a finite dimensional inner product space, namely Euclidean space.

A new norm is defined as follows:

$$N(f) = \sqrt{2^{-n}} Nr(f).$$

This norm is used in Section 4.

(3) Orthonormal system

The orthonormal system is as follows:

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

$$\text{where } e(x_j) = 1 - x_j \ \text{ or } x_j.$$

It is easily understood that these orthonormal systems are expansions of the atoms in Boolean algebra (See the Introduction). It is also easily verified that the orthonormal system satisfies the following properties:

$$< \phi_i, \phi_j >= 0 \ (i \neq j)$$

$$= 1 \ (i = j),$$

$$f = \sum_{i=1}^{2^n} < f, \phi_i > \phi_i.$$

This Euclidean space is an expansion of Hasse diagram of classical propositional logic.

4. Correspondence between the vectorial representation of a probabilistic logical function and a probability distribution over possible worlds

4.1 The information of a probabilistic logical function

The information of a probabilistic logical function is introduced, which is called logical entropy. Logical entropy H_L is defined as follows:

$$H(f) = -log_2(N(f))^2 (= -log_2(\int_0^1 \tau(f^2)dx)).$$

Examples are shown below.

$H_L(1) = 0 \quad \Rightarrow$ This means that the information contained in tautology is 0 bit.

$H_L(x) = 1 \Rightarrow$ This means that the information contained in x (namely the affirmation of a certain proposition) is 1 bit.

$H_L(xy) = 2 \Rightarrow$ This means that the information contained in xy (namely the conjunction of a certain proposition and another proposition) is 2 bits.

$H_L(0) = \infty \Rightarrow$ This means that the information contained in contradiction is infinite.

As the above examples show, the information of a probabilistic logical function is reasonably defined.

4.2 Correspondence principle between the vectorial representation of a probabilistic logical function and a probability distribution over possible worlds

(1) Terms and notations

The following terms and notations are introduced.

1. f stands for a probabilistic logical function.
2. **f** stands for a propositional vector, which is the vectorial representation of a probabilistic logical function.
3. (w_i) stands for possible worlds.
4. **p** stands for a probability vector, which is a probability distribution over possible worlds.

(2)At first, $X \vee \overline{X}$ of 1 variable is considered. The propositional vector of $X \vee \overline{X}$ is (1, 1). Then, this proposition is tautology. Therefore, this proposition has no information, namely the information in this proposition is 0 bit. The probability distribution whose information is 0 bit is (1/2, 1/2), where the information of a probability distribution (I) is defined as usual:

$I = n - H \; (H \doteq -\sum_1^{2^n} p_i log_2 p_i)$,

where n is the number of variables and p_i is probability. Therefore, the propositional vector (1, 1) corresponds to the probability vector (1/2, 1/2). Similarly, the following correspondences are obtained :

tautology of n variables $(1,...,1) \Leftrightarrow (1/(2^n), ..., 1/(2^n))$,

X of 2 variables $(1, 1, 0, 0) \Leftrightarrow (1/2, 1/2, 0, 0)$.

The latter correspondence follows from the following two facts. The first fact is that X denies $\overline{X}$, therefore the probability of possible worlds corresponding to $\overline{X}$ must be 0. Another is that the propositional vector (1, 1) corresponds to the probability vector (1/2,1/2).

These correspondences can be generalized to the following principle. Let (a_i) be a propositional vector and let (b_i) be a probability vector. Suppose m elements of (a_i) are 1.

If $a_i = 1$, then $b_i = 1/m$, and if $a_i = 0$, then $b_i = 0$

The above correspondence principle means that the direction of propositional vector is the same as that of probability vector and that they differ in the norm.

(3) The information of a probabilistic logical function is equal to the information of a probability distribution.

The information of a probabilistic logical function was defined in 4.1, and the information of a probability distribution was described in 4.2 (2). $H_L = I$ (namely $H_L = n - H$) can be obtained from the above two informations. Proofs can be found in [Tsu90b].

Thus two important relations between propositional vectors and probability vectors are obtained:

1. The directions of two vectors are the same,

2. The information of two vectors are the same.

Hereinafter, these relations are extended to nonclassical logics including probabilistic logic. In other words, we will use the above principle in nonclassical logics too.

4.3 Transformation formula

Probability distributions are transformed into vectors which represent probabilistic logical functions. This transformation will be executed by the above correspondence principle. Transformation from a probability vector to a propositional vector is obtained from the above two relations. From 2., $N_r = 2^{H/2}$ is obtained. Since $H_L = I \to H_L = n - H \to -log_2(2^{-n/2}N_r(f))^2 = n - H$. The directions are the same from 1.. Therefore, the transformation formula is

$\mathbf{f} = (2^{H/2}/|\mathbf{p}|)\mathbf{p}$, where $H = -\sum_1^{2^n}(p_i log_2 p_i)$, $|\mathbf{p}| = (\sum_1^{2^n} p_i^2)^{1/2}$, $\mathbf{p} = (p_1, ..., p_{2^n})$.

By this transformation formula, we can transform a probability vector to a propositional vector. Therefore, when we know a probability distribution over possible worlds, we can obtain the proposition corresponding to the probability distribution over possible worlds.

5. An application to inductive inference

5.1 Inductive inference from a probability distribution over possible worlds

We can infer a proposition from a probability distribution over possible worlds by the formula in 4.3. This inference is a kind of inductive inference. It is desired that a probabilistic proposition is expressed by natural language. The solution is to approximate the probabilistic proposition by an appropriate proposition of classical logic. The approximation method is as follows. Let (f_i) be the propositional vector which represents a probabilistic proposition, and let (g_i) be the propositional vector which represents a classical proposition. Then, if $f_i \geq a$, then $g_i = 1$, otherwise $g_i = 0$.

5.2 An example

An example is shown in Table 1. Here, X stands for rain and Y stands for cloudy. Although w_2 is very scarce, w_2 is considered, since probability p_2 is not 0. This example is about the weather. Therefore, the question will be 'Which classical propositions about the weather will be obtained from which probability distribution over four possible worlds?' Table 2 shows two specific instances of probability vector in Table 1 and the results of this inductive inference. Here, **p** stands for a probability vector. **f** stands for the propositional vector obtained by the transformation formula. g stands for the classical propositional vector obtained by approximation. (Let a be 0.5.)

Table 1. Four possible worlds

possible world	W1	W2	W3	W4
rain	rain	rain	not rain	not rain
fine/cloudy	cloudy	fine	cloudy	fine
probability	P1	P2	P3	P4
logical function	xy	$x\overline{y}$	$\overline{x}y$	$\overline{xy}$

Table 2. Two instances

No.	p	f	g	classical proposition
1	(0.23.0.02.0.3.0.45)	(0.99.0.09.1.29.1.94)	(1.0.1.1)	$\overline{X}$vY(=X→Y)
2	(0.45.0.3.0.23.0.02)	(1.94.1.29.0.99.0.09)	(1.1.1.0)	XvY

Instance No.1 means that fine and cloudy are even, and it rains one day per four days, and it very seldom rains, when it is fine (However the real probability of rain in fine weather is less than 0.02). Instance No.2 means that it is rather cloudy than fine, and that it is very seldom fine and that fine and cloudy are nearly even when it rains.

The result of Instance No.1 is $\overline{X} \vee Y (= X \rightarrow Y)$, which means 'whenever it rains, it is cloudy'. As probability$(X \rightarrow Y)$ is 0.98, this result is reasonable.
The result of Instance No.2 is $X \vee Y$, which means 'it rains or it is cloudy'. As probability$(X \vee Y)$ is 0.98, this result is reasonable. Naturally, $\overline{X} \vee Y (= X \rightarrow Y)$ cannot be obtained, which means 'whenever it rains, it is cloudy', since the probability of w_2 is 0.3, which is very big compared with that of Instance No.1 and is not much smaller than that of w_1.

These instances show that this inductive inference is reasonable and effective. Why we can obtain classical propositions from a probability distribution is mainly attributed to the fact that topology(distance) and information are introduced into the space which is a model for an expansion of classical logic. We will apply this inference to larger problems. However, this method has the problem of computational complexity, which will be argued in another paper.

6. Conclusions

The author has presented a topological model for probabilistic logic and applied it to an inductive inference. This topological model can be viewed as a functional analysis of logical functions, which enables logical functions represented as vectors. The author has also presented a correspondence principle between propositional vector (probabilistic logical function) and probability vector(probability distribution over possible worlds), by which the vectorial representation of a probabilistic proposition has been possible. Furthermore an inductive inference from a probability distribution has been obtained as an application of the above correspondence principle. The essence of this paper can be an introduction of distance and information into logics. In addition, it is characteristic that logic and probability are treated together and separately.

References

[Bra90] R.J.Brachman, The Future of Knowledge Representation, *Proceedings of 8th AAAI*, 1990, pp1082-1092.

[Bun90] A.Bundy, Incidence Calculus, Edinburgh Research Paper No.497, 1990.

[FH88] R.Fagin and J.Y.Halpern, A Logic for Reasoning about Probabilities, *Proceedings of 3rd IEEE annual symposium on Logic in Computer Science*, 1988, pp410-421.

[Gir87] J.Y.Girard, Linear logic, *Theoretical Computer Science 50*, 1987, pp1-102.

[Nil86] N.Nilsson, Probabilistic logic, *Artificial Intelligence 28*, 1986, pp71-87.

[Shv90] H.Shvaytser, Probabilities That Imply Certainties, *Proceedings of 8th AAAI*, 1990, pp665-670.

[Tsu90a] H.Tsukimoto, A Topological Model for Propositional Logics, *Transactions of Information Processing Society of Japan 31*, 1990, pp783-791 (in Japanese).

[Tsu90b] H.Tsukimoto, A Topological Model for Probabilistic Logic and Its Applications, working paper, 1990.