

# 形態素解析を用いた大規模停電に関するツイート分析

- 北海道胆振東部地震におけるケーススタディ -



グループ演習 4班  
児島 怜, 段 煙煙, 新田 和樹, 楊 嵐

アドバイザー教員: 庄司 学

## 1. はじめに

- 1.1 研究背景
- 1.2 既往研究
- 1.3 研究目的

## 2. 停電情報の可視化

- 2.1 停電戸数の遷移状況
- 2.2 停電情報の時空間分布センシング

## 3. 形態素解析を用いたツイート分析

- 3.1 形態素解析による頻出単語の集計
- 3.2 共起ネットワークの可視化
- 3.3 停電戸数と共起関係強度の時系列変化

## 4. まとめ

# 1.1 研究背景



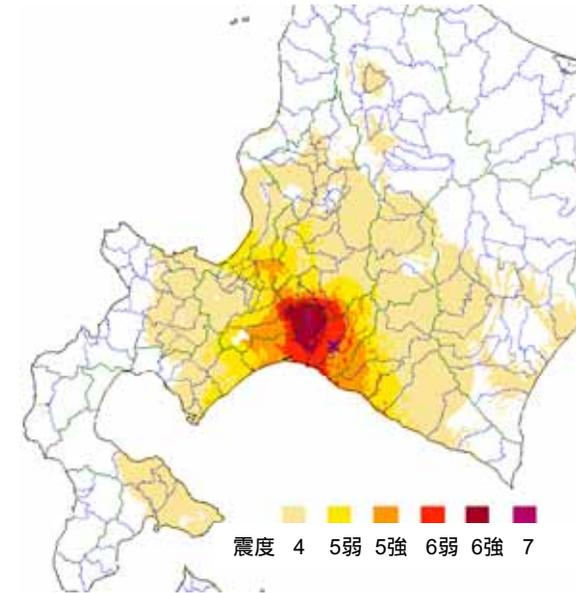
## 北海道胆振東部地震

発生日時: 2018年9月6日3時7分

最大震度: 7      マグニチュード: 6.7

死者 : 43人      負傷者 : 782人

北海道全域約295万戸の大規模停電



### ➤ 歴史的な大規模停電

兵庫県南部地震(1995年)      約300万世帯

東北地方太平洋沖地震(2011年)      約845万世帯

インフラ・産業に大きな影響

### ➤ 大規模停電による横断的な影響

特に北海道は,

国際的な観光都市

日本を代表する農業地域

観光産業

農林水産業

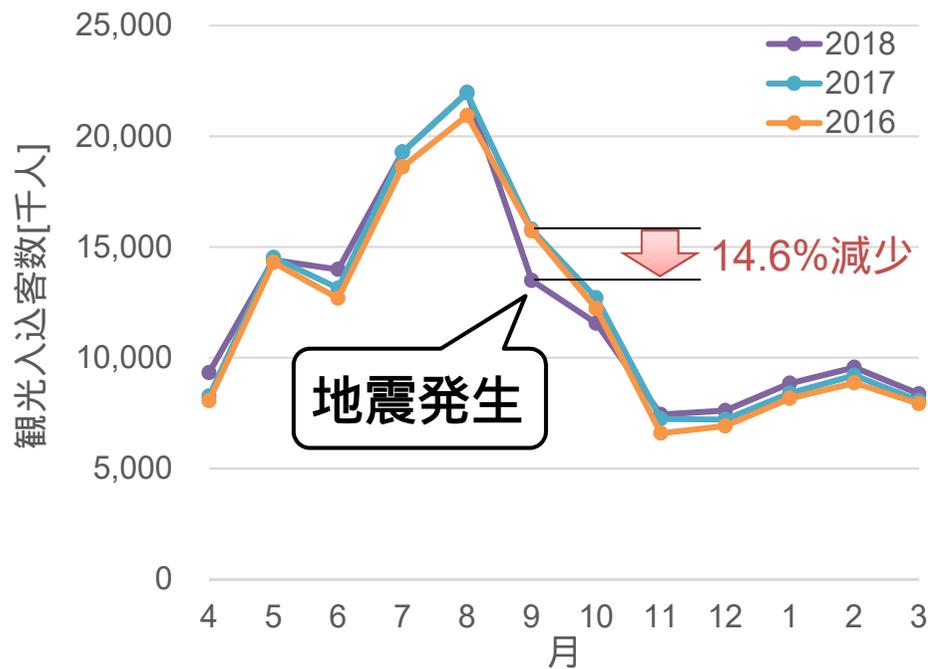
にも影響



# 1.1 研究背景

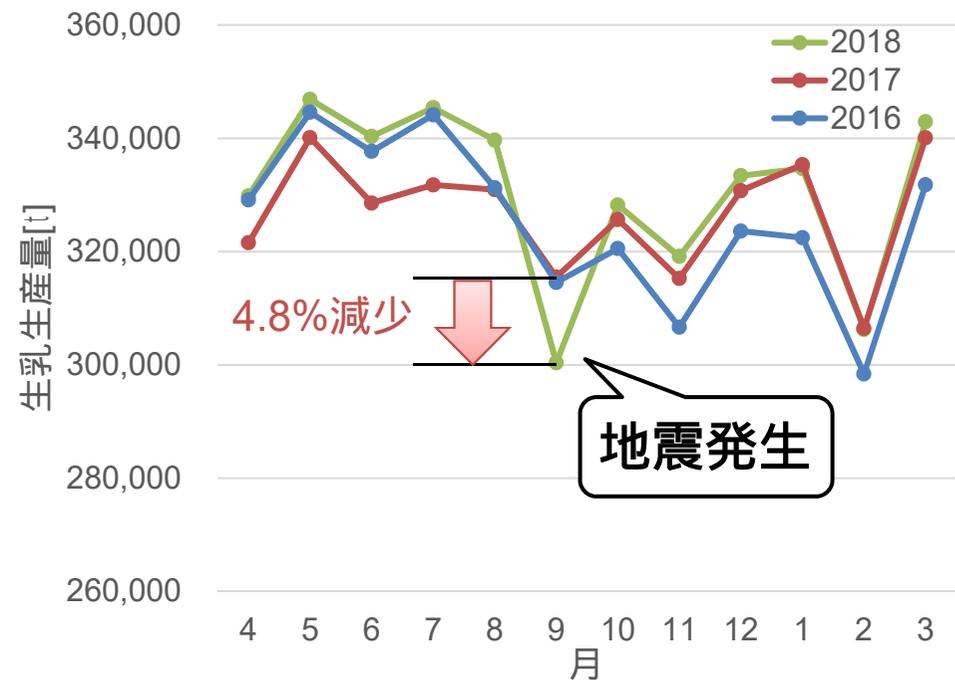


### 北海道における観光入込客数の推移



9月の観光入込総数  
前年比 14.6% 減少

### 北海道における生乳生産量の推移



9月の生乳生産量  
前年比 4.8% 減少

北海道胆振東部地震による影響が横断的に拡大していることがわかる

## 1.2 既往研究



西村ら(2013)

宿泊旅行統計データおよび宿泊業者を対象としたアンケート調査結果を用いて、東日本大震災後における**宿泊需要減少**に関する実証分析

震災による需要の減少や、復興需要による被害の軽減効果が時間および空間的に明らかとなった

庄司ら(2015)

東北地方太平洋沖地震において、**Twitter情報**を基に停電領域および復旧過程の時空間分布の特徴を分析

地震災害時における停電情報の時空間分布のセンシングにおいてTwitter情報が極めて有用であることが示唆された

湊ら(2014)

著者の心理情報を具体化するために**形態素解析**により、事故や企業トラブルに関する(1)概要の記述(2)感想文の2種類の文書を評価

形態素解析では、概要と感想文を分類する手段として使用できる可能性があることが示唆された

# 1.3 研究目的



## 本研究の目的

北海道胆振東部地震における大規模停電の影響を定量的に評価する

### 研究概要

#### ◆ 停電情報の可視化

北海道電力が公開している停電情報を基に、停電地域(空間的)およびその遷移状況(時間的)を可視化する

#### ◆ 形態素解析を用いたツイート分析

北海道胆振東部地震に関するツイートデータを抽出し、

(1) 頻出単語の集計

(2) 共起ネットワークの可視化

(3) 停電戸数と共起関係強度の時系列変化

} の3つの分析を行う

## 2. 停電情報の可視化



北海道電力が公開している停電情報を時間的および空間的に表現

### 1. 停電情報の収集

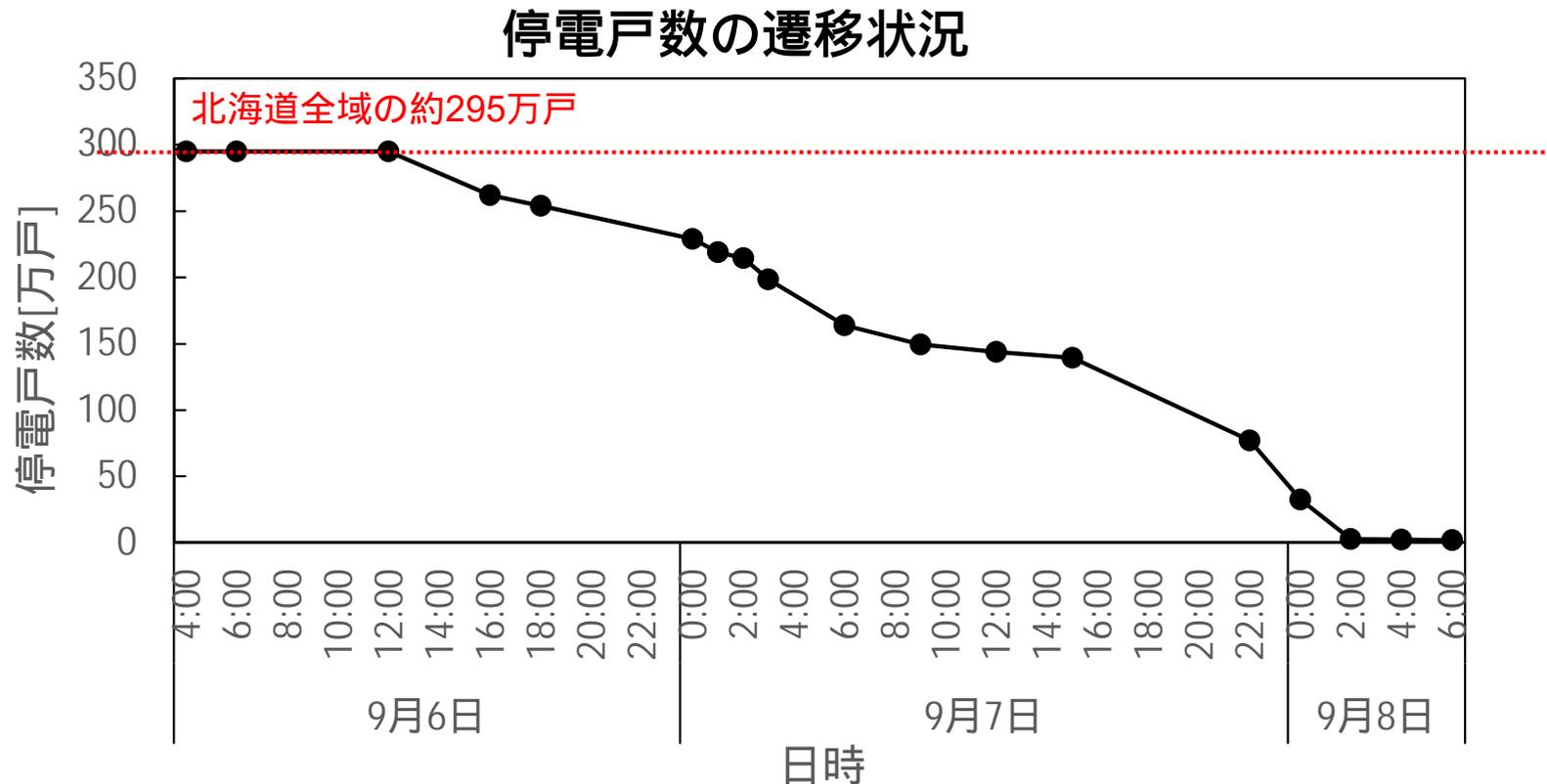
北海道電力が公開しているホームページ・Twitter・Facebookから、停電地域と停電戸数のデータを収集



### 2. 停電情報の可視化

停電戸数の遷移状況を示し、停電地域を時間的および空間的にGISで表現することで停電情報の可視化を行う

## 2.1 停電戸数の遷移状況



道内全域停電は1951年の北海道電力創設以来初の出来事である

・地震が3時7分に発生するが、当日の12時に電力を回復して始めた

・今回の停電は二日余りに渡った

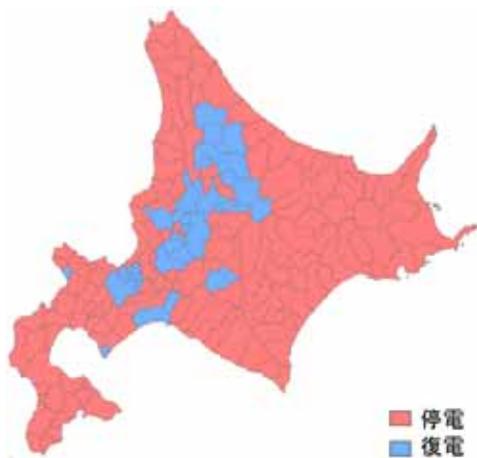
復電がはじめるのが遅かった

全域に電力を完全回復した時間が長い

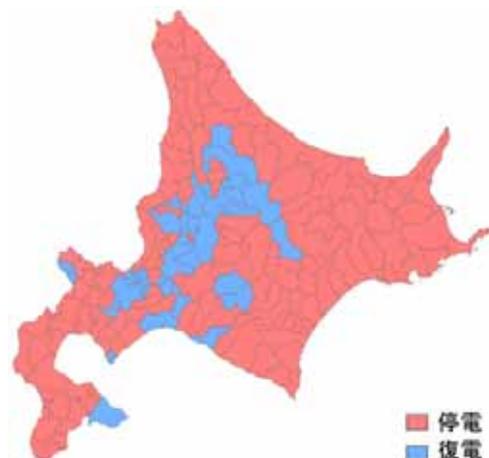
## 2.2 停電情報の時空間分布センシング



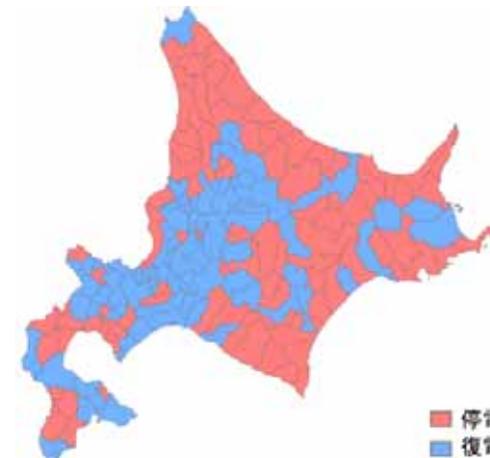
9/6 16時停電状況



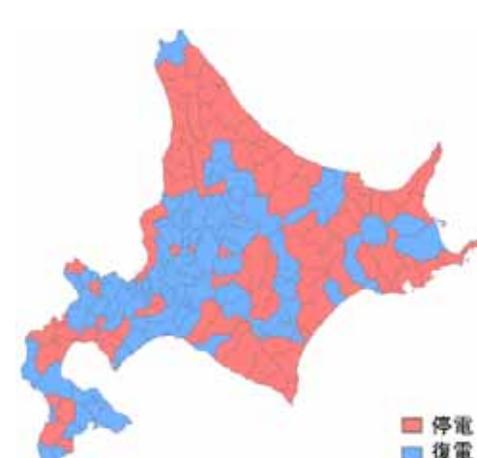
9/6 18時停電状況



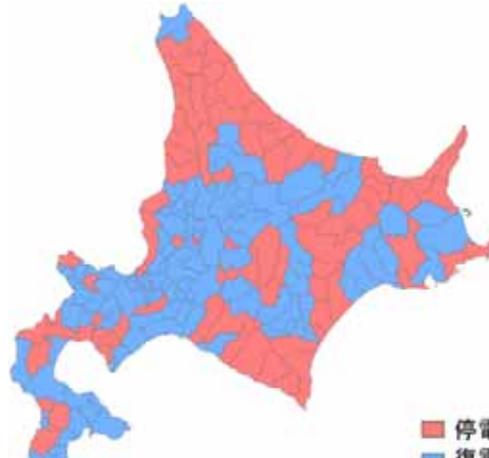
9/7 1時停電状況



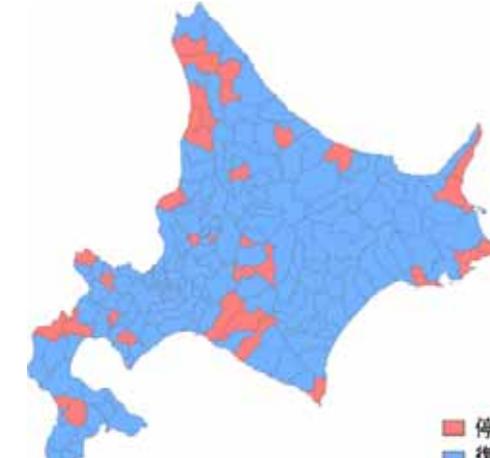
9/7 2時停電状況



9/7 3時停電状況



9/7 6時停電状況



参考: 9月6日16時[https://www.fepec.or.jp/about\\_us/pr/oshirase/\\_icsFiles/afieldfile/2018/09/06/h08.pdf](https://www.fepec.or.jp/about_us/pr/oshirase/_icsFiles/afieldfile/2018/09/06/h08.pdf)  
9月6日18時<https://www.facebook.com/534714416638119/posts/1807864872656394?s=100011333960968&v=&sfns=mo>  
9月7日 1時[https://www.fepec.or.jp/about\\_us/pr/oshirase/\\_icsFiles/afieldfile/2018/09/07/h04.pdf](https://www.fepec.or.jp/about_us/pr/oshirase/_icsFiles/afieldfile/2018/09/07/h04.pdf)  
GISデータ: [http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N03-v2\\_3.html](http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N03-v2_3.html)

9月7日 2時[https://www.fepec.or.jp/about\\_us/pr/oshirase/\\_icsFiles/afieldfile/2018/09/07/h05.pdf](https://www.fepec.or.jp/about_us/pr/oshirase/_icsFiles/afieldfile/2018/09/07/h05.pdf)  
9月7日 3時[https://www.fepec.or.jp/about\\_us/pr/oshirase/\\_icsFiles/afieldfile/2018/09/07/h06.pdf](https://www.fepec.or.jp/about_us/pr/oshirase/_icsFiles/afieldfile/2018/09/07/h06.pdf)  
9月7日 6時[https://www.fepec.or.jp/about\\_us/pr/oshirase/\\_icsFiles/afieldfile/2018/09/07/h07.pdf](https://www.fepec.or.jp/about_us/pr/oshirase/_icsFiles/afieldfile/2018/09/07/h07.pdf)

### 3. 形態素解析を用いたツイート分析



#### 1. 形態素解析による頻出単語の集計

平成30年北海道胆振東部地震に関するツイートデータを抽出し、頻出単語の集計を行った。

期間: 2018/09/06 ~ 2018/09/08の3日間の6, 12, 18, 24時のそれぞれ10分間  
Twitterの検索機能によって検索ワードを設定して収集



#### 2. 共起ネットワークの可視化

頻出単語の集計後、同一ツイート上にある単語同士の組み合わせをカウントし、共起関係を測った。

### 3. 形態素解析を用いたツイート分析



#### 研究地域: 札幌

##### 地域設定理由

- ・震央である北海道胆振地方中東部に比較的近い。
  - ・北海道の都市の中で最も人口が多い
- 地震による被害を様々な人々が受けた

##### 分析手順

- ・Twitterの検索機能によって検索ワードを設定して収集
- 検索ワード:「札幌」

例:「札幌 since:2018-09-06\_06:00:00\_JST until:2018-09-06\_06:10:00\_JST」

除外:引用ツイートやリツイート、自動生成ツイート

- ・形態素解析器MeCab

名詞に分類された単語を集計

# 3.1 形態素解析による頻出単語の集計



表 1 形態素解析による頻出単語の集計結果上位80

1. 「札幌」、「北海道」  
検索ワードと、それに関する地名
2. 「地震」、「停電」、「復旧」、「電気」、「断水」  
北海道胆振東部地震の被害
3. 「無事」、「大丈夫」、「情報」  
被災に遭った方々の安否確認  
北海道外の方々の、心配するツイート
4. 「実家」、「友達」  
3.に合わせて使用される
5. 「みたい」  
断定の回避
6. 「道路」、「信号」  
道路陥没や信号機の停止  
インフラへの影響

順位	単語	回数	順位	単語	回数	順位	単語	回数	順位	単語	回数
1	札幌	4526	21	大丈夫	267	41	時間	143	61	一部	106
2	北海道	838	22	心配	253	42	中止	142	62	電話	103
3	地震	776	23	中	253	43	何	139	63	函館	103
4	停電	711	24	被害	228	44	予定	139	64	東	103
5	復旧	626	25	連絡	207	45	余震	137	65	信号	99
6	市	538	26	店	185	46	ニュース	134	66	道路	99
7	電気	526	27	今	180	47	北	130	67	東京	99
8	日	480	28	月	178	48	地域	129	68	場所	96
9	区	422	29	今日	175	49	充電	126	69	再開	95
10	こと	385	30	震度	175	50	水	122	70	的	93
11	よう	371	31	私	167	51	旭川	122	71	病院	93
12	人	333	32	所	166	52	これ	120	72	影響	91
13	情報	306	33	ところ	166	53	確認	117	73	テレビ	91
14	方	302	34	断水	161	54	事	116	74	災害	90
15	市内	295	35	実家	155	55	うち	116	75	友達	89
16	無事	292	36	大変	153	56	駅	115	76	ドーム	89
17	みたい	290	37	明日	150	57	中央	112	77	代表	87
18	時	290	38	気	149	58	日本	109	78	前	87
19	そう	285	39	営業	149	59	一	108	79	笑	87
20	さん	272	40	状況	145	60	家	108	80	みんな	86

表の単語は数字と意味を持たないアルファベット、記号を除いたものに限っている

## 3.2 共起ネットワークの可視化



### 共起関係

1つの文中にある単語が現れたとき、その文中に限られた単語が頻繁に出現する関係  
本研究 同一ツイート上にある単語同士を共起関係であると定義

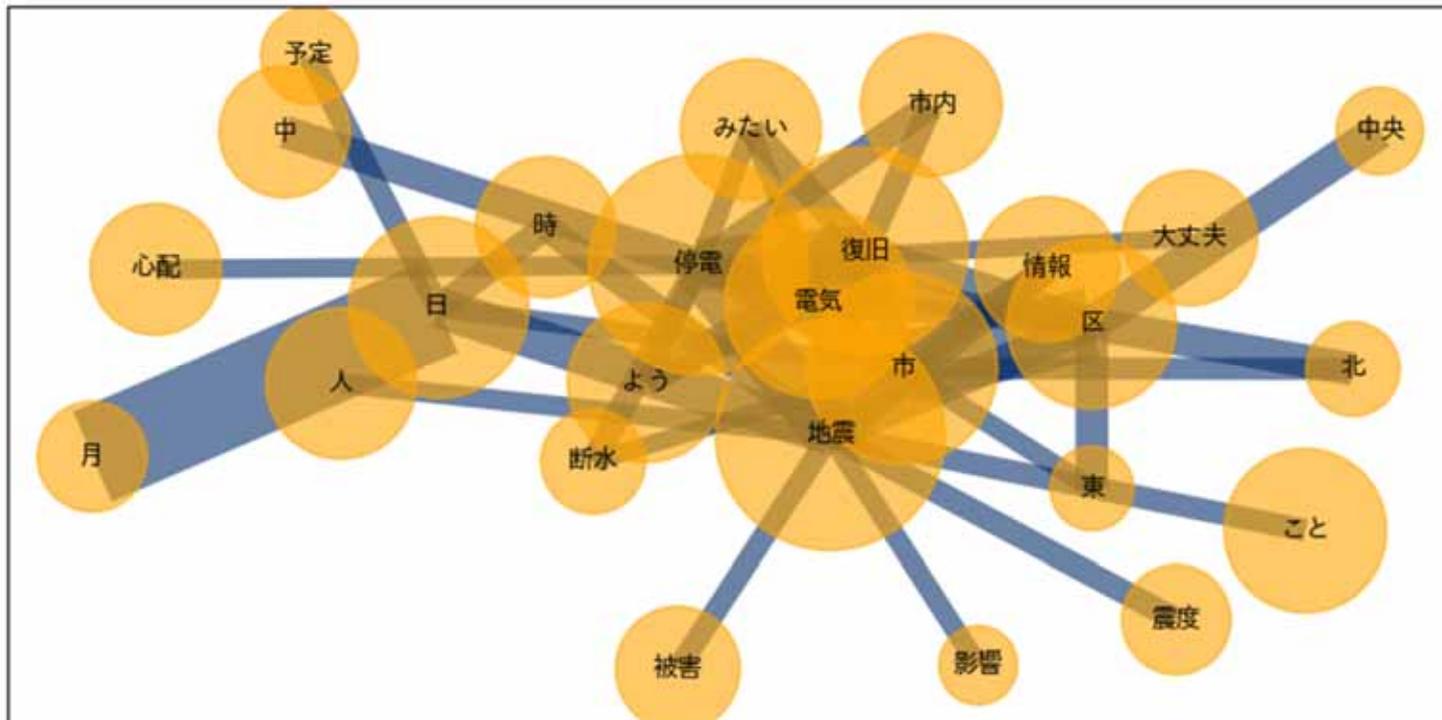
例:「今は**停電**と**水**は**断水**まではいってないけど濁ってて**不便**です。」

「**今**」「**停電**」「**水**」「**断水**」「**不便**」の5つの名詞が検出

(今, 停電)、(今, 水)、(今, 断水)、(今, 不便)、(停電, 水)、  
(停電, 断水)、(停電, 不便)、(水, 断水)、(水, 不便)、(断水, 不便)

の10種類の組み合わせが共起関係としてそれぞれカウントされる。

## 3.2 共起ネットワークの可視化



黄色い円の広さ: その単語の頻出度の強さ  
青い線の太さ: その単語同士が共起関係の強さ  
位置は意味を持たない

ツイートの共起関係が視覚的に明らかになった

## 3.2 共起ネットワークの可視化



停電 ←→ 地震、中、復旧、断水、市内、電気、みたい、影響、大丈夫、心配  
電気 ←→ 復旧、停電、区、よう、地震、市、みたい、こと、水道、水、市内  
地震 ←→ 停電、市、日、区、情報、時、震度、こと、被害、影響、人、電気、復旧、東

矢印の右側の単語は共起関係の強い順に並べた

### 考察

1. 「停電」、「電気」、「地震」はそれぞれ共起関係  
「復旧」ともそれぞれが強い共起関係

復旧にかかる時間への非常に強い関心

2. 「電気」 「水道」、「水」  
「停電」 「断水」

停電に伴う断水の被害、水への不安

(停電 復旧) ,(電気 復旧)  
の共起関係が非常に強い

### 3.3 停電戸数と共起関係強度の時系列変化

(停電 復旧), (電気 復旧) の共起関係が非常に強い

時系列による共起関係の増減を調査

**停電戸数の解消とこれらの2つの共起関係の関連性を調査**

#### 分析手順:

期間: 2018年9月6日6時 ~ 2018年9月8日6時まで

6, 12, 18, 24時で取得した10分間のツイートデータ

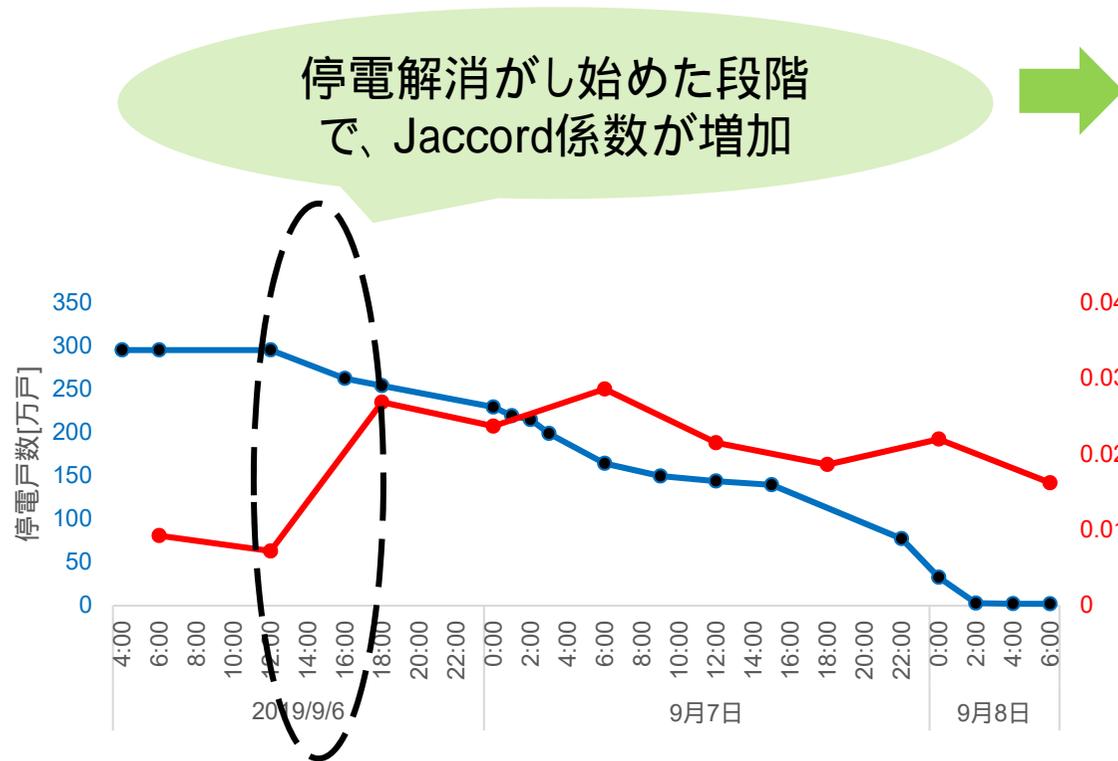
それぞれ共起関係を測り、それを集合の類似度を測るJaccard係数を用いて値を算出した。

$$\text{Jaccard係数} = \frac{A \cap B}{A \cup B}$$

A: 「停電」「電気」のいずれかを要素に持つ共起の発生回数

B: 「復旧」を要素に持つ共起の発生回数

# 3.3 停電戸数と共起関係強度の時系列変化



停電戸数と共起関係強度の時系列変化

青線は停電戸数[万户]の変化を示し、  
赤線はJaccord係数の量を示す

停電解消がし始めた段階  
で、Jaccord係数が増加



(停電,復旧)と(電気,復旧)の  
共起ツイート量が  
停電解消し始めた段階で**増加**



Twitterが停電復旧時、  
重要な情報発信・収集源

現実状況の変化によって  
共起関係の強度が変化

## 4 まとめ



1. 北海道電力が公開している停電情報に基づいてデータを収集し、GISを用いて停電地域の遷移状況の可視化を行った。停電戸数および停電地域の遷移状況を明らかとなった。

2. 地震に関するツイートデータを抽出し、頻出単語の集計を行った。集計結果では、「地震」「停電」「復旧」「電気」が上位を占めており、北海道胆振東部地震による停電の影響の大きさが伺えた。

3. 同一ツイート上にある単語同士の組み合わせをカウントする共起ネットワークの可視化を行った。「地震」との共起関係が最も強い単語は「停電」であることが示され、地震による停電の影響が甚大であることが明らかになった。

1. 2018年北海道胆振東部地震・大阪府北部の地震被害調査報告書；  
土木学会，2019
2. 西村泰紀，梶谷義雄，多々納裕一；大規模災害による宿泊業への影響評価  
—市町村宿泊旅行統計とアンケート調査に基づく東日本大震災のケーススタディ；  
土木学会論文集D3(土木計画学)，Vol.69，No.5，2013，pp.I\_217-I\_227
3. 庄司学，高橋大；Twitterを利用した地震災害時における停電情報の時空間  
分布のセンシング；平成27年度電気学会電子・情報・システム部門大会
4. 湊淳，鈴木慎吾，村上雄太郎，伊多波正徳，小澤哲；形態素解析を用いた  
事故・リスクに関する記述の分析；日本感性工学論文誌，Vol.13，No.2，  
2014，pp.341-346
5. 国土交通省；国土数値情報；<http://nlftp.mlit.go.jp/ksj/index.html>
6. MeCab: Yet Another Part-of-Speech and Morphological Analyzer；  
<https://taku910.github.io/mecab/>
7. Z. S. Harris；Co-Occurrence and Transformation in Linguistic Structure；  
Vol.33，No.3，Part 1，pp.283-340
8. 吉川武時，長尾真編；言語の機械処理；三省堂，1984