

オーロラの出現・形状の予測に向けた オーロラの観測画像の分類とデータベース化

田中孝宗*、佐藤由佳十、田中良昌十、池田大輔*

*九州大学 システム情報科学{府、研究院}

十 国立極地研究所

第4の科学：シーズから

専門：情報学

データの解析と利用基盤に関する研究

解析：テキスト(ブログ、Twitter等)、ログ、ゲノム配列、文学作品等

基盤：情報検索、マルチメディア検索、データベース、I/F、分散等

所属：システム情報科学研究院

国立情報学研究所

九大 附属図書館研究開発室

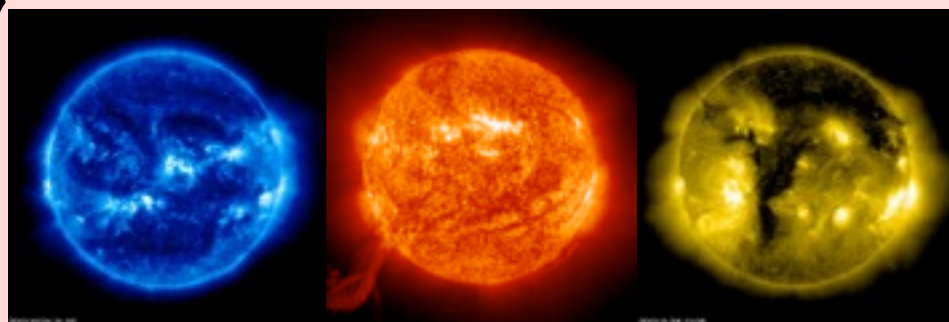
機関リポジトリ(文献の流通)→データリポジトリ(データの流通)

九大 国際宇宙天気科学・教育センター

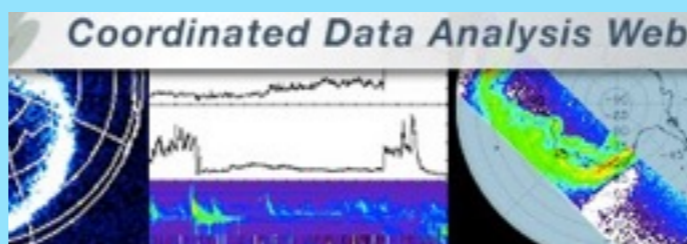
第4の科学：ニーズから

現象が巨大・複数の要素が相互作用し、様々なデータが存在

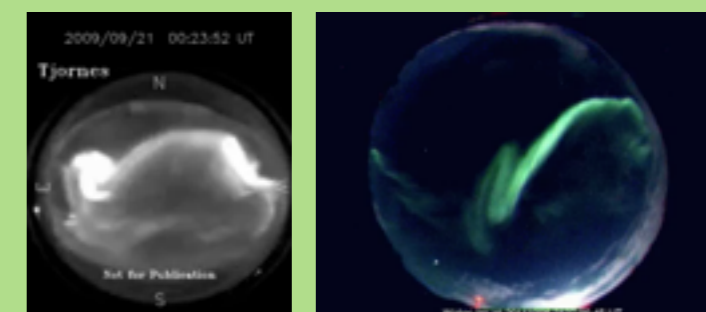
→ データ科学の専門家が必要



各周波数帯での太陽画像



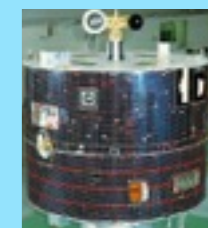
磁気圏磁場



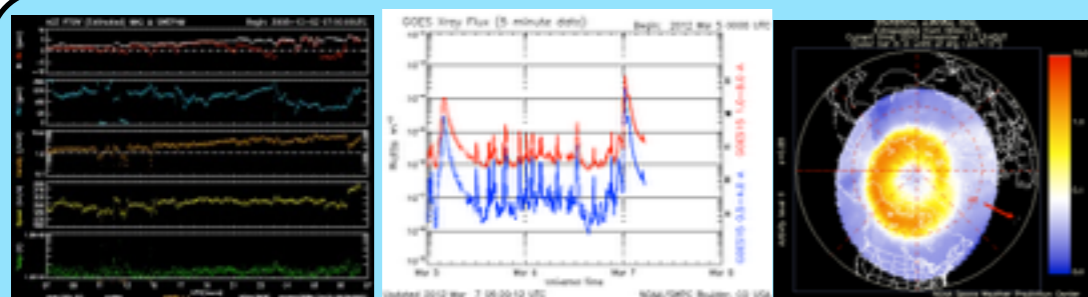
全天オーロラ画像

開始	最大	終了	Q	タイプ	場所/周期
0026	0032	0039	5	XRA	1-8A

太陽イベント



磁気圏尾部磁場



太陽風,高エネルギー粒子太陽X線,オーロラ活動指数



世界各国の地磁気,イオノゾンデ..

第4の科学って？

大量のデータを科学に直接利用する

例

文献から偏頭痛の潜在的な要因を特定(Swanson & Smalheiser'97)

ペアで出てくるキーワードを発見

ペアが出現する文から仮説を生成

(A, B), (B, C) => (A, C)

(ストレス、偏頭痛)、(ストレス、マグネシウム損失)、…

=> 偏頭痛にマグネシウムが効く

ジャック・アンドレイカ: 有望な膵臓がん検査

膵臓癌に関与する8000種類のタンパク質をリストアップ

4000 種類を調査し、検査に有効なものを特定

文献リポジトリを利用

研究の大きな目標

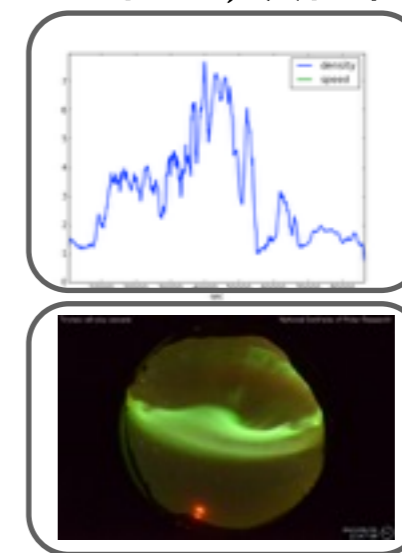
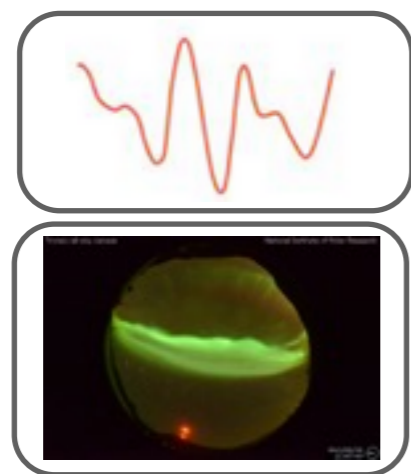
オーロラに関する超高層物理学のデータリポジトリ構築

リポジトリ上のデータを用いたオーロラ予測のモデル構築

物理学的な知識をなるべく用いない情報学的なモデル(分類など)

入力:時系列,画像データ

出力:類似時系列,類似画像データ



シンボル化

変化点検出

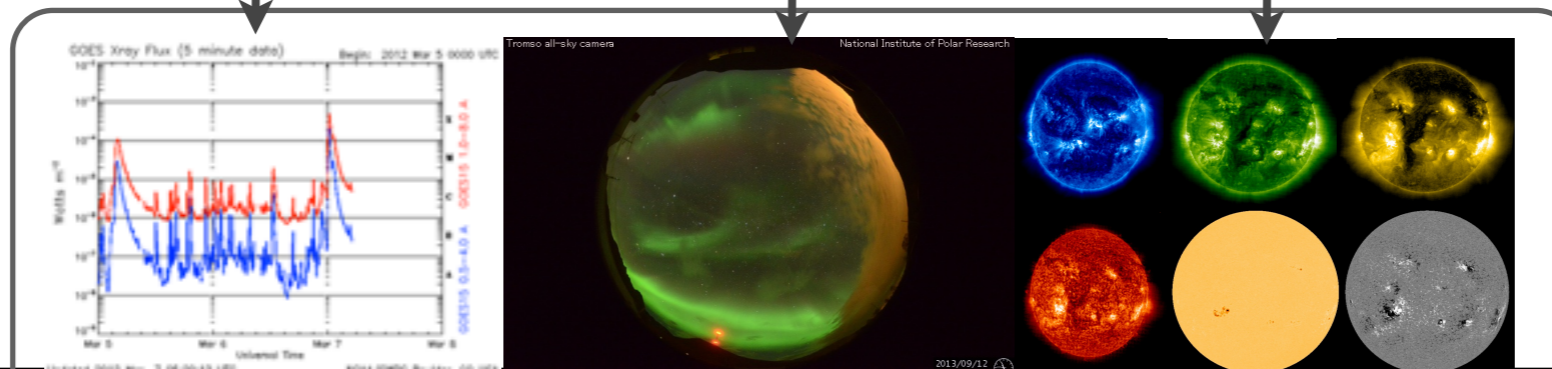
出現判別

色判別

形状判別

色判定

形状判別



研究の方針

機械学習、時系列解析を軸に進める

オーロラの有無による特徴の違いを自動的に抽出
オーロラと似ている部分的な時系列データを特定する

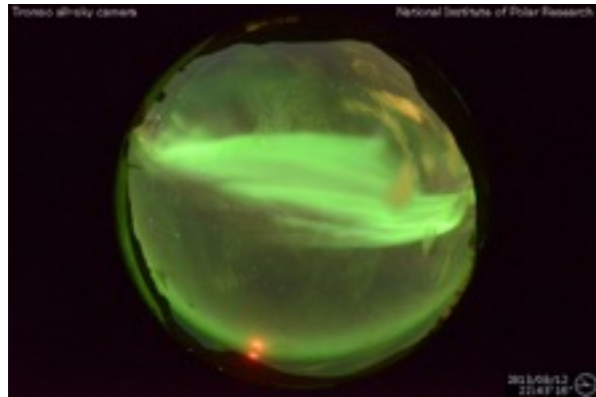
そのためには、まず、正解データが必要

いつどこでオーロラがでていたのか

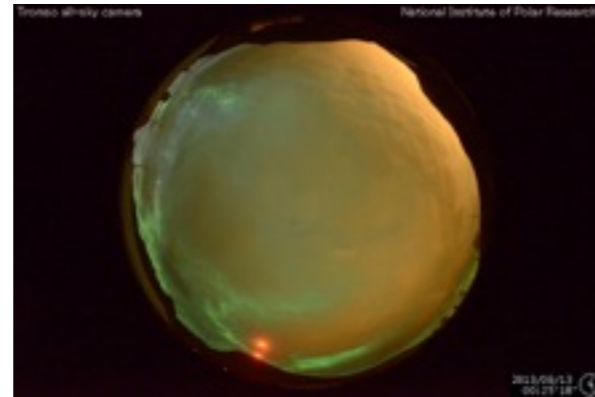
**地上からの全天画像を
機械的にラベル付け**

オーロラの出現判別

判別境界



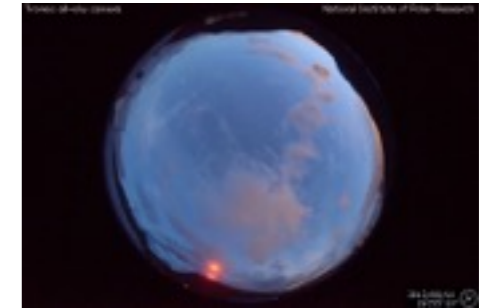
オーロラ有



雲の隙間に出現



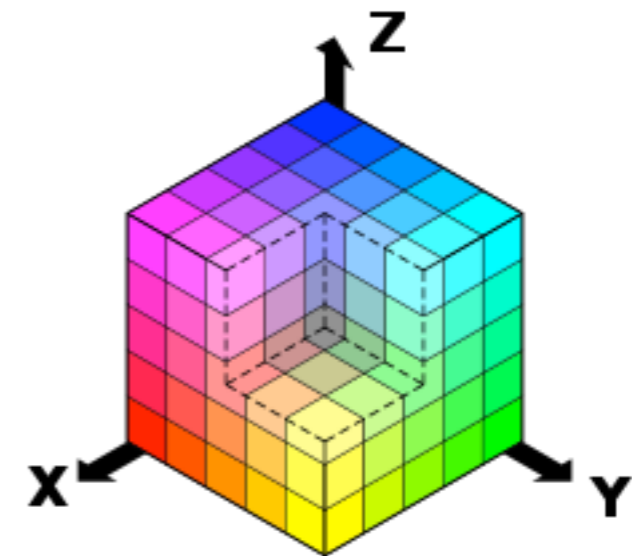
薄いオーロラ



悪天候で見えない
出現していない

オーロラの持つ**緑色の画素**を検出する

- ・ **RGB**カラーモデルは加法混色であり色の定義が複雑
→色の定義が容易な**HSV**カラーモデルを利用する。

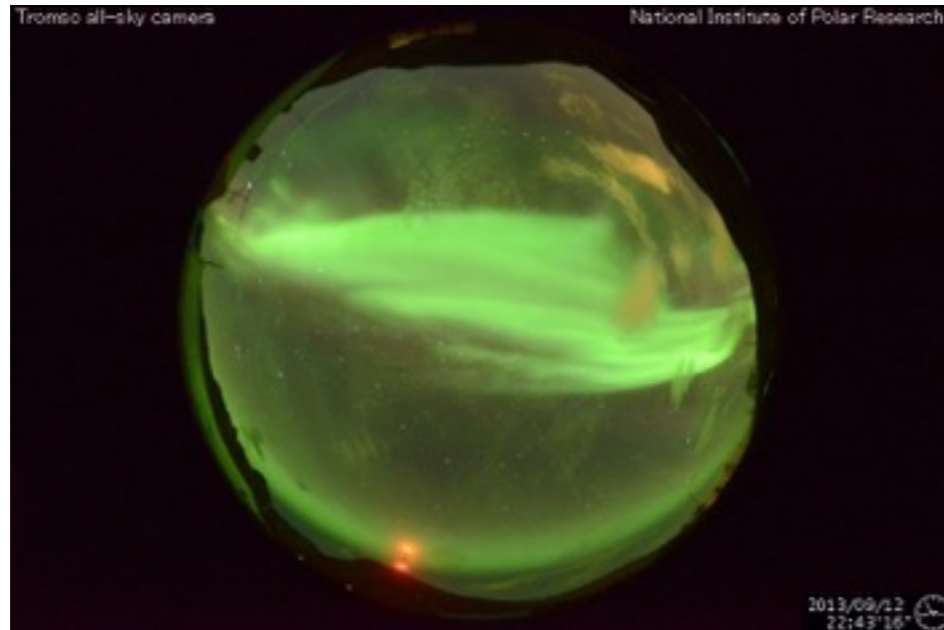


出典:<http://ja.wikipedia.org/wiki/RGB>

判別アルゴリズム

(今回は機械学習ではなくヒューリスティックな分類)

1. 各画像に対し、緑色画素の出現個数をカウント



カウント数 = 124463



カウント数 = 0

2. あらかじめ定めたしきい値以上あればオーロラあり！

実験

データ

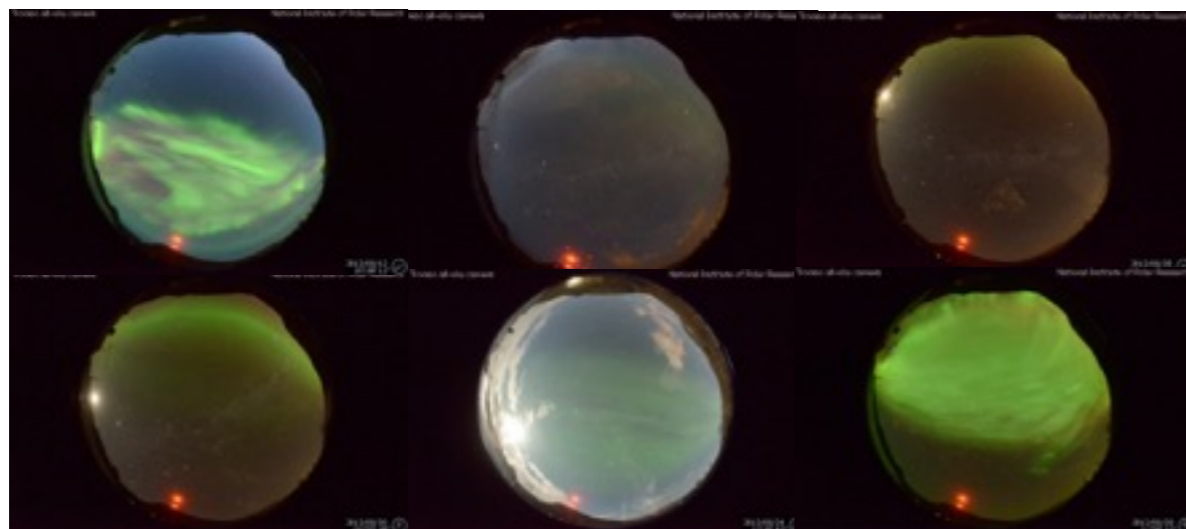
392枚の画像(720x480)

撮影地：トロムソ(ノルウェー)



評価方法

あらかじめ専門家にラベル付けしてもらった正解と比較



結果

識別率

	正答率	誤識別率
オーロラ有	94.3%	5.7%
オーロラ無	91%	9%

例

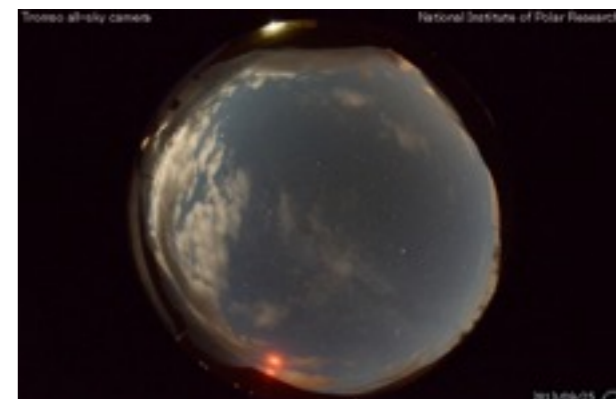
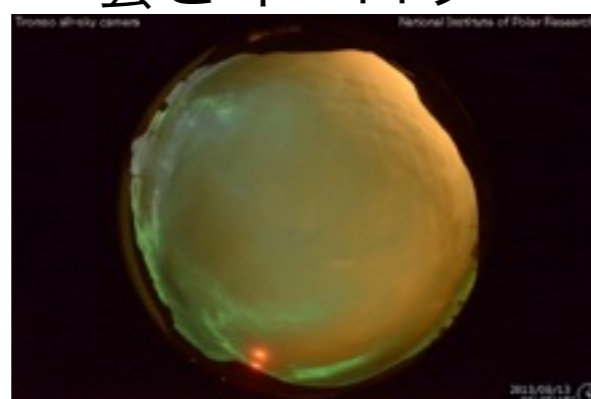
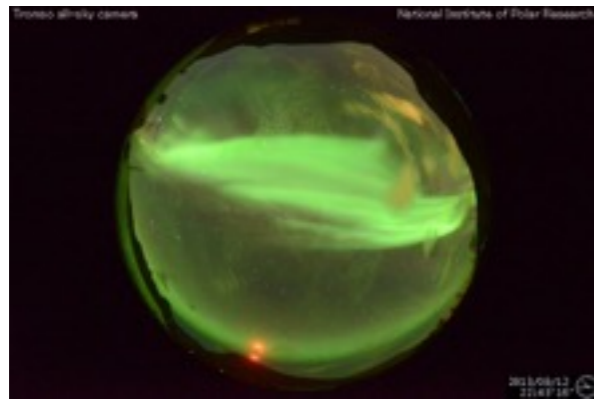
正答例

誤識別例

雲とオーロラ

薄いオーロラ

有



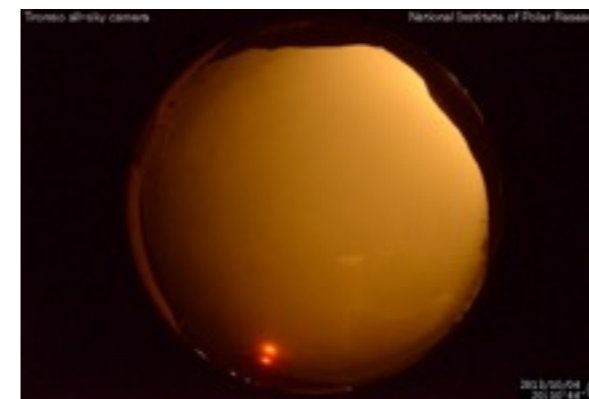
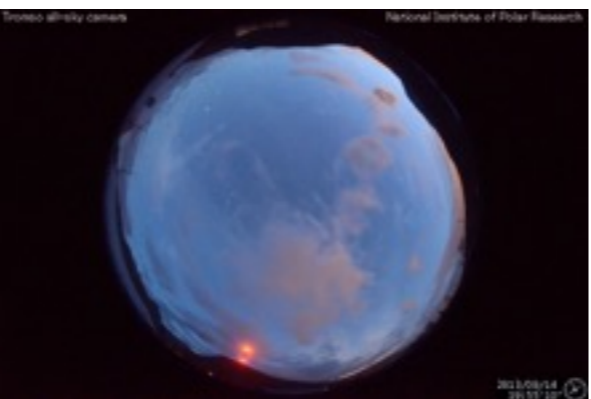
暗闇

夕方

悪天候

上端に暗いオーロラ有

無



まとめと今後の方針

まとめ

データリポジトリに向けたビジョン
ラベル付け作業の報告
簡単なルールで意外によい結果

今後の課題

他の特徴を用いてより高精度に
細かな色情報、動き(動画)、局所特徴量など
学習アルゴリズムによる自動的な分類学習