

「デジタル台風」プロジェクト：大規模時系列画像データベースのマイニングにむけて国立情報学研究所
北本朝展

気象衛星画像が毎日の天気予報に登場するようになってから、台風の渦巻き状の雲は誰もが知る存在となった。しかしその形状を仔細に観測してみると、台風ごとに多様な形状を示すことがわかる。その形状の多様さは見ているだけでも興味深いものであるが、実際の応用においても、台風の形状と台風の勢力には相関があるとの仮定が台風の勢力推定に利用されている。つまり台風の形状には、台風の特徴が埋め込まれて現れていると考えてよいだろう。このような複雑な形状と、個々の台風の気象学的な特徴とはどのように関係しているのだろうか、という疑問を出発点として「デジタル台風」プロジェクトはスタートした。

プロジェクトの基本的な方針は、1)関連データの網羅的な収集、2)データベースの構築、3)データベースからのマイニングである。まず最初に、台風経路データは 1951 年以降の 1400 個以上の台風経路、静止気象衛星画像（ひまわり・ゴースト）は 1995 年以降の 5 万件以上の台風画像を網羅的に収集し、現在も更新を続けている*1。これらのデータを次にデータベース化し、さまざまな方法で検索したり解析したりすることが可能な状態に整備した。そして、このデータベースに隠されたパターンを、アルゴリズム的に発見しようというのが、プロジェクトの最終的な目標となる。ここで重要となるのが、画像パターンを扱うための画像解析・パターン認識技術であり、また時間的に変動する対象を扱うための時系列解析技術である。そこでこれらの領域を中心に研究を進めてきたのがこれまでの流れである。

これらの基本方針は、著者が今後推進しようと考えている「メテオ・インフォマティクス (meteo-informatics)」という方法論に基づくものである。その基本的なアイデアは、データベースに基づく科学的探究というスタイルを、気象学に持ち込もうというものである。以前の科学的探究では、少量のデータを大事に利用するというスタイルが主流を占めていたが、現在では大量のデータを取得しデータベース化した後、そこから必要なデータを取り出すというスタイルが増加してきた。両者の研究の本質的な違いは、特定の研究対象をまず選んで観測するか、網羅的な観測の後に必要な研究対象をコンピュータ上で選ぶか、という違いにある。センサ技術とコンピュータ技術の発達により、現在ではむしろ後者の方が能率的な場合も増えてきたことが、いわゆる「X-インフォマティクス」研究の隆盛の背景にある。このような方法論を支えるためのデータ品質の一貫性保持、データ検索言語の開発などのテーマも、本プロジェクトの重要な研究課題である。

本プロジェクトはこのようにデータ中心主義のスタイルを採っているが、観測されたデータを後付けで説明するだけでは研究として不十分であるとも考えている。観測されたデータには常に「たまたまそうなった」という側面があり、観測データを絶対視して説明するだけではモデルとして不十分であろう。換言すれば、観測されてデータとなった値は何らかの確率変数の実現値に過ぎず、観測されたデータを生成するプロセスに潜む確率的な構造を推定することが、本来やるべきことであると考えている。本プロジェクトでは台風の雲パターンを

特徴空間中の軌跡と捉え、そのダイナミクスを確率的にモデル化することでこの問題を追究する計画である。

上記の方針に基づき、現在までのところ、台風パターンの類似性検索とその時系列比較、台風雲パターンによる発達 / 衰弱の分類、台風ライフサイクルの類似時系列検索などについての結果を得た。その結果の詳細については発表で紹介する。これらのアルゴリズムはまだ実用的な台風解析に使えるものとはなっていないが、その大きな原因は、台風雲パターンの特徴解析と類似性に関する研究が不十分な点にあると考える。今後はこのテーマをさらに深く掘り下げることで、大量データを根拠とした新たな発見に結び付けていきたい。また、気象学はもともとシミュレーションが特に盛んな分野であるが、データベースとシミュレーションは、帰納過程と演繹過程というコインの表裏のような関係にあり、本来はお互いにフィードバックしあえる存在でもある。こうした二つの分野の融合も新たな発見に結び付く可能性がある。

最後に、研究のデモンストレーションとして作成したデータベース「デジタル台風」は、準リアルタイム更新の台風データベースとして多くの利用者を集めており、累計のページビュー数は 920 万に達している（2004 年 10 月中旬現在）。このような活動は研究としては評価されにくい面もあるが、社会への還元として続けていきたいと考えている。

<注釈>

1：その他にアメダスデータやニュース記事、多数の人々による情報発信なども収集しているが、本発表では軽く触れるにとどめる。

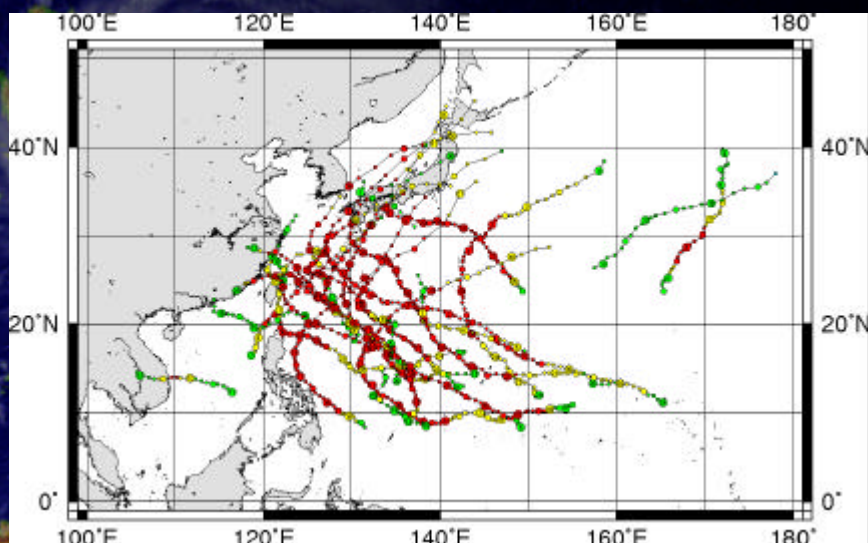
デジタル台風」プロジェクト: 大規模時系列画像データベースの マイニングにむけて

国立情報学研究所

北本 朝展

<http://www.digital-typhoon.org/>

2004年の台風経路図 (速報)



災害が続発した2004年

- 相次ぐ台風の来襲
 - －日本上陸個数：10個（従来記録 6個）
 - －沖縄接近回数：14回（従来記録 12回）
 - －台風 23号 死者不明 89人（兵庫県 23人）
- 規模が大きいものが多く、しかも勢力が弱まる前に日本列島に到達した。
- 相次ぐ台風による大量の降雨が、新潟中越地震の土砂災害の原因にもなった。

阪神・淡路大震災から約10年

- 災害時における情報システムの役割とは何か？
- インターネットは飛躍的に普及したが、それによって状況は改善したのか？
- 公共機関が発信する1次情報の充実に比べ、個人が発信するローカルな1次情報の集約・データベース化は？

新潟中越地震における 緊急時情報システム

	1次情報	2次情報
公共的	公共機関からの情報 発信は大幅に増加し 有用な情報源に成長。	マスコミ報道は強力 な情報源。バイアス もあるが改善も。
個人的	安否情報については NTT等が活躍した。し かしその他は掲示板 等による情報交換が 主流。期待のウェブ ログは十分に役割を 果たしているか。	リンク集やコピーペ ーストサイトが主流 のよう。10年前から それほど進歩してい ないような印象を受 ける。

緊急情報のための データベースシステム

- 緊急的な情報の流れを集約し データベースを即時的に更新。
- 現在の状況と過去の事例とを、即座に比較できるようにする。
- この問題の解決には、マルチメディアデータベース技術が不可欠である。

ウェブサイト「デジタル台風」

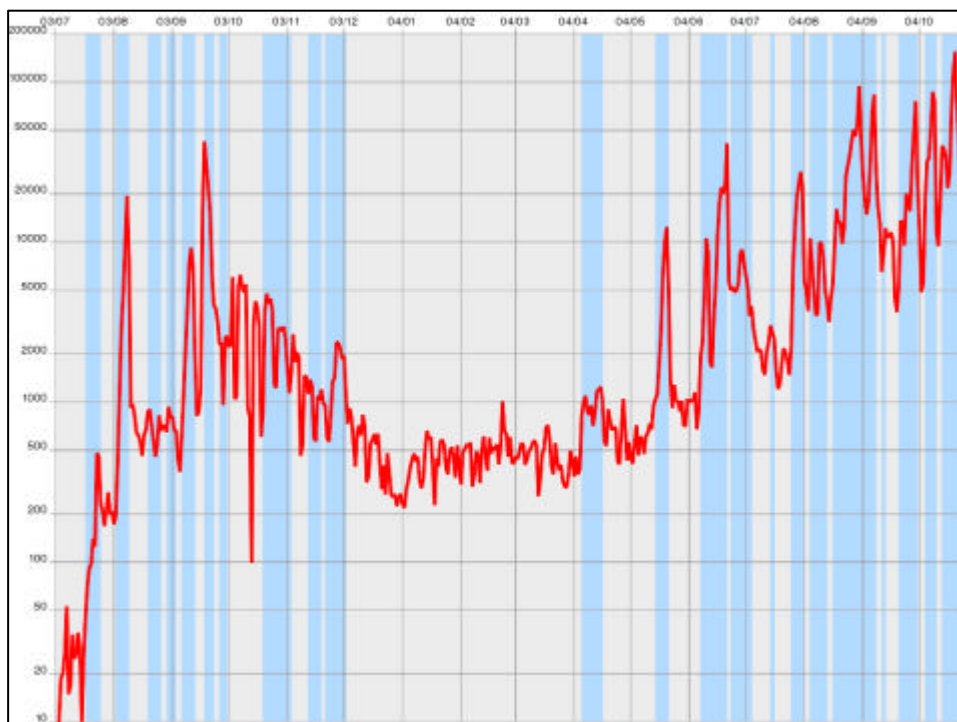
	1次情報	2次情報
公共的	気象庁からの台風情報と気象衛星画像の処理とデータベース化。	ニュースサイト記事の自動言語処理と組織化。
個人的	多数の個人が発信するウェブログからのトラックバックを集約したデータベース。将来的には携帯機器対応。	台風ニュース・ウェブログによる台風関連情報の概要。

<http://www.digital-typhoon.org/>



「デジタル台風」の概要

- ・台風情報に関するポータルサイト。
- ・台風画像と台風ニュースをほぼリアルタイムで更新しデータベース化。
- ・大部分の情報を日本語と英語で提供。
- ・1日のトップページビュー:
 - Weekday 5,000-20,000
 - Peak 50,000-150,000
 - サイトページビュー 最高420,000/日**



なぜ台風なのか？

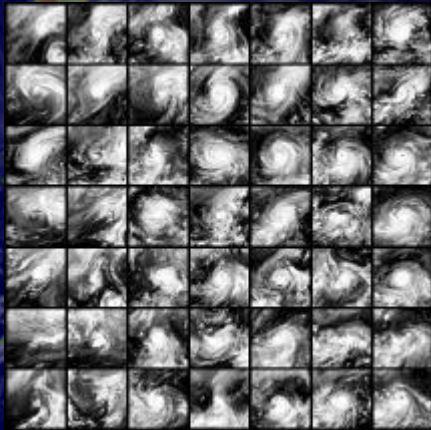
1. 地理的・時間的に複雑なふるまいを見せる実世界の重要現象。メテオインフォマティクスのモデル現象となる。
2. 事例に基づく台風解析と予測に利用。
3. 情報学の研究者としてできることに取り組むが、その結果が気象学的に意味があるかどうかを常に念頭におき、Cross Communityな研究を目指す。

台風画像コレクション

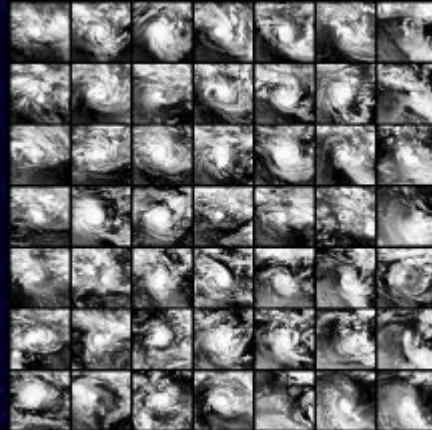
2004年10月現在	北西太平洋	オーストラリア
	ベストトラックデータ	
担当機関	気象庁 (JMA)	豪気象局 (BOM)
緯度範囲	赤道以北	赤道以南
経度範囲	100~180	90~170
台風画像コレクション (累計 53,150)		
シーズン数	9 シーズン	6 シーズン
台風系列数	232	86
台風画像数	41,278	11,872
1系列あたり画像数	53~433	25~513

台風画像コレクション

画像データコレクションのK-平均クラスタリング



北半球



南半球

網羅的画像コレクションとは

- データの重要性に関する予見を排除し、できるだけ大規模にデータベース化。
- 重要なデータは、ユーザがデータベース上で選択する。
- 高品質が望ましいのは当然だが、均質性を保つ品質管理にも留意すべき。
- 地図投影法のような重要な前処理については、事前の慎重な検討が必要。

前処理の一つ 地図投影法

- 地図投影法が満たすべき条件とは？
 1. 地理的位置によらず面積を比較できる。
 2. 円形の物体に対する歪みが小さい。
- 以上の条件を満たす地図投影法として、**ランベルト正積天頂図法**を選択。
- 上記の性質により、**複数の台風雲パターンの形態をダイレクトに比較**できる。

研究課題

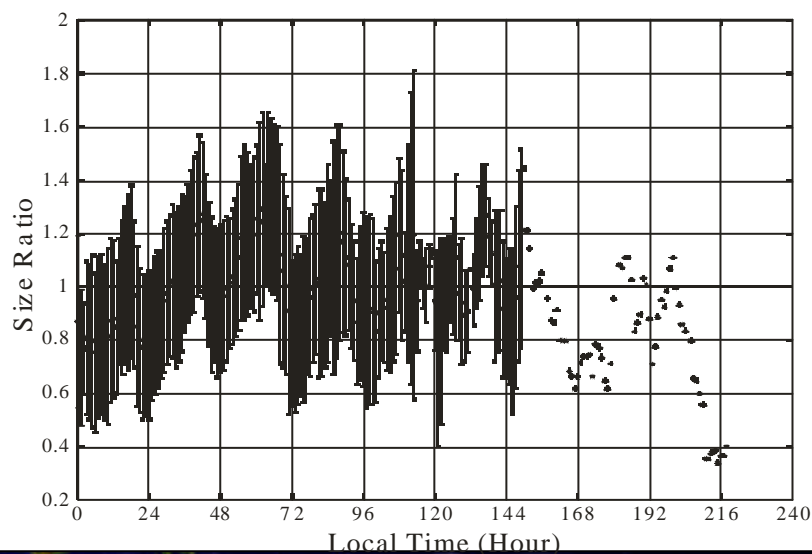
- **「台風の形態学」**が重要な研究課題。
- **画像認識および画像データベース、マイニング技術**を中心とする。
 1. 画像特徴から台風発達の予兆をどのように発見するか？
 2. 台風の時空間的发展パターンをどのように数理モデル化するか？

台風の形態学

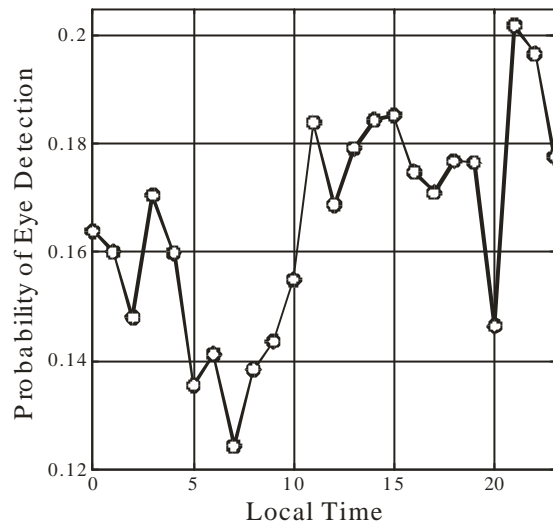
- 形態学的方法による台風解析（ドボラック法）。
- 専門家による気象衛星画像雲パターンの目視解析を基本に、他の観測データを加えて総合判断する。



台風サイズの日周変化



台風眼検出確率の日周変化



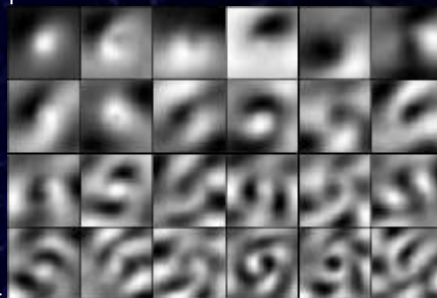
画像特徴空間

- 画像そのものをデータとするのではなく、**重要な画像特徴を抽出して空間を構成**し、その空間で検索などをおこなう
- **画像特徴空間の次元を小さくし、かつよい距離尺度を定義できれば、そこから有用な情報を発見することができる。**
- この部分は画像解析の**核心的な問題**。

固有表現に基づく画像特徴空間

- 枠取りした台風画像の地図投影
- 雲画像分類から雲量ベクトルの計算
- 主成分分析による次元削減
- 特徴ベクトル算出

北半球の台風に対する固有台風表現

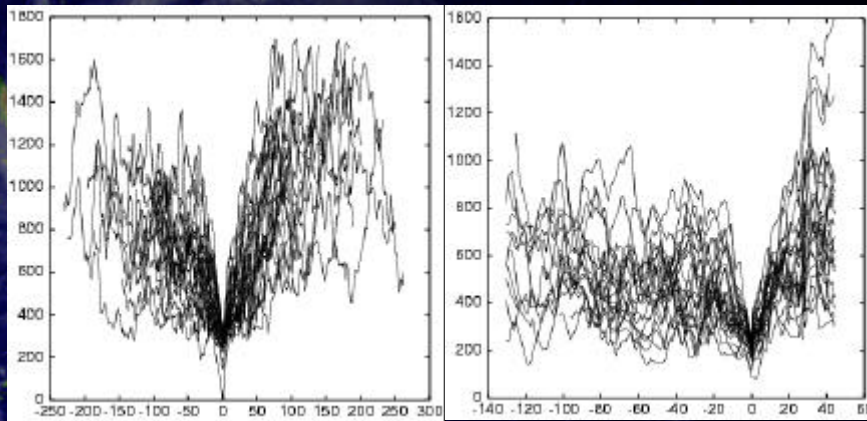


類似性検索

- 過去の事例を類似性に基づき検索。
- 画像特徴空間の選択により、適切な事例が検索できるかが決まる。



カオス的性質の現れ



- 適切な画像特徴空間をとってもなお、データ間の距離は初期値から拡大する。

Technical Result (1)

- Problem Statement:
Classification of Developing / Weakening Typhoon
- Goals:
 - Recognize some “signs” of cloud patterns before it actually occurs?
 - Solve a relaxed problem → classify moderately developing / weakening typhoons from cloud patterns.
- Threshold = $|\Delta 10\text{hPa}| / 24\text{hour}$



Rapid Deepening

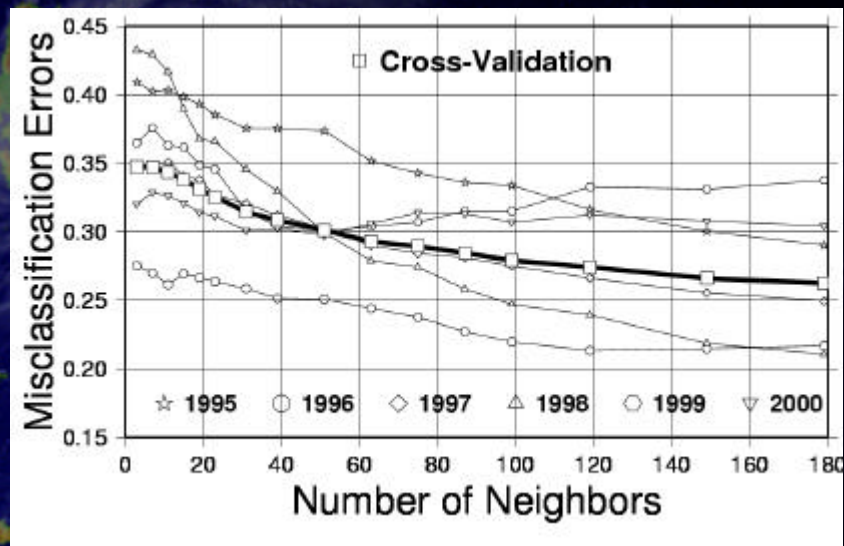
- **Rapid deepening** – intensifying +30-40 hPa in 24 hours.
- A very significant event from the viewpoint of **disaster management**.
- Meteorologically **the prediction of rapid deepening** is difficult.



Our Approach

- **Kernel PCA** -> Extraction of non-linear principal components.
- The effectiveness is still not clear.
- **Support Vector Classifier** -> This part is discussed in this presentation.
- Many thanks to Libsvm (V2.35) from Prof. Lin@National Taiwan University.

Classification by K-NN



Classification by nu-SVM (1)

Kernels	Nu	6-fold CV
POLY d=1	0.3	33.9
	0.5	26.9
	0.7	25.4
POLY d=2	0.3	33.1
	0.5	32.6
	0.7	34.0
POLY d=4	0.3	34.6
	0.5	35.0
	0.7	37.6

Classification by nu-SVM (2)

Kernels	Nu	6-fold CV
RBF $g=0.1$	0.3	28.4
	0.5	26.3
	0.7	25.6
RBF $g=1$	0.3	26.7
	0.5	25.8
	0.7	25.5
RBF $g=10$	0.3	26.2
	0.5	24.8
	0.7	25.8

Comparison

- From the result of K-NN, the **baseline classification error is about 25%**.
- The polynomial kernel achieved this performance with only a limited combination of parameters.
- The RBF kernel is relatively stable and showed **comparable performance**.
- This error rate (25%) may be close to **Bayes error rate?**



Summary

- We can classify and capture the signs of typhoon development from image features.
- Advanced image features such as
 - Shape, texture-based image features.
 - Dynamic (temporal) features.
- Advanced kernel design such as a HMM-based kernel.



Technical Result (2)

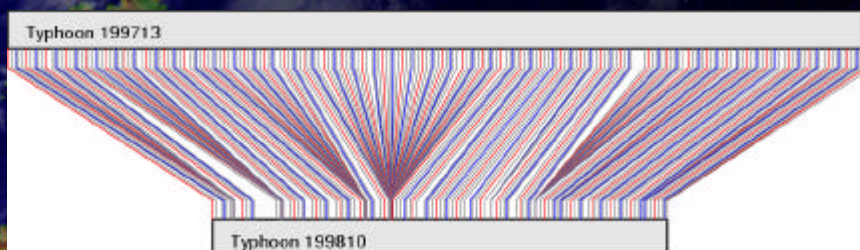
- Problem Statement: Multiple alignment and meaningful fragment extraction of typhoon image sequences.
- Goal:
 - Extract meaningful temporal fragments based on pattern similarity.
 - Natural boundaries between fragments are usually unknown.

Our Approach

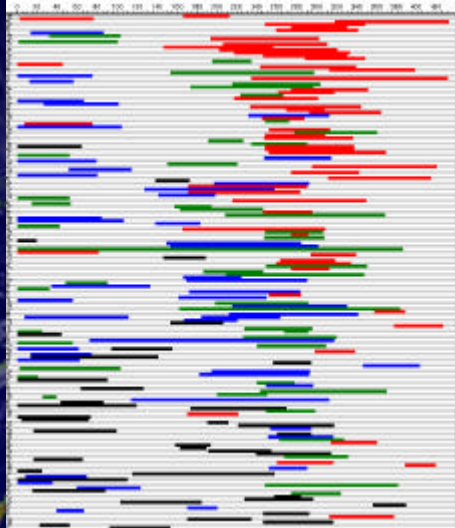
1. Apply **global alignment** between the pivot sequence and another sequence.
2. Apply **multiple sequence alignment** to find similar temporal patterns.
3. **Dynamic programming** and **clustering** are key components.

Pair-wise Alignment

1. **Global alignment** of two typhoon sequences and extract fragments.
2. **Dynamic programming** is a typical solution for computing the **minimum matching score**.

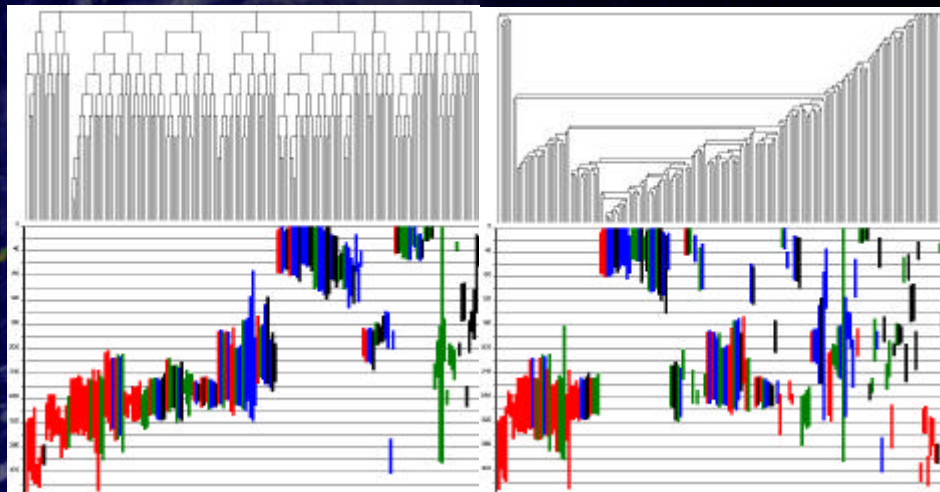


Alignment with the pivot sequence



- Alignment of sequences with the pivot sequence.
- Parallel matching fragments are extracted.
- Similar sequences are shown upper.

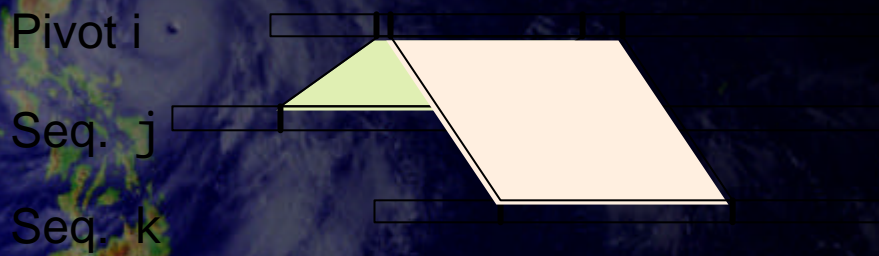
Hierarchical Clustering



Ward method

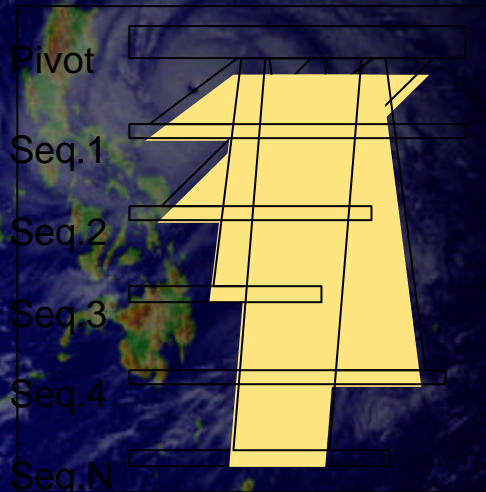
Min-Dist Method

Distance Measure between Fragments



1. Large overlap on the pivot sequence?
2. Ratio of matching scores is small?

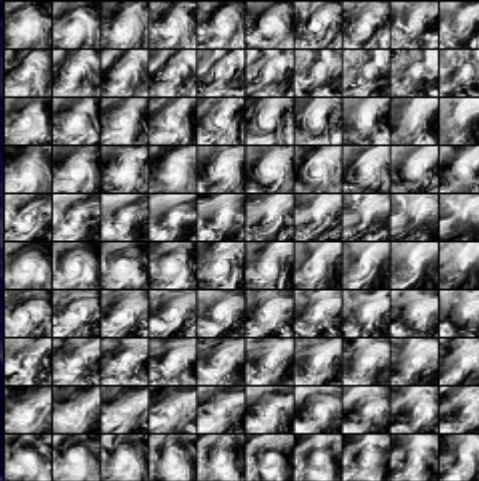
Consensus Fragments



- Global alignment results are combined.
- Average position is computed as consensus fragments.

Matching Results

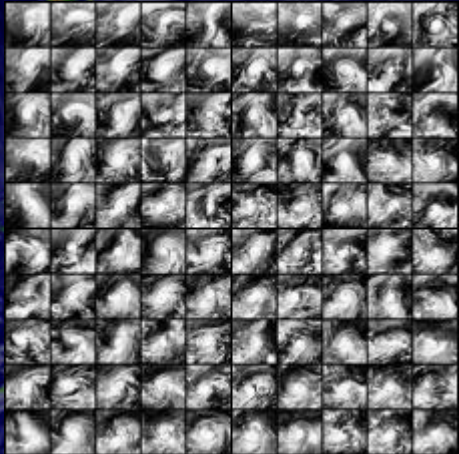
- Fragment sets that illustrate the diversity of typhoon life cycle.
- Comparison of cloud patterns is intuitive after alignment.



Summary

- Similar temporal fragments can be extracted using **dynamic programming** and **clustering techniques**.
- Advanced **distance measures** and **alignment schemes**.
- Advanced **clustering methods** for "semantically" similar fragments.

クラスタリングと時間分布



K-平均クラスタリング



時間分布

データベースに基づく科学研究

- 生物情報学
 - ヒューマンゲノムプロジェクト
- 天文学
 - スローンデジタルスカイサーベイ
- X-Informatics と呼ばれる分野を形成。
- 計算的アプローチと情報学的アプローチ

メテオ・インフォマティクス

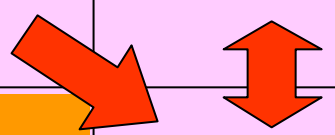
- メテオ・インフォマティクス (気象情報学) があってもよいのでは？
- データの検索や比較を繰り返すことにより、知見を発見し蓄積していく。
- どんなデータベースが必要か、データからどんな知識が発見できるか？
- 数値シミュレーション等の演繹的アプローチに対し帰納的アプローチを追究。

計算的アプローチと 情報学的アプローチ

- 気象学は例外的に (?) 計算的アプローチ (数値シミュレーション) が非常に発達し、しかも成功した分野。
- ENIAC以来、計算的アプローチが他を圧倒し発展してきた歴史がある。
- しかし、それだけでいいのか？

メテオインフォマティクスの領域

	データ	モデル
説明	事例解析 統計解析	数値天気予測
一般化	メテオ・インフォマティクス	大気 (地球) シミュレーション



事例」概念の拡張

- シミュレーション結果は厳密には「事例」ではないが、データに基づく方法の根本的な問題である事例不足を軽減できる可能性がある。
- リアルデータとシミュレーションデータを比較できるような、巨大科学データベースの設計について考えるべき？

二つのアプローチの融合

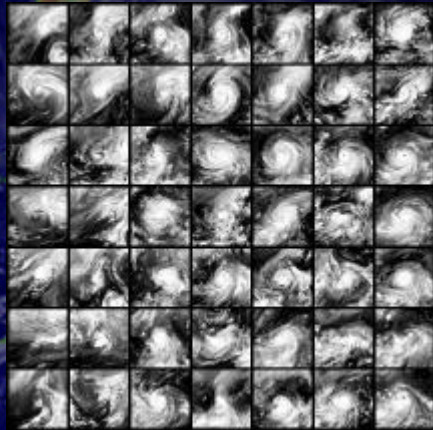
1. シミュレーション結果に類似した現象を過去のデータから検索する。
2. シミュレーション結果と過去のデータとは、どのように一致し、どのように異なるのかを分析する。
3. シミュレーションに使えるような観測結果や法則を、論文データベースから検索する。

今後の研究課題

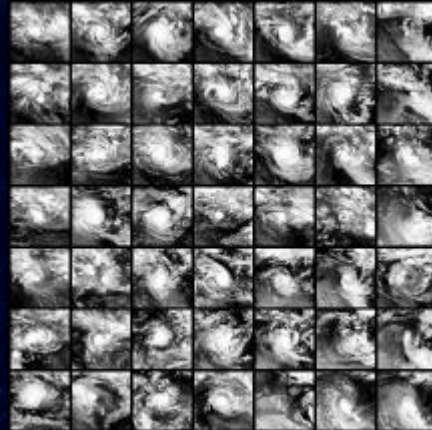
1. ダイナミックに変動する画像 (空間) パターンの時系列モデリング
2. パターン発見に適した問合せ言語を備えた高速画像検索エンジンの開発
3. 大規模な分散データ収集システムの構築と集約した情報の即時的組織化

比較台風学

画像データコレクションのK-平均クラスタリング



北半球



南半球

全世界台風情報ネットワーク

- 全世界に発生する熱帯低気圧の網羅的なデータベースを構築し、科学的な研究に役立てる。
- 全世界で台風を見つめる人々の出来事などを即時的に集約し組織化し配信するようなネットワークを形成し、災害情報などに役立てる。

おわりに

- 「デジタル台風」プロジェクトの概要とウェブサイトを紹介した。
- このプロジェクトは、メテオ・インフォマティクスモデルプロジェクトに位置づけることができる。
- 事例に基づく方法を現在は追究しているが、将来的にはシミュレーションに基づく方法と融合できるとよい。

謝辞

- GMS画像を研究目的にご提供くださる、東京大学生産技術研究所の安岡善文教授、喜連川優教授に感謝いたします。なおGOES画像は(財)気象業務支援センターによるものです。
- 台風中心位置等のデータは気象庁の観測に基づくものです。
- 本発表に関連する情報や画像などに関しましては、以下のウェブサイトをご覧ください。

<http://www.digital-typhoon.org/>