

## 時系列データ解析をより手軽に

### 問題

時系列データ解析はハードルが高く、データの活用が進んでいない。

### アプローチ

時系列データの解析を自動的に行う技術を開発する。

時系列データの活用を促進させることで、科学・産業を発展させる。

### 成果

イベント時系列データの解析技術を開発し、

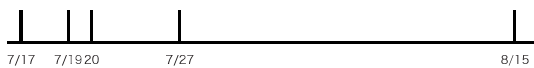
Web データ分析における有用性を示した。

#### ・ イベント時系列データ

あるイベントが起きた時刻のデータ

イベントはなんでもいい： Youtube のビデオ視聴 (情報学)  
商品購入 (マーケティング)  
地震発生 (地震学)

例：2016年7,8月の東京における地震データ (震度4以上)

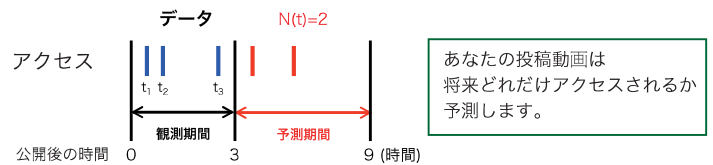


イベント時系列データは分野を超えて現れる。  
情報学、マーケティング、地震学、経済学、脳科学...

#### ・ Web データ： 常に増加しており、人手で解析するのは困難。



#### 応用例： Web コンテンツのアクセス数予測

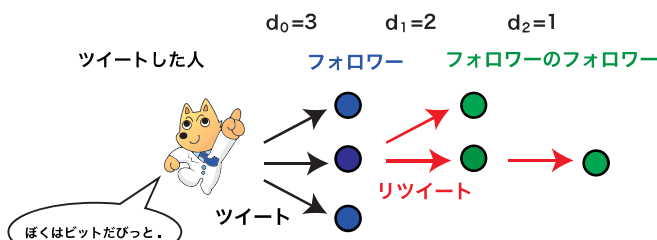


## 研究成果の詳細

### ・ Twitter

(Kobayashi & Lambiotte, ICWSM 2016)

ツイートが拡散する様子

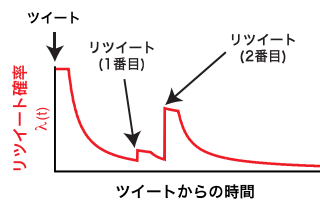


今後 (1, 2, 4日) にどれだけリツイートされるか予測した。

#### モデル： TiDeH

$$\lambda(t) = p(t) \sum_{i:t_i < t} d_i \phi(t - t_i)$$

リツイートする確率      ツイートを覚えている人の数

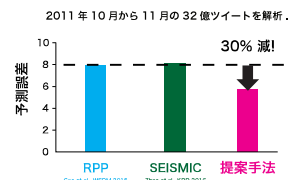


#### 予測公式

$$\hat{\lambda}(t) = h_T(t) + \bar{d}p(t) \int_T^t \hat{\lambda}(s) \phi(t-s) ds$$

観測期間の効果

#### 予測精度



### ・ 署名サイト

(Proskurnia, Grabowicz, Kobayashi, Castilo, Cudre-Mauroux, Aberer, WWW 2017)

署名サイトの署名数を予測する

署名サイトとは？

インターネット上で署名を集めて  
行政機関や大企業に抗議の声を届けるサイト

本研究で収集したデータ  
The Petition Site  
(動物愛護に関する署名)

日本における代表例  
Change.org

今後 (1, 2, 4日) にどれだけ署名されるか予測した。

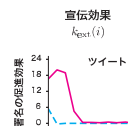
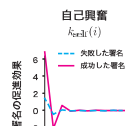
#### モデル

$$n(t) = h(t) + \sum_i k_{self}(i) n(t-i) + \sum_i k_{ext}(i) f_{ext}(t-i)$$

将来の署名数  
(時刻 t から t+1)

自己興奮効果  
(参加ユーザによる宣伝)

宣伝効果  
(ツイート, トップページ閲覧)



#### 予測精度

収集したデータ

署名：2016年8~10月の5,000署名。  
<http://thepetitionsite.com/>

ツイート：2016年8~10月の関連ツイート。

