

高知大学大学院 理学研究科 情報科学専攻  
平成14年度 学位論文発表会 プログラム

2003年2月17日(月) 於 高知会館

発表者	論文題目	時間
井谷 健一	土地情報管理システムに関する研究	10:00 - 10:20
菊池 史尚	映像情報データマイニングに関する研究	10:20 - 10:40
解 文明	分散オブジェクト技術に基づく電子商取引システムの構築	10:40 - 11:00
中島 健造	メタデータ RDF に基づく XML データ格納法に関する研究	11:00 - 11:20
小藤 憲邦	Application of Network Time Protocol for LAN/WAN Performance Evaluation	11:20 - 11:40
昼 休 み		
谷 誠	GMS-5 衛星画像と気象記録の統合データベース構築に関する研究	13:00 - 13:20
前田 薫孝	光量子ゲートの基礎	13:20 - 13:40
西村 友宏	有限束の分類	13:40 - 14:00
Mao Sopath	Development of Chinese Input System for Mobile Phone	14:00 - 14:20
王 冬青	日経平均株価データのフラクタル解析	14:20 - 14:40
宗石 有平	気象データの非正規性の定量化	14:40 - 15:00

# G I S 運用に向けた土地情報管理システムに関する研究

情報科学専攻 計算機科学講座 井谷 健一

地方行政重点施策のなかで、住民のニーズに対応した、質の高い行政サービスが求められている。この要求に応えるべく誕生したのがG I Sである。しかし、G I Sを導入できる業務とできない業務が存在している。その導入できない理由には、県や他市町村との兼ね合い等もあり、システムを統合することは極めて困難なこと。その他にも、市町村独自のシステムの存在や市町村合併に関することなど、数々の問題があり、一つの業務であっても複数・異種のシステムが混在しており、G I Sのように、システムを1社に特定してしまうというのは非常に困難である。又、補助金の少ない事業などは、たとえデータ共有型G I Sを有しても維持・管理費が捻出できないのが現状である。

そこで、この問題点に当てはまる生産調整業務をモデルに選び、G I Sにデータを供給できる土地情報管理システムの開発を行うことにした。

本研究の目的は、システムを複数使用しているために、G I S導入に踏み込めない小規模事業を支援するシステムを研究・開発し、G I Sを幅広い業務で活用することを目的としている。

問題点を解決するためには、G I Sと地方自治体の多種多様な業務の間に隔たっている異種データという壁を無くし、スムーズに流通させる仕組みを持ったシステムの構築が必要である。

本研究で行ったことを以下に示す。

1. G I Sのデータ構成を研究した。
2. 今後、相互流通の面で期待されるG - X M Lについて研究を行った。
3. 生産調整業務における日常のワークフローを分析した。
4. 数種のシステムデータより生産調整データを新たに作成した。
5. G I Sやインターネットで活用できるシステムとした。
6. 生産調整業務の目標面積計算、達成予測が行えるシステムとした。
7. データの抽出・更新・削除が行えるシステムとした。

このシステムは、G I Sにデータを供給することによって、土地利用図を活用して、今まで困難であった分析や検索などを可能にするものである。また、システム間のデータの違いを乗り越えて、柔軟に流通・活用でき、生産調整業務のサポートの役割を果たすことができるシステムとなっているので、実際に職場での運用において、非常に効果的なシステムである。

映像情報はそれ自体で様々な情報を含んでおり非常に利用価値が高く、効率的な索引付けと検索利用に対する要求が高まっており、そのためには蓄積された情報をよく理解し、有用な知識を的確に判断する必要がある。このようなデータベースから知識を発見する方法としてデータマイニングがある。データマイニングは大量の事例データを扱い、自明でない規則性やパターンを半自動的に取り出す手法である。これらの従来のデータマイニングにおける多くの研究や商用システムは、リレーショナルデータベースにおける知識発掘として発展してきたが、マルチメディア情報におけるデータマイニングの研究は、まだ途上段階である。

本研究ではデータマイニング法の一つであるクラスタリングと相関ルールを組み合わせることによって知識発見の手続きを半自動的に行う。そして **Semantic Model**(対象物の知識情報)、**Cinematic Model**(シーン、イベントルール)、**Physical Model**(物理世界での制約条件)により構成された知識モデルを構築する。さらに確率論的なグラフィカルモデルである隠れマルコフモデル(HMM)で知識モデルを再編成する事により知識の蓄積を行う。さらにそのモデルのマッピングによりセマンティックかつダイナミックな索引付けをし、コンテンツベースのビデオの検索を可能とする情報システムを開発し、その有用性を確かめる。

本研究では、映像情報としてプロ野球の1試合のビデオ(約80分・21844フレーム)を実験対象とした。本研究の特徴を以下に示す。

1. 「クラスタリング→時系列変換→時間依存相関ルール」といったデータマイニングの手法を組み合わせたハイブリッド学習アルゴリズムを提案した。
2. 映像情報を連続した静止画像(フレーム)と考え、フレームをクラスタの集合に分類するため、自己組織化マップを用いてクラスタリングを行い、クラスタの特徴づけを行った。
3. 連続した静止画像に割り当てられたクラスタの値を時間軸上に列挙する。その時系列データからノイズを除去し、カットごとにまとめた時系列データとイニングの切れ目を抽出した時系列データを生成する。
4. 時間依存相関ルールを適用することで、時系列の中に存在する特徴パターンを抽出する。
5. そのパターン情報を、隠れマルコフモデルの手法によりモデル化し、より精度の高い知識モデルを作り出し、インタラクティブな索引付けを行う。
6. パターン情報をデータベースに格納し、時間軸を考慮した SQL クエリを行い、検索結果を動画として視覚的に表示する。

これにより、映像情報の意味を持った時系列データを知る事ができ、「先行のチームの選手がヒットを打ったシーン」のような情報の検索を行うことができる。また、変化パターン情報をつなぎ合わせる事により映像情報の要約が可能となる。我々の開発したこの手法は、他の様々なデータゲーム・試合の映像情報に適応できる可能性がある。

## 分散オブジェクト技術に基づく電子商取引システムの構築

情報科学専攻 計算機科学講座

解 文明

インターネットの急速な発展によって、電子商取引が盛んに行われており、企業では、商取引を電子化することにより、ビジネスの効率化が可能になり、コスト削減が期待できる。しかし、電子商取引では各企業間に独立して構築された異種情報源と異なるプラットフォームが分散して存在するという新たな環境が実現された。こうした環境でのデータ交換においては、人間が介在せずにコンピュータによる自動処理を行うことが困難であるという問題がある。このような問題の解決策の一つとして CORBA と XML のような技術が必要になる。CORBA はリモート環境で各々情報源に共通のインタフェースを持つラッパーを実装し、プラットフォーム・レベルでのデータの異種性を解決する。XML は利用者側で電子文書から一部を抽出したり、表現体裁を変えたりすることが可能である。

本研究では、CORBA と XML により、異なる企業間の情報統合、再利用、及び業務プロセスの連動の自動化を行う。企業間取引を想定して、受発注業務プロセスの他社との連動の自動化を実現し、企業間取引を介在する XML ファイルを用途に応じてさまざまな形式で表示することにより取引の効率化を実現する。応用例として、インターネットを利用して、多数の売り手、買い手企業間での取引を実現する電子商取引の商流部分を構築した。本システムで行ったことを以下に示す。

1. システムでは、エージェント型メデイエーションシステムである HI-AMS と CORBA 環境を組み合わせたアーキテクチャを基本モデルとする。XML 用メッセージングプロトコル SOAP を採用し、ラッパーを Java 言語で、JavaIDL にて実装した。
2. 情報源のドメイン情報の登録・選択機能や、CORBA IDL による各情報源のラッパーの共通インタフェースの定義により、異種分散情報源への統一的なアクセス機能を実現し、異種データ間の検索及び検索結果の XML ファイルの作成を可能にした。
3. 企業間で交換する XML ファイルを XML パーサーで解析し、DOM ツリー構造から必要なデータを取得しデータベースに登録する。また、XSLT プロセッサで企業の XSL スタイルシートを適用することで新しい XML ファイルを生成することもできる。

このシステムは電子商取引において、商品の検索、注文書・請求書等の XML 化及びプラットフォーム・レベル、データモデル・レベルでの異種性の解決によって、複数の企業間、また、金融、保険、医療業界等の組織間で、業務プロセスとの連動の自動化、取引の効率化を実現することができるように設計と実装がされている。

ネットワークを通じて交換される文書やデータの共通フォーマットとして普及している XML はデータ構造を自由に変更できる柔軟性を持っていることからその交換性という要件を超えて、データベースに格納し検索できるという要件が重要になってきている。XML 問合せ処理のベンチマークも開発されはじめているため、詳細な評価実験を中心とした各手法の比較が行われるようになることが予想される。XML データベースの実現手法としては、関係データベース、オブジェクト指向データベースや XML 専用のデータベースを用いる方法等が提案されているが、中でも関係データベースを用いる方法は、最も普及しているデータベース形式であり稼働中のシステムが多数存在し、大量のデータが蓄積されていることから重要である。また、過去 20 数年に渡る研究による、問合せ最適化や論理作業単位(トランザクション)管理等の技術の蓄積があることなどから、XML を既存の計算機資源、情報資源と連携して利用するのに有利である。

本研究では XML データを関係データベースに格納する以下の3種類の方法についての性能を評価しその有用性を検証する。

- ・XML データをそのまま格納する方法(ファイル格納方式)
- ・XML の要素をフィールドに対応させて格納する方法(テーブル写像方式)
- ・XML の木構造に基づきノードとリンクに分解し格納する方法

(木構造写像方式)

また、Web ページのデータ(メタデータ)を記述する枠組みとして提案されている RDF(Resource Description Framework)モデルを XML により表現する。

本研究で行ったことを以下に示す。

1. 上記の3種類の格納方式のそれぞれに対して格納、検索するシステムを構築した。
2. 格納された XML データに対する問合せは XQuery にて行い SQL 式に変換して関係データベースで処理させるモジュールを実装した。
3. 実験データに書誌情報を使用し、RDF と DublinCore メタデータを組み合わせ XML 形式で表記した。

本研究によりそれぞれの格納方式による特徴を発見することができた。ファイル格納方式では階層構造が深い場合に有利、テーブル写像方式では複数の XML 文書から検索する場合有利、木構造写像方式では文字列照合を有した問い合わせの場合に有利という結果が得られた。

## Application of Network Time Protocol for LAN/WAN Performance Evaluation

情報科学専攻 計算機科学講座 小藤憲邦

### **Objective**

The use of special devices is needed to monitor network traffic. Usually, the access to these devices is restricted by many means, as security for example. This work is based on NTP (Network Time Protocol) to evaluate network traffic without access to routers, switches or any other device with network traffic information.

### **Method**

NTP servers synchronize with each other asking time, putting timestamps between interactions and calculating the delay between them. It uses many algorithms to try to minimize the effect of network traffic in the time. But, if both servers have a precise time source, it is possible to see the effects of network traffic as a positive or negative time offset between these NTP servers.

A NTP server was built using a JJY receiver as time source. It gets the Japanese standard time (JST) through 60 kHz radio waves and was modified to increase accuracy by the addition of a PPS (pulse per second) signal out port. NTP uses this PPS signal to calibrate their clock. This server was configured to synchronize with some other NTP servers accordingly to the networks specified to analysis. A perl script, activated by crontab every 5 minutes, was used to acquire data from the NTP server and put it into a database. Then network traffic graph was built from that database using rrdtool and shown in a webpage.

### **Results and Discussion**

Using a costless and easy-to-install JJY receiver plus a few scripts and the rrdtool, it is possible to measure any network traffic flow, including the direction of congestion without access to routers, switches or any specialized device. Furthermore, it also makes it possible to measure the traffic between locations even outside the local network, which by usual ways is quite difficult or even impossible. However, care should be taken in order to avoid misinterpretation of data. A good understanding of the network topology is essential to correctly comprehend the generated reports.

GMS-5 衛星画像は、雲の分布、高さ、高度別の風の状況、海面の温度分布の情報がえられる他、台風、前線、気圧の波などの動向をとらえ解析するための資料として重要なものである。高知大学菊地研究室でも Web ページ高知大学気象情報頁において GMS-5 衛星画像を取り扱った情報を発信しており、そのデータ量はおよそ 200 GB 以上に及ぶ。

GMS-5 衛星画像情報発信の手法として雲画像と気象記録を利用した統合データベース構築を試みた。情報源には気象情報頁に蓄積されている日本付近の雲画像と、日本気象協会発行の「気象」に毎月掲載の天気図日記を気象記録として利用した。構築手順として、データベース部には PostgreSQL を利用し、天気図日記から日付、単語、風速や降雨量等で表される天候状態と、衛星画像に関する情報を、PHP スクリプトを用いてデータベースに挿入した。情報検索は、Apache と PHP を用いて、Web 上から日付、単語、天候状態のような検索キーワードを入力後、PHP スクリプトで AND 検索となる SQL 文を生成し PostgreSQL 上で実行することで、条件を満たす天気図日記と衛星画像を取得する。統合データベースでの応用実験として、カルマン渦列等の天気事象が起こった日の気象記録から特定の単語を選び検索をかけることで同様の天気事象をもつ衛星画像が得られることを検証した。

また、気象衛星画像検索データベースの拡張として気象衛星画像間の類似性からの検索を試みた。気象衛星画像の特徴をデータベース化するにあたり、気象衛星画像の形状解析を行った。解析手順は、①画像の雲域部分とそれ以外の部分とで 2 値化、②2 値化された雲域に対して八近傍収縮膨張による図形変換処理、③雲域の輪郭線を八近傍に基づき順に追跡し連続した座標の集合として気象衛星画像の特徴を抽出、の順で行った。抽出した座標は多角形の幾何データ型として気象衛星画像の撮影日時と共にデータベースに格納した。指定された日時の気象衛星画像の幾何データをデータベースから取得後、データベースに対し幾何演算を行うことで類似画像が得られることを目標に今後のアルゴリズム改良が必要である。

# 光量子ゲートの基礎

情報科学専攻 情報処理学講座 前田 薫孝

コンピュータの構成要素の半導体素子をどんどん小さくすると、量子力学の法則に見られる電子の波動性が強まり、現在の古典物理学の法則に従った素子の動作原理ではコンピュータは使えなくなる。そこで、その問題を逆手にとり、量子力学の性質をそのまま素子の動作原理として利用し計算させる「量子計算」の分野が発生し、従来のコンピュータの能力を超越するであろう量子コンピュータについての研究が始まった。

この量子コンピュータを実際に実現するためには、従来の古典コンピュータにおいて全てのブール論理 (AND、OR、NOT など) が、トランジスタを用いて表現できるのと同様に、量子コンピュータにおいてもある量子状態を利用した qubit による量子論理ゲート (1 qubit の位相シフトと 2 qubit の制御 NOT) が必要であり、様々な実現方法の研究が行われている。この論文では、量子コンピュータの実現のための量子論理ゲートの基本的な理論と、その量子論理ゲートの動作原理を検証していく。

量子コンピュータに必要とされている制御 NOT を、異なる 2 つの 2 準位系原子を用いて、図 1 に示すような 2 つの原子の間に働くクーロン相互作用による準位の変化と、図 2 に示すような光パルス照射によって起こるラビ振動を利用し量子状態を変化させ実現させる。

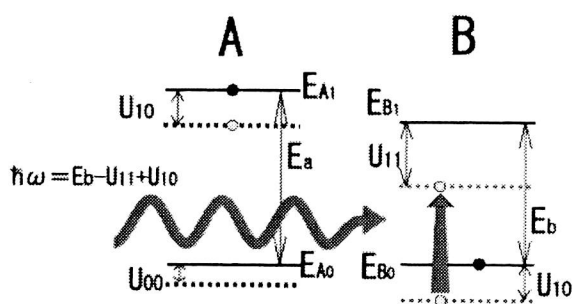


図 1 クーロン相互作用による準位の変化

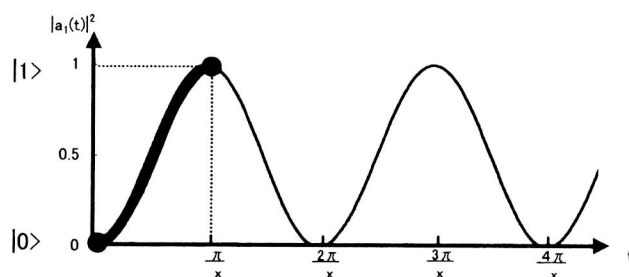


図 2 光パルス照射によって起こるラビ振動

今後の課題として、コヒーレンス時間の保持、クーロン相互作用以外の摂動によるラビ振動の減衰、標的の準位差に等しいエネルギーを持った光パルスの周波数が実現可能な範囲かどうかの検証があげられる。

## 有限束の分類

有限束を調べる上で、特に束の個数に注目した。今回は元の数が 8 個までのものについてその個数を求めた。結果は以下の様なる。

元の数	1	2	3	4	5	6	7	8
束の個数	1	1	1	2	5	15	53	222

これを求めるに際し、元の数が 5 以下の場合は束の定義より直接導き出せる。元の数が 6 以上についてハッセ図による分類によってその個数を求める。一般に元の数が  $n$  の束の個数を  $h(n)$ 、その中で極大元を  $a$  個、極小元を  $b$  個持つものの個数を  $h(n;a,b)$  とすれば、

$$h(n) = \sum_{a,b} h(n;a,b)$$

さらに、極大元かつ極小元の個数が  $c$  である束の個数を  $h(n;a,b;c)$  とするとき、

$$h(n;a,b) = \sum_c h(n;a,b;c)$$

また、 $c$  が正の数のとき

$$h(n;a,b;c) = h(n-c;a-c,b-c;0)$$

が成り立つ。

これらの式から  $h(n)$  を求めることは  $h(n;a,b;0)$  を求めることに帰着する。

以下に  $h(n;a,b)$  ( $n$  が 8 以下) の表を記す。

1	2	3	4	5	6	7	8
1	1	1	1 0	1 1 0	2 2 1 0	5 6 3 1 0	15 21 12 4 1 0
		0 1	1 1 0	2 3 1 0	6 9 5 1 0	21 35 22 7 1 0	
			0 0 1	1 1 1 0	3 5 3 1 0	12 22 17 5 1 0	
				0 0 0 1	1 1 1 1 0	4 7 5 3 1 0	
					0 0 0 0 1	1 1 1 1 0	
						0 0 0 0 1	1 1 1 1 0
							0 0 0 0 1

また、次のことも分かる。

$$h(n;1,n-3;0) = 1 \quad (n \geq 4)$$

$$h(n;2,n-4;0) = n-4$$

$$(n \geq 6)$$

# Development of Chinese Input System for Mobile Phone

The using Internet through mobile phone is becoming an increasingly important component of modern day communication and information change. Critical to satisfying the text input and display needs of Chinese language expect to encounter when they input text, but the situation is even worse for Chinese where thousands of characters are used and typing recognition is extremely difficult. To solve this problem, we approach to develop the Chinese input system for mobile phone that is able to input rapidly and easily by using Chinese Pinyin and Chinese conversion technique then achieve it as utilize e-mail and web page into mobile terminal.

The most common input methods are the ones that let user type text in Chinese language that use thousands of different characters on a regular-sized keypad. The text is typed in a form that can be handled by keypad, for example, in a Roman form, and then converted into the form that's really intended. Typically a sequence of several Chinese characters needs to be typed and then converted in one chunk, and conversion may have to be retried because there may be several possible translations.

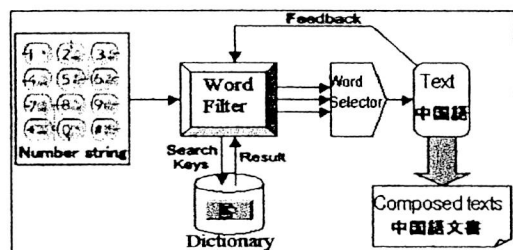


Fig. 1 Input System Architecture

Fig.1 shows an overview of text input systems. It shows how a Chinese text is composed using a standard keypad. Roman character strings are first converted to Chinese which represent the pronunciation of Chinese words, and then they are converted to Chinese characters by a Pinyin converter. Since multiple Chinese characters often have the same pronunciation, the user must choose the correct one by using the selector.

**Filtering Step** First, a user provides search or input keys for a word he/she wants to enter. Search keys can be the spelling, pronunciation. As soon as he/she enters search keys, the system dynamically uses the keys to look for the word in the dictionary and shows candidate words to the user for selection.

**Selection Step** Second, the user selects a word from the candidate list and the word is placed in the composed text. Next input words are predicted from the context and are used in the next filtering step.

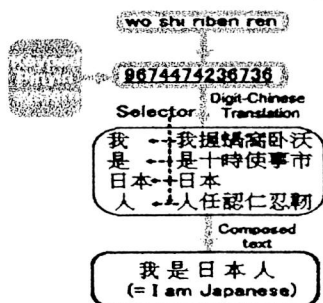


Fig.2 Input System Example

When input string consists of more than one word, the Conversion System looks up each word parts in the words-selector associative list and delivers each part to the selected separately. Following example, a string consists of Pinyin Chinese part:

To input character '日', first input pinyin 'ri' by typing in '74' and then select it from the candidate list '日 和 鉅 駟' and to input '本', input 'ben' by typing in '236' and the select it from '本 奔 笨 笨 奔 盆 沝 濟 鑄'. So we input digit '74236' to convert a word "日本"

This manuscript has introduced the possible input system for mobile terminal by utilizing Pinyin input system. We have appropriated the Digit-Chinese dictionary for mobile terminal conversion-system. The dictionary is possible to search and converse digit to Chinese character. The experimental result shows that this kind of conversion method achieves the satisfying conversion accuracy and hence can be used in applications where only a small number keypad is available for the Chinese character input.

## 日経平均株価データのフラクタル解析

情報科学専攻 情報基礎学講座 王 冬青

株価は毎日時々刻々不規則に上下して、投資家を悩ませている。株価は統計的に、次のような性質を持っていることが知られている：

株価の変動は、時間の単位によらず、統計的に相似であり、時間に関するフラクタル構造を持つ。換言すれば、株価の頻度分布がデータのサンプリングの区間によらず自己相似性を示すこと、即ち、フラクタル特性を持っていることである。

時間的な相似性を用いて、未来の株価予想を行う試みは未だ成功した例はない。また、株価変動のフラクタル的な特徴の詳細についての知見も不足している。

本研究では、12年間(1991.1.4-2002.12.30)の日経平均株価データを用いて、この株価データのフラクタル性を解析する。まず、標準的なデータ解析として、株価の原始データの頻度分布とフラクタル次元  $D=1.4784$  を決定した。次に、株価変動の差分データの解析を行ない、ここにもフラクタル的な性質があることを示し、合わせてフラクタル次元を決定した。フラクタル次元の定量的な決定には次の手法を用いた：サンプルデータを  $I$  日飛びで測ったときの変動量の絶対値の総和

$$N(I) = \sum_{j=1}^{n-I} |k_j - k_{j+I}| \times \frac{1}{n-I} \times \frac{1}{I} \times \frac{n-1}{I}$$

の対数  $\log N(I)$  を  $\log I$  に対してプロットしたとき得られた直線の勾配をフラクタル次元とした。株価と株価の1,2,5,10日間の差分データのフラクタル次元の大域的挙動を求め、株価の多重フラクタル構造についても考察した。

# 気象データの非正規性の定量化

情報科学専攻 情報基礎講座 宗 石 有 平

気象データの中でも気温は最も日常的な気象情報のひとつである。気温は一年を周期として変動しており、日本ではその周期は四季によって細分され夏と冬の間では約30度の気温差がある。しかしながら、人は一日単位の気温差を肌で感じ取ることは一年を通じてもそれほど多くはない。そのことから、気温というものは人にとって心地よい変化をしているのではないかと考えられる。そこで、本研究では気温変化がどのような性質を持っているのかということについて、秋田県の12年間の平均気温のデータ(1990年1月1日~2001年12月31日)を用いて、以下の研究を行った。

1. 時系列の  $n+1$  vs  $n$  の回帰写像の作成とその性質の解析。
2. 気温変動の差分データのフラクタル次元を解析し、 $D = 2.0058$  を得た。これは、ほとんど乱数であることを示す。
3. 気温の原始データのフラクタル解析を行い、その次元として、 $D = 1.259$ (長周期成分)、 $D = 1.863$ (短周期成分) を得た。これは多重フラクタル構造の存在を意味し、その境界点は約16日であり、24節気との対応を予想させる。
4. フーリエ変換を行い、気温データのパワースペクトルを調べた結果、一年周期、半年周期、四季節に対応する3ヶ月周期の他に4ヶ月間隔でもある種の周期変動があることを発見した。しかもこれは、四季節(3ヶ月)周期より約3倍の強度を持っている。また、パワースペクトルからこれらの周期成分を除いたものは一様なスペクトル分布を示し、一様乱数に類似していることが判明した。
5. 気温の1日差分データの頻度分布において、データの非正規性を定量化するために、正規分布関数

$$P(x) = C \exp[-ax^2]$$

から誘導される疑似分布関数

$$Q(x) = C \exp[-ax^\alpha]$$

を導入し、その指数  $\alpha$  を決定した。ここで、 $\alpha$  は実際のデータ分布と関数  $Q(x)$  との誤差の2乗積分が最小となるように定めた。その結果、指数  $\alpha = 1.472$  を得た。また、データ分布は分布の左右両側で非対称であるため、左右各側別々に決定し： $\alpha_{右} = 1.641$ 、 $\alpha_{左} = 1.256$  を得た。

また、今後の課題としては、時系列の  $n+1$  vs  $n$  の回帰写像からの有効なデータの抽出、4ヶ月周期の意味すること、非正規分布の指数  $\alpha$  を用いてその他のデータ(地域や季節ごとの気温データ、気温以外のデータ)について調べることなどである。