

# 企業に対するニュースの 影響予測のための重要語抽出

東京大学大学院工学系研究科

技術経営戦略学専攻

松野紘明

# 背景

- Webを通じたニュースの配信が定着した
- 日々の新聞記事や速報だけでなく、過去の記事もWeb上で閲覧可能である
- しかし、数が多く、個人では情報を生かすきれない



## 目的:

- ✓大量の記事の閲覧補助システムの作成
- ✓まずは記事に関係する企業名を抽出したい
- ✓関連企業を上げることで、影響を受ける企業  
が分かる

# 従来手法による重要語抽出

- 使える文書
  - ニュース記事、決算書、企業のホームページなど
- TFIDFによる重みづけ
$$tfidf(t, d) = tf(t, d) \cdot idf(t)$$
- 内積による類似度の評価

しかし、

- 同じ単語が出現しない場合には類似度が下がる
- 形態素解析の区切り位置が適切とは限らない

# 語ベクトルの内積値による 関連企業の抽出

- ロイターのニュースと決算書を使用
- mecabを用いてTFIDFを計算し語ベクトルを作成
- 上記の記事と全上場企業との内積を計算した

結果は、右の表のとおり。  
1つの単語のみで類似度が決まっていることが分かる。

title 米GM、2010年までにインドの市場シェア10%獲得へ = インド部門社長  
time 2008年01月9日 16:18 JST  
[ニューデリー 9日 ロイター] 米ゼネラル・モーターズ(GM)(GM.N: 株価, 企業情報, レポート)インド部門のカール・スリム社長兼マネジングディレクターは9日、GMは2010年までにインドの市場シェアを10%に拡大するとの目標を達成できると確信していると述べた。  
同社長は、当地で行われたオートショーで「キャプティバ」を紹介。2008年はインドのディーラー網を「大幅に」拡充するほか、新モデルを導入してシェア拡大につなげたいと述べた。  
同社は、年間生産能力を22万5000台超に増やすため、第2工場の建設に着手している。07年のインドでの同社販売台数は6万台超で、前年比68%増となった。  
同社は、今後エンジン・トランスミッション工場を建設する可能性があり、インドが小型車の輸出拠点になることもあり得ると表明している。  
<http://jp.reuters.com/>より引用

社名	内積値	共起語
関西ペイント	0.0101958	インド
東京製鉄	0.0036625	工場
熊田組	0.0030044	建設
高松建設	0.0021667	建設
福田組	0.0021646	建設
鹿島建設	0.0020622	建設
福田組	0.0019284	建設
熊田組	0.0018363	建設
鹿島建設	0.0018283	建設
鹿島建設	0.0018067	建設
青木あすなろ建設	0.0017855	建設
鹿島建設	0.0017446	建設
高松建設	0.0017342	建設
青木あすなろ建設	0.0016465	建設
ナカノフドー建設	0.0013395	建設
松井建設	0.0013015	建設
あおみ建設	0.0012893	建設
松井建設	0.0012563	建設
あおみ建設	0.0012277	建設
東亜建設工業	0.0011708	建設
松井建設	0.0011119	建設

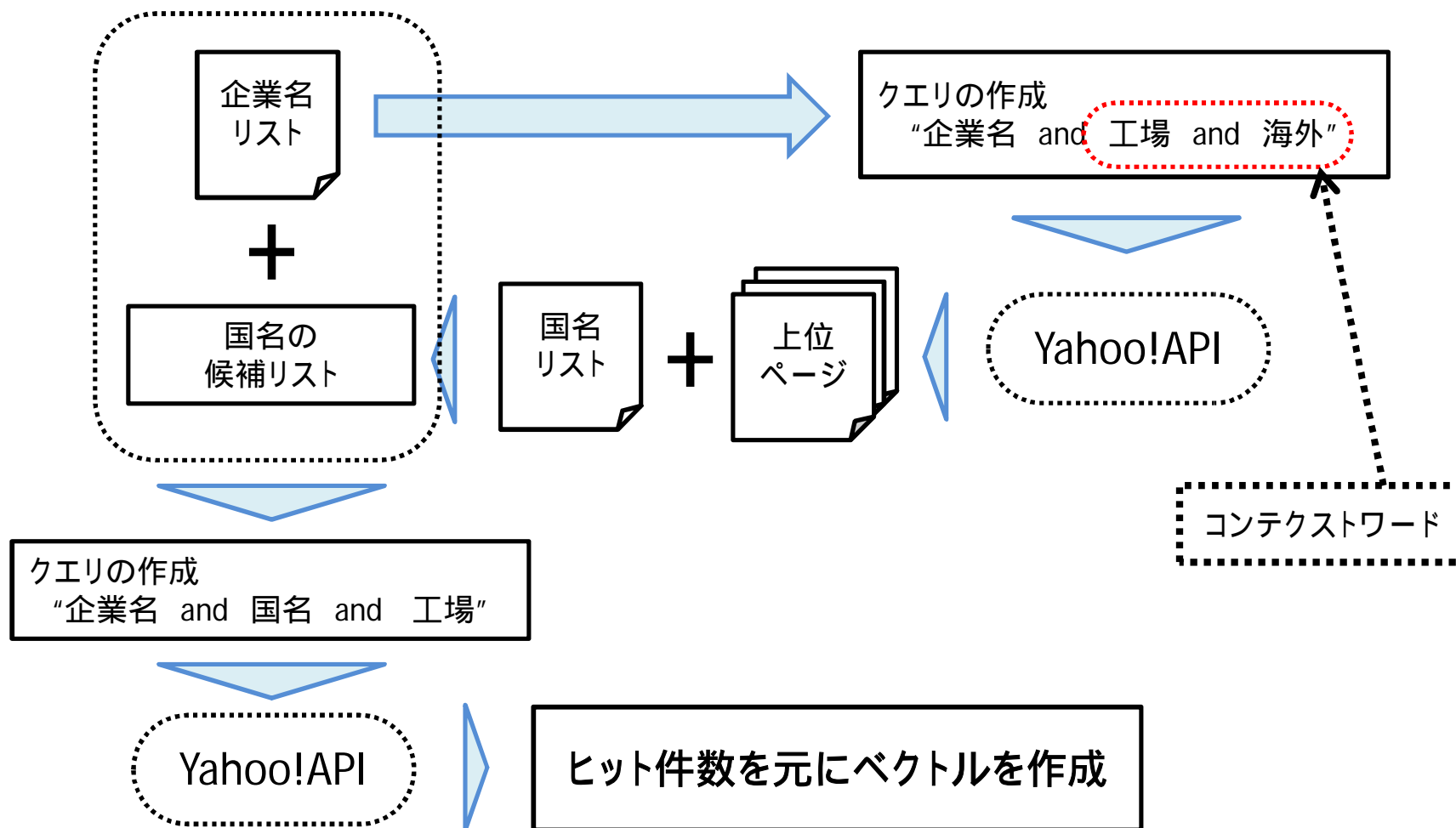
# 研究の目標

- 検索エンジンを利用する
  - 十分な公開情報にアクセスできる
  - 表記ゆれに強くなる
- 企業に関するベクトルを作成
  - まずは国名に着目
  - 企業ごとに“工場”に関するベクトルを作成

# システムの概要

- 国名リストを作成
- 企業名リストを作成
- 検索エンジンを使って企業ベクトルを作成
  - コンテキストワードは“工場”と“海外”
  - 「“企業名” and “工場” and “海外”」をクエリとして、上位のページから企業に関連する国名を取得
  - 「“企業名” and “国名” and “工場”」をクエリとし、再検索
  - 検索エンジンでのヒット件数を重みとしたベクトルを作成
- ニュース内に辞書内の国名がある場合に、企業ベクトル内でその国の重みが重い企業をランキング

# システムの概要



# 企業ベクトルの作成

- 上場企業の中でも電機業界に属するもののリストを作成
- 国名の辞書を作成
  - Wikipediaから引用
- yahooAPIを使ってヒット件数を取得
  - クエリは「"企業名" and "国名" and "コンテキストワード"」
- 企業ごとにヒット件数を重みとしたベクトルを作成
  - 国別工場ベクトル コンテキストワードは "工場"
  - 国別売上ベクトル コンテキストワードは "売上"



# 語ベクトルの結果

企業名	国		ヒット件数	実在するか
ソニー	台湾	工場	895000	
鈴木	台湾	工場	839000	×
東芝	台湾	工場	779000	
シャープ	台湾	工場	673000	
富士通	台湾	工場	564000	
キヤノン	台湾	工場	467000	
三洋電機	台湾	工場	344000	
三菱電機	台湾	工場	328000	
パイオニア	台湾	工場	320000	×
日立製作所	台湾	工場	260000	
松下電器産業	台湾	工場	241000	
京セラ	台湾	工場	204000	×
オムロン	台湾	工場	188000	
サンクス	台湾	工場	184000	×
リコー	台湾	工場	154000	
アクセル	台湾	工場	146000	×
松下電工	台湾	工場	146000	
TDK	台湾	工場	119000	
エルピーダメモリ	台湾	工場	108000	
ローム	台湾	工場	106000	×
セイコーエプソン	台湾	工場	104000	×

- 企業の国ベクトルから、台湾に工場を持つであろう企業のリストを作成
  - 左は電気機器業界のみである
  - 有価証券やWebから実際に進出しているかどうかを調べた
- 適合率は66.7%(14/21)

# 今後の課題

- 国名以外への展開
  - 技術名、商品名、地域名(県名など)
- 重みの最適化
  - ヒット件数のみでいいか
  - 他の指標が使えないか