

# 人工知能による著作物の創作性尺度に関する アルゴリズム情報理論から見た考察

金子 格†

東京工芸大学†

**概要**：これまでより知能的な処理(人工知能)による、著作権の生成、改変の利用が拡大している。今後これまで、表現とみなされてきた著作物の同一性、独立性の判断の基準が揺らいでいく可能性がある。知的なデータ処理によってもある作品を従来の基準では独自の作品とみなされる程度に変換することが可能となりつつある。機械的に変換した作品はもちろん著作権法の目的を考えれば、引用または改変とみなされるべきという考え方もある。その基準はどうあるべきか、本論ではアルゴリズム情報理論をその判断基準とする場合の諸問題について検討する

## On measure of originality of production of AI using algorithmic information theory.

ITARU KANEKO†

### 1. はじめに

数々の AI の成功が続きその実用化が加速している[2][3][4]。こうした状況において AI が生成した情報の知的財産権に関する制度設計は喫緊の課題である。

アン女王法以来、知的財産法制度は社会の変化、技術の進歩とともに、大きな発展と変貌を遂げた。その制度設計において、社会における知的財産の効用の最大化が当然重視されてきた。ここ 20 年、デジタル技術、インターネット技術が普及する中で知的財産の効用を拡大するために多くの改訂が行われた。

AI の登場はさらに大きな改訂を必要とするように思える。その議論には、多方面からの検討が必要と考えられるが、本稿では AI による創作の参照モデルを構築しモデル上で操作される情報の量と性質を検討し、さらにアルゴリズム情報理論を適用して AI が生成した創作物の知的財産権を論ずる。

### 2. AI による創作の参照モデル

#### 2.1. 知財化候補データ

本稿では、個別の事例においてあるデータが法的に知的財産権であるか、だれに帰属するかという議論を行うことが目的ではない。そのような議論を避けるため、解釈により知的財産とみなしうる情報を、本稿において特にその法的属性を決定

せず「知財化候補データ」と呼ぶこととする。つまりこれらは外形的な要件として知的財産権が認められうる内容と形式を持つあらゆるデータ、つまり、音楽、特許、文学作品、レシピ、デザインなどに相当するデータであり、もしそのデータの生成手段が慣習的に通常の知的財産の生成手段と認められたものであれば(通常の方法で主に人間が創作したものであれば)、知的財産権を主張すれば知的財産として認められうるために十分な、内容と形式を有しているデータである。

#### 2.2. AI による創作の参照モデル

本稿における AI による創作の参照モデルを図 1 に示す。本モデルは筆者等が提唱している[1]。

F:知財化候補データはすでに説明したように知的財産の外形的な条件を有しているデータである。人間の作者は A:指令と、B:作者入力情報を入力し、D:補助機構を操作して創作を行う。補助機構は C:参照情報として、外部の情報を参照し、内部に E:蓄積情報を持ち、事前に収集した情報を蓄積加工して用いる。AI による創作は、D:補助機構が高度に知能的な動作をし、C, E を用いて人間の指示をほとんど受けなくても人間の創作と同程度の F:知財化候補データを生成しうる状況としてモデル化する。

† 東京工芸大学

Tokyo Polytechnic University

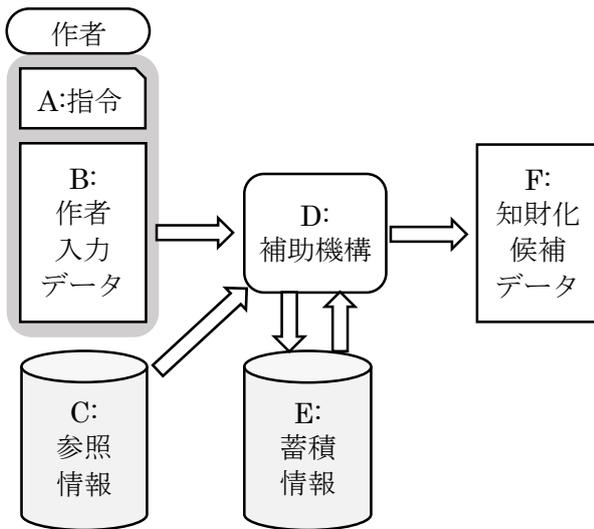


図 1 AI による創作の参照モデル

まず議論の出発点として、補助機構が AI ではなくパソコン上の単純な編集ソフトである場合について考える。表 1 に、今日の著作権法上の扱いをまとめる。ただし議論の出発点のための簡単な考察が目的であるから、法的には若干厳密性を欠くが一部単純化した。

表 1 著作物における情報源と著作権上の扱い

A	B	C	D	E	
◎					創作
	◎				創作
		◎			引用
			◎		D に帰属
				◎	様々

創作が A, B からなされた部分は通常の創作活動による部分である。部分的に C に依拠する場合引用と考えられる(引用の要件が満たされない場合もある)。D のみに依存する部分は D の付随物とみなされる(たとえばパソコンの起動画面)。E のみに依拠する場合、E に蓄積された情報源の性質により引用、創作、D の付随物のいずれかの場合があり得る。

### 3. 情報量の影響

#### 3.1. 均等確率の蓄積情報

以下では情報量が、F:知財化候補データの生成にどう影響するかを考察するためいくつかの例を示す。

図 2 に再び AI による知的財産候補データの生成の参照モデルを示す。ここでは均等確率の蓄積情報:E1 と指令 A1 から F1:知的化候補データを生成する。

この例における蓄積情報 E1 を表 2, 指示:A1 を表 3 に示す。蓄積情報:E2 には、様々な長さの文字列が集積され、指示:A1 はその蓄積情報をピックアップする指示でありその結果出力が得られる。

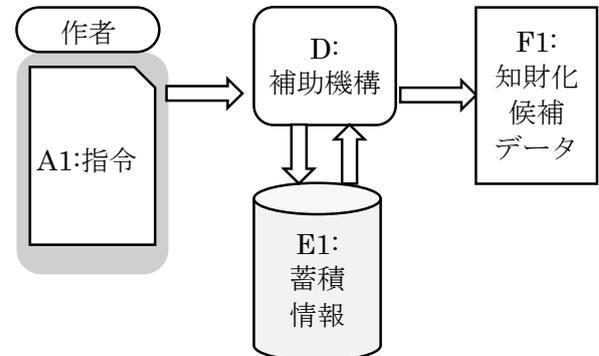


図 2 均等確率分布による蓄積情報 E1:により F1:知財化候補データを生成する場合

表 2 E1:蓄積情報. E1 では各 s と l について以下の文字列が蓄積情報として蓄積されている。文字列の分布はすべての文字列パターンで均等である。

	l=1	l=2	l=3	l=4
H1[bit]	4.75	9.5	14.25	19
S				
1	A	AA	AAA	AAAA
2	B	AB	AAB	AAAB
...	...	...	...	...
26	Z	AZ	AAZ	AAAZ
27		BA	ABA	AABA
28		BB	ABB	AABB
...		...	...	...

表 3 指示 A1, A1 では以下のような指示により文字列が出力される

Command	(s,l)	出力	H12[bit]
1	1,1	I	4.75
2	2,12	AM	9.5
3	2,374	OK	9.5

指示により、任意の長さの任意の文字列が選択できるから任意の英文を生成することができる。また E1 はすべての文字列の並びを列挙しているにすぎず知的財産として価値のあるデータとは考えられない。したがって F:知財化候補データの価値に E1 は寄与しておらず A1 が知財化候補データに価値を与えていると考えられる。

### 3.2. 不均等確率の蓄積情報

次に図3に蓄積情報 E2 が表 4 に示す不均等な確率分布によって構成されている場合を示す。この場合、E2 には英文が蓄積されているがその出現順序は英文中で利用される頻度の順に並べられている。したがって有名なフレーズほど上位になり、上位のフレーズは短い記号で指定することができるとする。

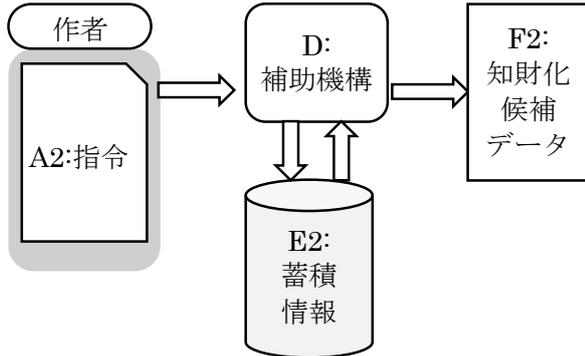


図 3 不均等確率分布による蓄積情報 E2:により F2:知財化候補データを生成する場合

表 5 に示す指示 A2 は均等確率の場合と同様に、蓄積情報の選択を示す指示である。ここでは 2 つの文章を選んで、そしてこの 2 つの文章を接続すれば Simon & Garfunkel, “The Sound Of Silence” の冒頭の歌詞になっている。

この場合 E2 の上位には単純な文が多く収容されているが、やがて大量の、しかし英文の中で最も有名なフレーズが連続してあらわれることになる。したがってまったくでたらめな指令を行っても、出力には有名なセンテンスが多く出力されることになる。

ここで着目すべきは、A1, A2 いずれの場合も指示次第で任意の文字列を生成し得ることであると思われる。E1 の蓄積情報は確率的に均等に分布しているが指示 A1 の与え方次第でいかなる単語、文字列も出力することができる。A2 の場合確かに命令の数値的な大きさが小さいほど、有名でよく知られたセンテンスが出力されるが、だからといってあまり使われないセンテンスもやがて表に出現するのでそのような文字列を出力する可能性がないわけではない。

このような場合に F1 の場合は指示 A1 が、F2 の場合は蓄積情報 E2 が知財化候補データにより大きく寄与することは情報の観点から解釈することができる。

まず E1 の文字列は、指示がでたらめであれば、E2 の全文字列は均等な確率で選ばれる。したがって任意の文字列の分布に対する E2 の情報量は

少なく、F1 の情報量はすべて A1 の情報量である。

F2:知財化候補データにおいては逆に事前に出現頻度の高い英文に数値の小さい、桁数の少ないコマンドが割り当てられている。したがって出力 F2 の情報量はそのほとんどが E2 の蓄積情報から得られたものであるということが出来る。

表 4 蓄積情報 E2. E2 では出現頻度を考慮した英文の表の一例。頻度が多い文に短い可変長符号を表した場合の bit 数を H1 に示している。

S	H1[bit]	Sentence
1	1	HELLO
2	1.5	YES
3	2	NO
4	2.3	I LOVE YOU
5		WHO ARE YOU
...		
2145656	21	HELLO DARKNESS MY OLD FIRIEND
...		
369310405	28	IVE COME TO TALK WITH YOU AGAIN
...		

表 5 指示 A2. Command として示す指示により出力文字列が出力される。出力 bit 数として空白を含む文字数×4.75bit を H2 として示す。

Command	出力	H2[bit]
2145656	HELLO DARKNESS MY OLD FIRIEND	137.8
369310405	IVE COME TO TALK WITH YOU AGAIN	152.1
...	...	

### 3.3. 情報量の算出

それぞれの持つ情報量はどう算出すべきだろうか。

図 2 に示した A1-E1-F1 のシステムの場合、A1 が持つ情報量はだいたい指示 A1 から通常の情報量を求めた情報量に等しいと考える。したがって A1 のコマンド一つあたりの情報量をコマンドの数だけ集計すればよい。一方指示 A1 により E1 から出力を選ぶことによって追加される情報量があるかを考えると、たとえば 3 文字の文字列パターンは全く均等に選ばれるような蓄積情報となっている。たとえば A1 の Command はだいたい 10bit であるが、出力される文字列の情報量もだいたい 10bit であり、E1 を経由することで追加される情報量はほぼゼロである。

一方図 3 に示した付均等確率の E2 を持つ A2-E2-F2 のシステムの場合は Command の持つ情報量は、文字数に比例する出力文字の情報量より

は少ない。この情報はどこから来たかという、たとえば表 5 の#2145656において Command の情報量は 20bit 程度であるが、出力の情報量は 25 文字×6=150bit 程度ありほぼ全部の情報量が E2 かけ得られていることがわかる。

かりに 25 文字の出力を指定する Command が 25 文字の文字列と同程度の情報量、すなわち 150bit = 10 進数で 40 桁程度の数であれば、E2 を参照することによって情報量の増加がない。E2 によって情報量が増加するのは E2 の文章の並びが均等確率ではないことが理由である。

## 4. アルゴリズムと知的財産権

### 4.1. アルゴリズム的情報量(AlgIM)

アルゴリズム的情報量(以下 AlgIM)の提案はチューリングにさかのぼることができる[5]。チューリングは円周率 $\pi$ の小数点以下の桁の連続が、一見ランダムなデータで無限の情報量を持つように見えるが、一定のアルゴリズムにより「計算可能」である点でランダムな数字の列とは異なることに着目し、「計算可能数」という概念を提唱した。

G. J. Chaitin は、これをさらに発展させアルゴリズム的情報量という概念を提唱した [6]。  $\pi$  の桁の並びはほぼランダムであるが、  $\pi$  の生成アルゴリズム自体はコンパクトである。したがって  $\pi$  を記録再生するには無限のメモリは必要ではなく、  $\pi$  のアルゴリズムと計算領域があればよい。データの情報量を、そのデータを生成し得るアルゴリズムのサイズとして定義すれば、  $\pi$  のような計算可能数の情報量は有限でありアルゴリズムに基づいた情報量の測度を構築できる。宮部も計算理論の中でアルゴリズム的情報理論に言及している[7]。

### 4.2. AlgIM と知的財産権

$\pi$  の桁はほぼランダムであることが経験的に知られている。したがって  $\pi$  を 27 進法で表現し、各桁にアルファベットと空白を割り当てれば、Simon & Garfunkel, “The Sound Of Silence”の歌詞や、シェイクスピアの小説も、いずれかの桁において出現する可能性が非常に高い。

そこで  $\pi$  を蓄積情報とし、蓄積情報の位置を指定することで出力を行う場合を考える。先の議論において蓄積情報から得られる追加の情報量を問題にしたが、蓄積情報から得られる正味の情報のみが重要であるという立場であれば、この場合知的財産はあくまで蓄積情報から派生したとみなせる。しかし文字列が  $\pi$  という数値の一部であることを考えると、そのような解釈は不自然とも考えられる。そこには AlgIM の評価が関係してい

ると考えられる。

蓄積された  $\pi$  の AlgIM はごくわずかである。  $\pi$  のすべての桁が、表面的に多くの情報量を有しているとしても、 AlgIM は少ない。したがってアルゴリズム歌詞の情報量は  $\pi$  から得られたものではなく、「桁数指定」という指示によって与えられたと解釈される。  $\pi$  の中に無限のコンテンツが含まれているとしても、 AlgIM としては  $\pi$  の計算アルゴリズムの長さに対応する情報量でしかないから、  $\pi$  に帰属する知的財産権は少ないと評価することが可能になる。

$\pi$  と同様に、表面的に多くの情報量を生成する手段は数多く存在する。そのように多量の「表面的」情報を生成することで多くの知的財産権を与えると有用な情報へのインセンティブにならない。より有用な情報を創出する作者や作業者にインセンティブを与える制度設計上、アルゴリズム的情報量を導入することは有益であると考えられる。

## 5. AlgIM の導入

### 5.1. AlgIM と AI の創造性

次に補助機構がより高度な AI となった場合に AlgIM が指標としてどのような特徴を有するかを検討する。

AI をアルゴリズムと考えると AI 自体の AlgIM は AI のコードサイズに比例する。 AI のコードサイズは E:蓄積情報に比して非常に小さい。すべての情報はほぼ外部から得られたものであるから、 AI が出力する知財候補データの AlgIM はそのほとんどが入力となる外部情報に起因することになる。すなわち AlgIM により評価すると、 AI が創作した知財候補データはほぼ外部の情報に依拠し、 AI の創造性はほとんどないということになる。これは一見極端な結論と思われるが、しかし AlgIM による評価の原理を考えると当然である。知財候補データに含まれる情報は元をただせばすべて外部の情報でありそれを計算により加工した結果であるから、アルゴリズム情報量として増加する情報量はない。

しかしこの結論に対し以下の反論が当然予想される。人間自身もまったく情報を得ずに生得的に創作を行うことはない。人間自身を生体的アルゴリズムとみなし、人間の創作物を AlgIM で評価すれば、人間の創作もすべて入力した情報に依拠すると結論づけることになる。 AlgIM による依拠性で創造性を判断すれば人間の創作活動の価値も否定することになるかもしれない。

したがって、 AI の創造性を評価する場合、 AlgIM による依拠性の評価だけでは不十分である、と考えられる。

とはいえ、4.1 で示したような単純な編集操作において AlgIM は有益である。単純な編集操作のような計算処理と、高度な AI のみが行う計算処理の違いをどこに見出すかが次の課題となるだろう。

### 5.2. 付加価値ゲインの考慮

そこで情報量が増えないが創作による付加価値の増加が認められるような場合を考える。図 4 に示すシステムでは E3:蓄積情報は表 6 に示すような内容であり A3:指令は

このフレームワークにおいて C<sub>31</sub> には非常に大きな価値があり他はほとんど無価値であるとする。すると同じ情報量であっても指示 C<sub>31</sub> の情報の持つ平均価値がきわめて大きく指示 C<sub>31</sub> 以外は価値がほとんどない。このような情報の価値に大きな差が生じている場合、その情報を生成している人工知能は高い付加価値を生成していると考えられることにする。

表 7 に示すような内容であるとする。知財化候補データはそれらを組み合わせた知財化候補データを構築するものとする。

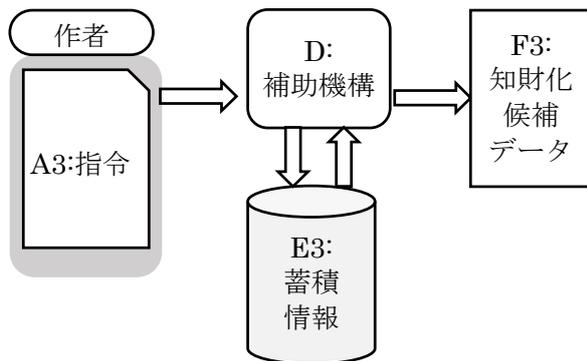


図 4 付加価値の低い蓄積情報 E3 から指令 A3 により付加価値の高い知財化候補データ F3 が生成する場合

表 6 蓄積情報 E3. E3 は付加価値の高い英文の表の一例。頻度が多い文に短い可変長符号を表した場合の bit 数を H1 に示している。

s	Sentence
1	HELLO
2	YES
3	NO
...	
k <sub>1</sub>	HELLO DARKNESS
...	
k <sub>2</sub>	MY OLD FIRIEND
...	

k <sub>3</sub>	IVE COME TO TALK
...	
k <sub>4</sub>	WITH YOU AGAIN
...	

このフレームワークにおいて C<sub>31</sub> には非常に大きな価値があり他はほとんど無価値であるとする。すると同じ情報量であっても指示 C<sub>31</sub> の情報の持つ平均価値がきわめて大きく指示 C<sub>31</sub> 以外は価値がほとんどない。このような情報の価値に大きな差が生じている場合、その情報を生成している人工知能は高い付加価値を生成していると考えられることにする。

表 7 指示 A3. Command として示す指示により出力文字列が出力される。出力 bit 数を空白を含む文字数×4.75bit として H2 に示す。

#	Command	出力	Capital Value
C <sub>31</sub>	k <sub>1</sub> ·k <sub>2</sub> ·k <sub>3</sub> ·k <sub>4</sub>	HELLO DARKNESS MY OLD FIRIEND IVE COME TO TALK WITH YOU AGAIN	100
...	それ以外		平均 0.001

このような例を検討すると、情報量を加えることなく付加価値が高まる場合があるといえる。人間や AI の創作活動の中にはそのように単純な情報量という意味では情報量の増加をともなわずに価値だけが上がる場合があり、そのような付加価値の寄与が人間や人工知能の創作性において重要であると考えられる。

### 5.3. 付加価値の安定性

そのような場合に付加価値は安定しているだろうか。人工知能はアルゴリズムであるから一旦実現してしまうと同等の結果を大量に生成することが可能である。したがって、人工知能が生成した知財候補データは価値の下がるということも容易に想像できる。

したがって、ある評価尺度において望ましい AI とは、その生成する知財候補データの付加価値が平均的に高いと期待される AI であるが、そのような AI の開発には大きなコストがかかり得るし、またそのような AI が実現した結果、計算処理自体のコストは安いので、AI による付加価値が低減し、AI が生成する知財候補データの平均価値と AI が提供するサービス自体の価値が下がること

が予想される。その結果、AI に提供される E3 の蓄積情報の希少性が高まり、市場価値としては E3 の価値が高まるという現象も起こりうるのではないかと考える。

このように AI の開発、利用自体が知財候補データの経済価値に大きな影響を与えつつ、AI はより付加価値の高い知財候補データを生成することを補助するために使われていく、と予想することができる。

したがって、AI による著作物の創作性指標を検討する上では、AlgIM による依拠性のみをその指標とするのはもちろん、人間の創造活動に対する AI がもたらす便益の最大化のためには適当ではない。知財候補データの経済的価値の増加を考慮すべきであるが、その経済的価値自体が AI の利用によって影響を受けることを反映することも必要であると筆者は考える。

一方、AlgIM の経済価値を評価する公平で信頼できる方法が必要と考えられるが、それについては今後の研究課題としたい。

## 6. まとめ

本稿ではいわゆる「人工知能」(以下 AI)による著作物の創作性尺度をアルゴリズム情報理論の観点から考察した。まずすでに筆者等が提唱した AI による創作の参照モデルを用い形式的な議論の準備を行った。

第二に、創作モデルで交換される情報量を分析し、情報量が著作物の依拠性に対応する場合と、情報量のみでは依拠性の評価に問題が生じる場合を示した。

第三に、アルゴリズム情報量(AlgIM)を導入し、第二の例の問題の一部が解消されるが、人間が著作物を作成する場合を含めた、創作活動の価値すべてを評価するには不十分であることを示した。

最後に AlgIM と「付加価値」の関係を議論し、AI による著作物の創作性を論ずるには、おそらく付加価値の評価が不可欠になるだろう理由を示した。また AI の利用自体が、付加価値に影響を与える可能性を示し、その場合には AI による著作物が普及することが、著作物の創作性に影響を及ぼし、さらには知的財産権の定義にも影響を及ぼす可能性を示唆した。

現在 AI が各分野で利用されることにより、人間の仕事の多くが AI に奪われるのではないかとの議論があり著作物においても全く同様であると考えられる。しかし同様の議論は、印刷、録音録画が実現した際に起こった議論と同じ側面を持つ。すなわち、録音を実現する前はすべての演奏に演奏者が必要だったか、録音により単純な演

奏の付加価値が下がる。その結果演奏者は、複製の元となる高度な品質を持つ演奏者。録音では対応できない、個別の条件に合わせたアドリブ演奏ができる演奏者。作曲家などのように分化していったと考えられる。

AI の利用拡大においても、同様に「創作活動において人間でなければできない部分」に人間の仕事に移り、そこには十分な報酬も支払われる、という楽観的な予想がある一方、そのような仕事は少なく、新たな創作がほとんど生まれなくなる、という危惧もある。

しかし AI のコストが安いという前提で、システムを AI による価値ある著作物の生成が行われるように最適化していけば、価値あるあたらしい著作物の生成に必要な、人間のみが生み出しうる情報に最大限のインセンティブを与えるシステムが構築できるのではないだろうか。そうしたシステムは社会に最大の効用をもたらすのではないだろうか。

付加価値の、確率的、統計的評価は機械が最も得意とする計算であり、AI をその部分に合わせて適用することで、人類に最も貢献する知的活動にインセンティブが与えられるような AI-人間共生システムを構築できる可能性は十分にあるのではないだろうか。

## 参考文献

- [1] 金子 格, AI, ML の産業応用の拡大における知的財産の扱いに関する考察, 情報処理学会研究報告, 電子化知的財産・社会基盤 (EIP), 2015-EIP-69(8), 1-4 (2015-09-03)
- [2] IBM, 質問応答システム“ワトソン”がクイズ番組に挑戦!, <http://www.ibm.com/smarterplanet/jp/ja/ibmwatson/quiz/>, (2011)
- [3] Quoc V. Le, et al., Building High-level Features Using Large Scale Unsupervised Learning, <http://research.google.com/pubs/pub38115.html>, (2012)
- [4] 高村 誠之他, 高度並列計算による進化的画像符号化の高速化, 電子情報通信学会技術研究報告. IE, 画像工学 110(148), 65-69, 2010-07-19(2010)
- [5] A.M. Turing, On computable numbers, with an application to the entscheidungsproblem., Proceedings of the London Mathematical Society, 420:230-265(1936)
- [6] G. J. Chaitin, "Algorithmic Information Theory", IBM Journal of Research and Development Volume: 21 Issue: 4(2000)
- [7] 宮部賢志, ランダムネスの一般化 (形式体系と計算理論), 数理解析研究所講義録 1729, 84-94, 2011-02(2011)