

# 金融ビッグデータによるバブルの早期警戒技術の創出 - Detecting asset price bubbles -

水野貴之 国立情報学研究所・JSTさきがけ

## 概要

【目的・手法】バブルのナウキャストは可能か？  
・バブルとは単なる急激な成長ではなく、実体経済から乖離した過剰投機による上昇である



集中投機による不平等の拡大を観測

### 【株式・不動産バブル検出】

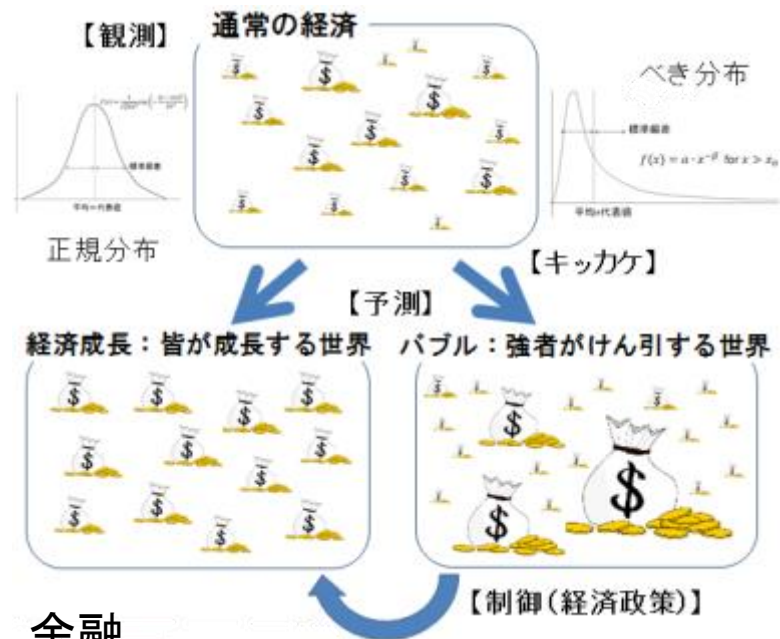
- ・人気の銘柄に投機マネーが集中するために企業業績の差以上に株価の差が生まれる
- ・人気の物件に投機マネーが集中するために物件の質の違い以上に価格に差が生まれる

### 【暴落・暴騰のキッカケの検出】

- ・ネガティブな情報ほど短期間に報じられる
- ・ニュースの「新規性」「話題性」を抽出

### 【金融ショックの波及】

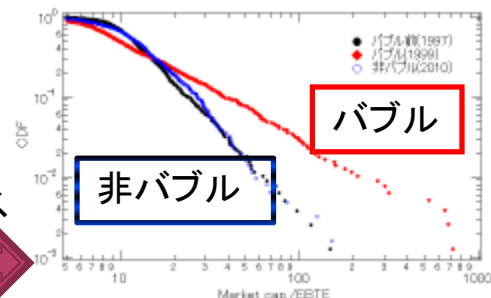
- ・4取引先まで金融ショックが波及



## 金融ビッグデータ

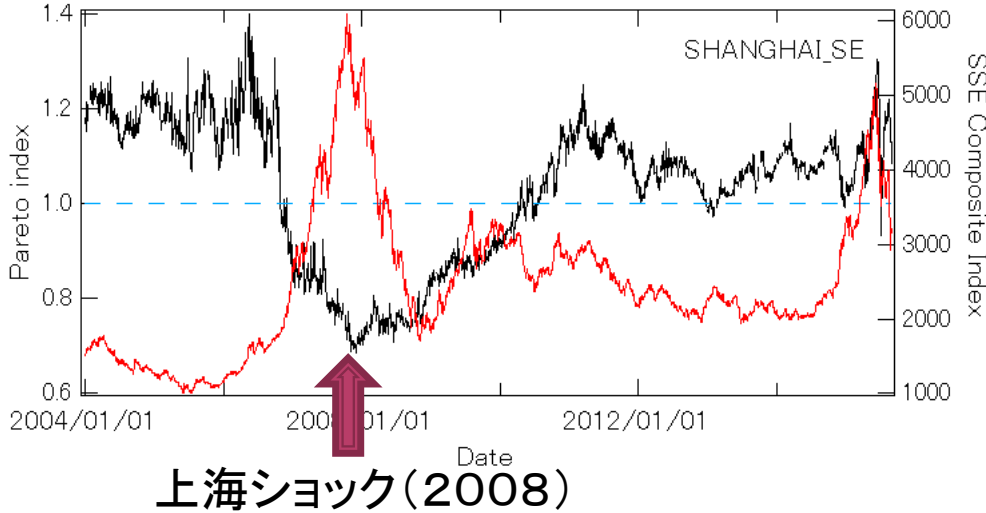
- 30万社の株価
- 1億社の財務諸表
- 100万件の不動産
- 3億記事のニュース

集中投機の可視化



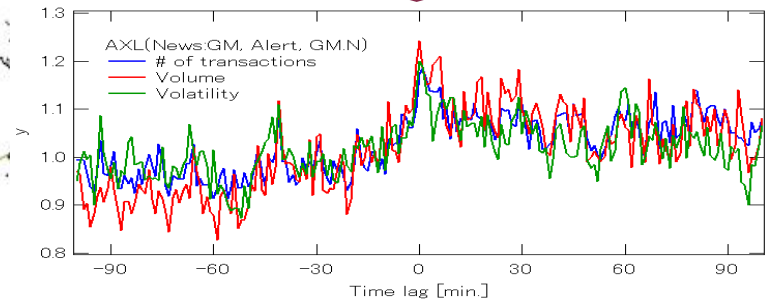
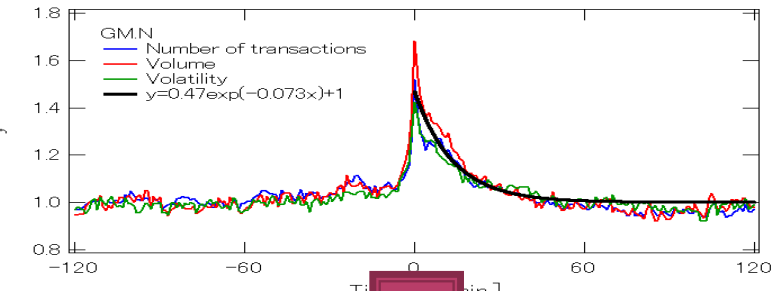
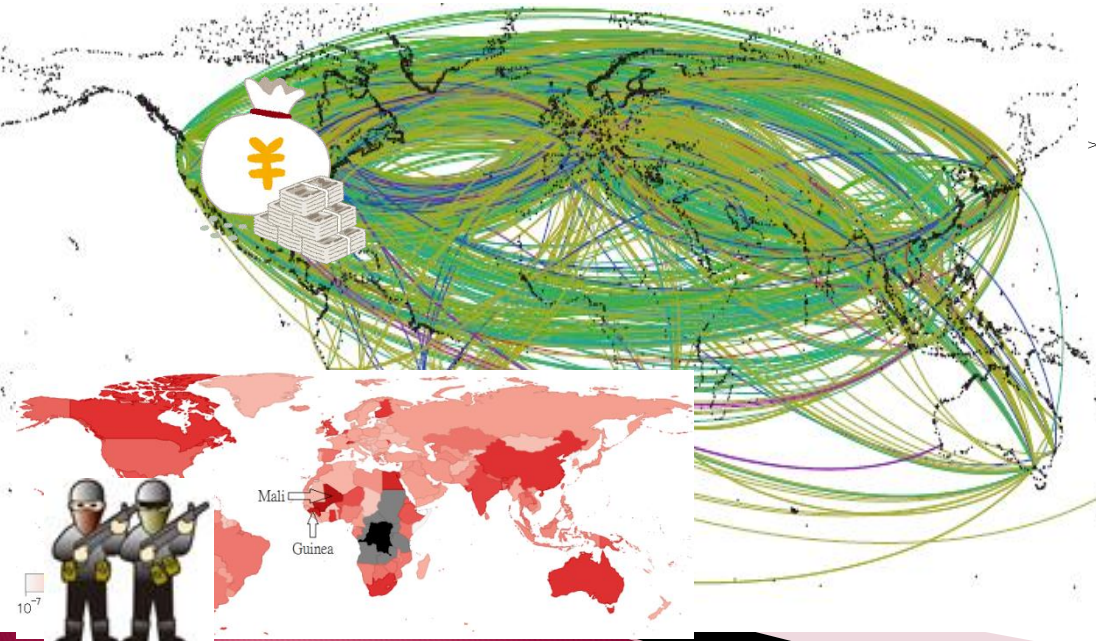
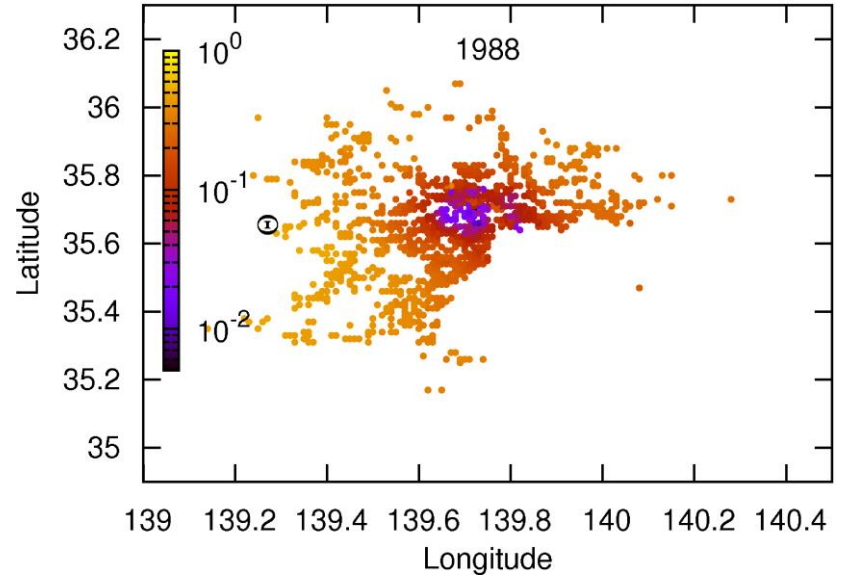
# バブルと拡散

## 上海株式市場での株価格差



## 日本の不動産バブル

[Ohnishi, Mizuno, et al, IJMP2012]



JSTさきがけ

「社会的課題の解決に向けた数学と諸分野の協働」

## 乱択アルゴリズム設計の技法と脱乱択化の数理

来嶋秀治

九州大学 大学院システム情報科学研究所

JST さきがけ研究者

ELC B01班分担

Fukuoka  
(福岡)



Exploring the  
Limits of  
Computation

# 高次元多面体の体積計算の決定性近似について

## Lovász [1986]

“membership oracleで与えられた  $n$ 次元凸体の体積に対して、  
近似比  $1.999^n$  の決定性多項式時間アルゴリズムは存在しない。”

## 乱択近似 (1980～)

- ✓ Jerrum, Valiant, Vazirani (1986): 近似数え上げ $\approx$ ランダムサンプリング
- ✓ Toda (1991):  $PH \subseteq P^{\#P}$  (Gödel賞)
- ✓ Dyer, Frieze, Kanna (1991): 高次元凸体体積の乱択近似可能 (Fulkerson賞)
- ✓ Jerrum, Sinclair, Vigoda (2004): パーマネントの乱択近似可能 (Fulkerson賞)

## 決定性近似 (発表者の研究 2014～)

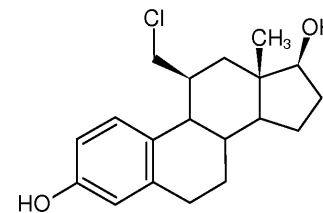
- ✓ ナップサック多面体の体積は決定性近似可能(FPTAS) [Ando, Kijima15+]
  - H多面体: 線形不等式系で記述される多面体
- ✓ ‘双対’ナップサック多面体の体積は決定性近似可能(FPTAS) [Ando, Kijima16+]
  - V多面体: 端点集合で記述される多面体
  - $n$ 次元 $L_1$ 球2個の交差領域(resp. union)の体積計算は#P困難

※ポスターでは「おねえさん問題の乱択近似」のはなしをします。

# Indexing billions of chemical libraries for scalable similarity searches

Yasuo Tabei (TITECH/JST-PRESTO)

- Chemical library is artificially generated compounds from functional ones
- The size of the whole chemical space is said to be approximately  $10^{60}$
- Huge chemical libraries are created in the near future (10 billions of chemical libraries)
- Currently, two billions of chemical libraries are generated



# Similarity search of chemical descriptors (integer vectors)

- Similarity search for 40 million compounds in PubChem
  - Find all descriptors  $x_i$  similar to query  $q$
  - Similarity = Generalized Jaccard: 
$$J(x_i, q) = \frac{x_i \cdot q}{\|x_i\|_2^2 + \|q\|_2^2 - x_i \cdot q}$$
- Many methods for fingerprints (0/1 vectors)
  - Multibit tree[WABI,10], DivideSkip[ICDE,08],  
b-bit minhash[WWW,10], SMBT[KDD'13] etc
- Recent progress on descriptor representation having rich information on compounds
  - KCF-S[GIW,13], SMILES-D[JCIM,11], many others [Wiley,09]
- Present novel compact data structures for fast similarity search of descriptors

# Similarity search on 40 million compounds, take less than one second and 4GB memory

