

# ソーシャルメディアに対する ジオタギングを利用した 降水状況の解析

北本 朝展<sup>1,2</sup>

<sup>1</sup> 国立情報学研究所, <sup>2</sup> JST 「さきがけ」

<http://agora.ex.nii.ac.jp/futtekitter/>

<http://agora.ex.nii.ac.jp/GeoNLP/>

# 基本的なアイデア



キーワー  
ド検索



 北本 朝展 @KitamotoAsanobu 5月6日  
ふってきたー [agora.ex.nii.ac.jp/futtekitter/ra...](http://agora.ex.nii.ac.jp/futtekitter/ra...) が各地の雨と雷で  
すごい状況になっている。15時現在、埼玉から栃木、茨城にかけて、巨  
大な雷雲が帯状に発生しつつあるようだ。 [agora.ex.nii.ac.jp/digital-typhoo...](http://agora.ex.nii.ac.jp/digital-typhoo...)  
開く

ソーシャルストリーム

GeoNLPによる地名ベー  
スのジオタギング



(time, latitude, longitude, situation)



降水状況に関するソーシャル・  
科学ストリームの比較。



科学ストリーム



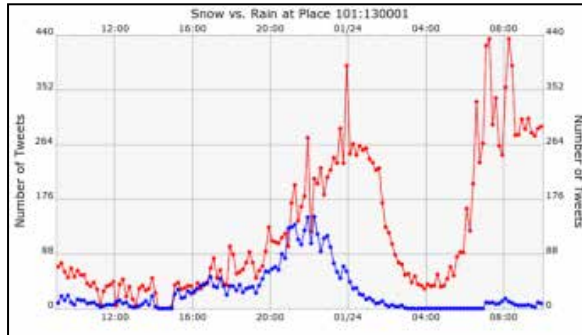
- 天気一般に比べて降水は実際に目で観察しやすいという利点がある。
- レーダーを使えば実際に降水があったかどうかを確認しやすいという利点もある。

# ソーシャル／科学ストリーム



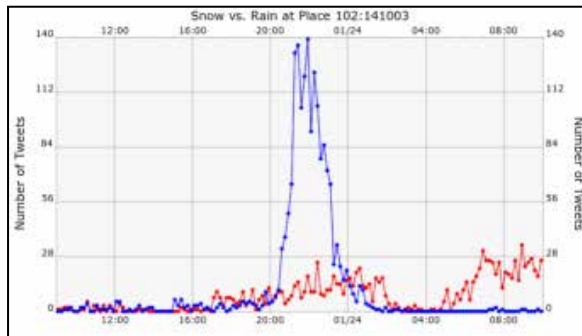
ツイート数（ソーシャルストリーム）を気象レーダー（科学ストリーム）に重ねたもの。左：リツイート除く、右：リツイート含む。

# 雨と雪の判別



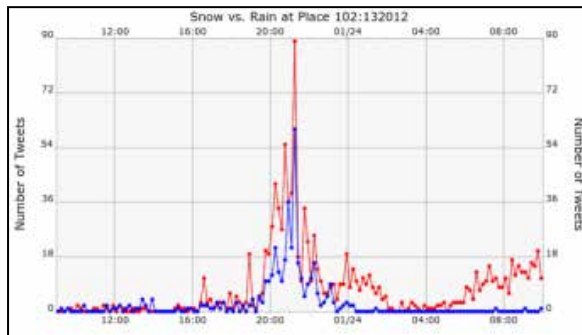
東京

- 東京は雪（赤）が多い。対して横浜は雨が多い。



横浜

- 雨雪判別は現在でも難しいため、新しい情報となる。



八王子

- 雪は「積もった状態」もある。

# ツイートに関する統計

種類	ツイート数	地名数	地名数/ツイート数
雨	3654564	4596602	1.26
雪	1738441	2207735	1.27
雷	478471	619218	1.29
台風	855120	1114163	1.30
ひょう	62369	70901	1.14

- 地名を含む100万以上のツイートが1年以内に収集できた。
- キーワード検索の制限で、気象現象文字列を含むツイートが結果に混ざる。
- 20-30%のツイートは複数の地名を含む。

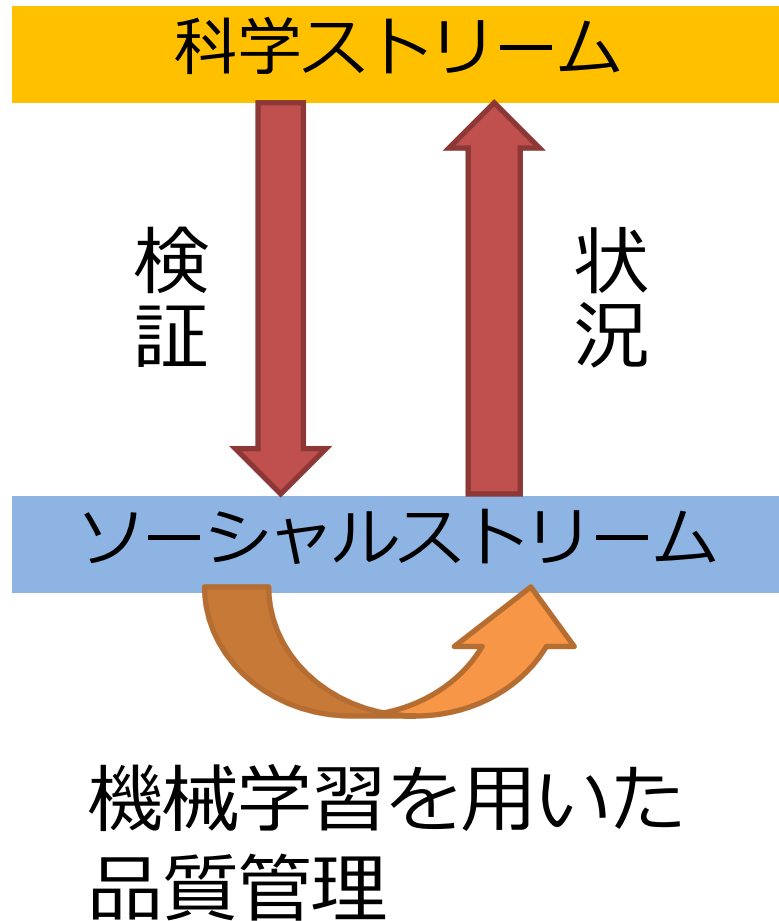


GPS情報が付与されたツイートは全体の0.15%しかない。



地名を一つ以上含むツイートは全体の10%程度はある。

# 今後の課題



- 科学データを検証に用いる。
- 機械学習をツイートのフィルタリングに用いる。
- ソーシャルメディアデータに対する品質管理。
- ソーシャルメディアを用いた状況把握へ。