

ソーシャルメディアからの社会予測 — ネットに映る実社会とは？ —

国立情報学研究所 客員准教授
東京大学 工学系研究科 准教授
松尾 豊

自己紹介



1997年 東京大学工学部電子情報工学科卒業

2002年 同大学院博士課程修了. 博士(工学)

産業技術総合研究所 研究員

2005年 スタンフォード大学客員研究員

2007年 東京大学大学院工学系研究科総合研究機構／技術経営
戦略学専攻／知の構造化センター 准教授

2010年 国立情報学研究所 客員准教授

2007年より国際WWW会議プログラム委員

2012年より、人工知能学会 編集委員長・理事

専門は、Webマイニング・社会ネットワーク・人工知能

人工知能学会論文賞(2002年)

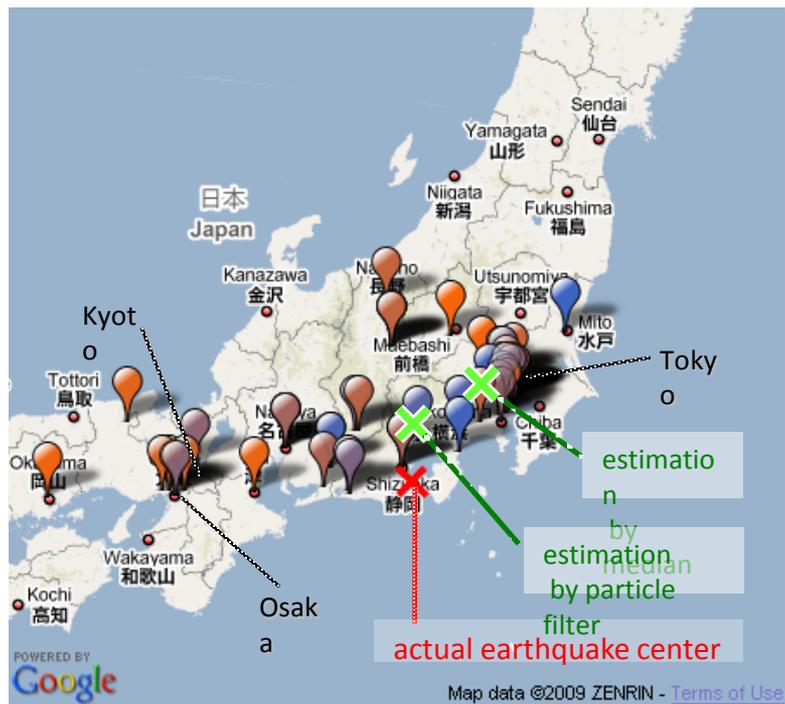
情報処理学会長尾真記念特別賞受賞(2007年)

人工知能学会 創立20周年記念事業賞、現場イノベーション賞
ドコモモバイルサイエンス賞(2013年)

オーマ株式会社技術顧問、経営共創基盤(株)顧問、Plug AI Asset
Management 技術顧問、国家戦略会議 叡智のフロンティア部会委員等

ウェブ上のビッグデータから、社会を観測する

- ブログから選挙結果の予測: 2008年～。世界でも最初期
- Twitterから地震の検知: 2009年～。WWW論文、400以上の被引用、渋滞情報の検知(トヨタ自動車との共同研究)に展開



	予測数	正解数	誤り数	混戦	的中率
ブログ分析	300	256	43	0	85.67%
ブログ分析 (混戦あり)	254	241	39	36	94.88%
朝日新聞	260	245	15	40	94.23%
日経新聞	288	264	24	12	91.67%



(クチコミ総選挙, 2009, <http://senkyo.kakaricho.jp>より引用)

実用化の試み

- SPYSEE (国内最大の人物検索)
- READYFOR (国内初、最大規模のクラウドファンディング)

SPYSEE 松尾豊 検索

ATSUSHI 横倉勝仁 松木謙公 菅直人 原口一博

松尾豊さんの関連図 人物詳細情報へ戻る

その他 20%
学業・研究 80%

松尾豊さんに関する人物検索結果のネットワーク図。中心には松尾豊さんのプロフィールがあり、周囲には長岡謙、友部博之、大平法雄、松尾直弘、大向一純、新司啓男、田塚行則、横田浩一、中島、山口広樹、沼見介、松村真生、三浦麻子、西田豊明、西村祐一、フェルディナンド・マ、安田善、向崎直毅、ミシェル・フォーコー、田原雅夫などの人物が接続されています。

松尾豊
研究者プロフィール 松尾 豊 東京大学大学院工学系研究科 准教授。東京大学工学部卒業、東京大学大学院工学系研究科 電子情報学専攻博士課程修了。博士（工

READYFOR?

プロジェクトを始める | プロジェクトを探す | READYFOR?をもっと知る

プロジェクトをウォッチする

ホーム | 新着情報 1 | スポンサー数 16

現在の達成金額 **62,000** 円
目標金額 **115,000** 円
スポンサー募集終了まで **64** 日
支援した人数 **15** 人

このプロジェクトは 8月31日 (水) 午前11:13 の時点で、115,000円以上集まった場合のみ、決済が完了されます。

すでに購入予約がされています。確認はこちら

¥1,000 円の支援で受取る引換券
在庫制限無し
製作するベンガル語雑誌を1部お贈りします。

7人が支援

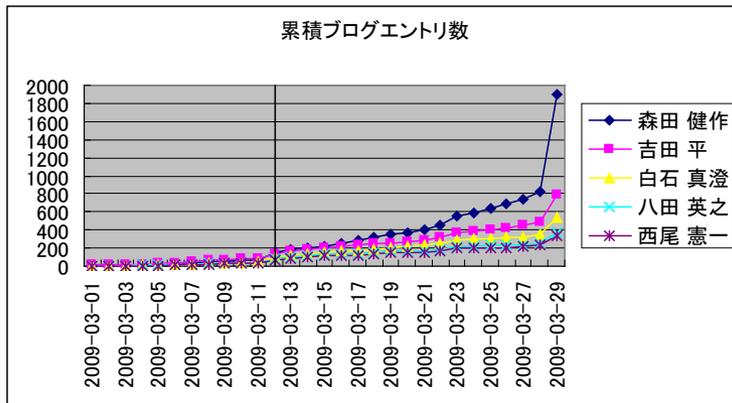
¥5,000 円の支援で受取る引換券
在庫制限無し
製作するベンガル語雑誌の裏表紙にスポンサー様のお名前を記載させて、1部お贈りし

プロジェクトについて
アジア最貧国バングラデシュ。たくさんのNGOが提供する奨学金やイヤーラーニングによって学習機会が十分でないながらも広まっていく一方、この国の15歳は将来を思い描かず、勉強するモチベーション、ひいては「人生を切り拓く」原動力を掛けずにいます。

ウェブにおけるビッグデータ

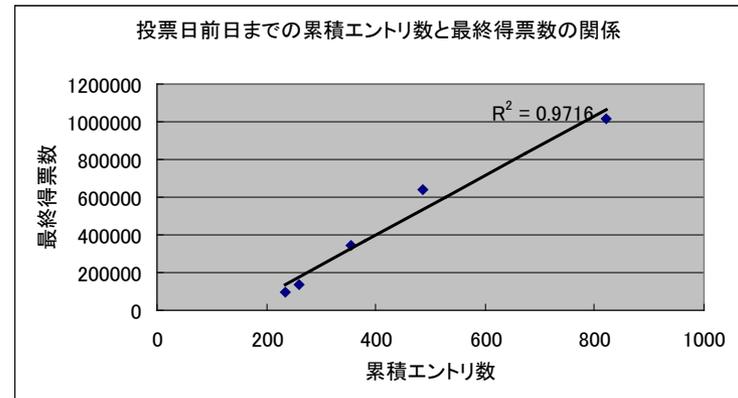
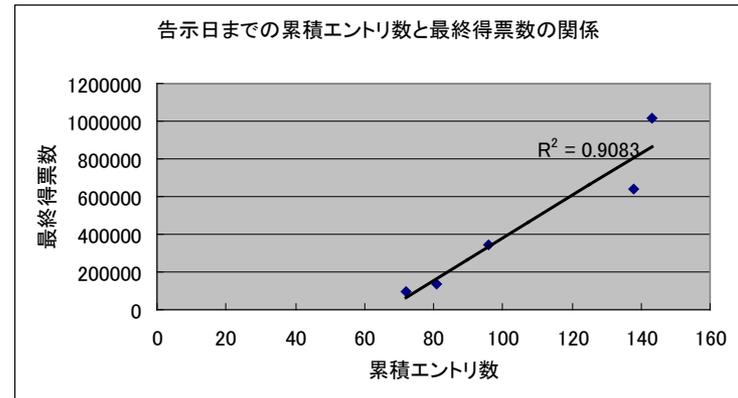
- 知能をいかに実現するか → ウェブ上の大量のデータをつかって、世界に関する知識をどのように取得するか。
- 要素技術
 - 検索
 - 自然言語処理
 - 機械学習
- 事例
 1. 選挙結果予測、地震の検知
 2. アイドルグループの分析
 3. クールジャパン
 4. 人と人のネットワーク抽出(人物検索)

ブログから選挙結果を予測する： 千葉県知事選挙

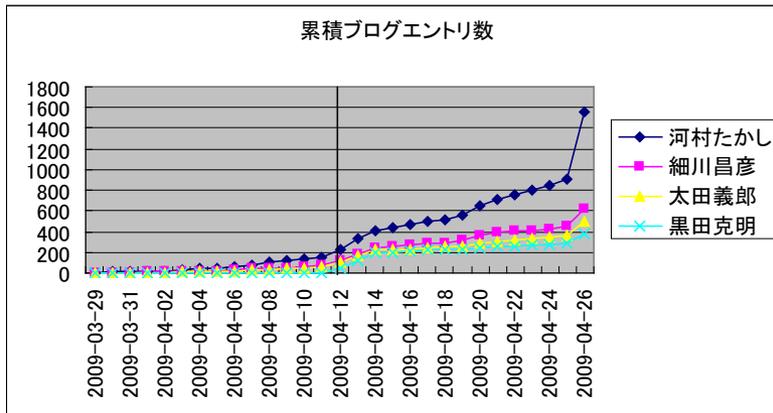


告示日: 2009年3月12日
投票日: 2009年3月29日

投票日前日までのブログエントリ数と最終得票数の間の相関係数: 0.98
時間とともに相関係数は増加し、告示時点でほぼ1に近い値まで推移

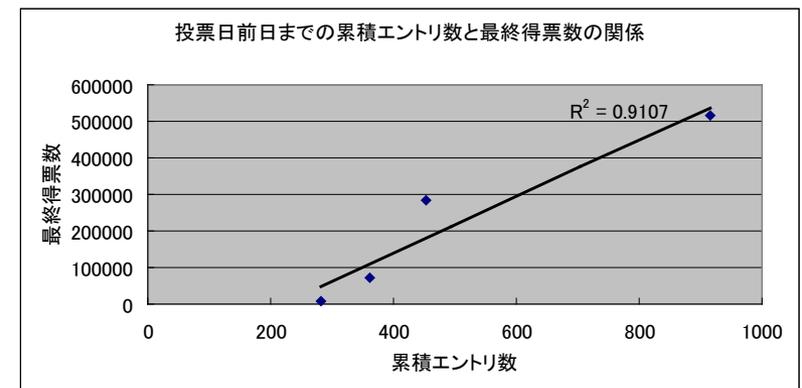
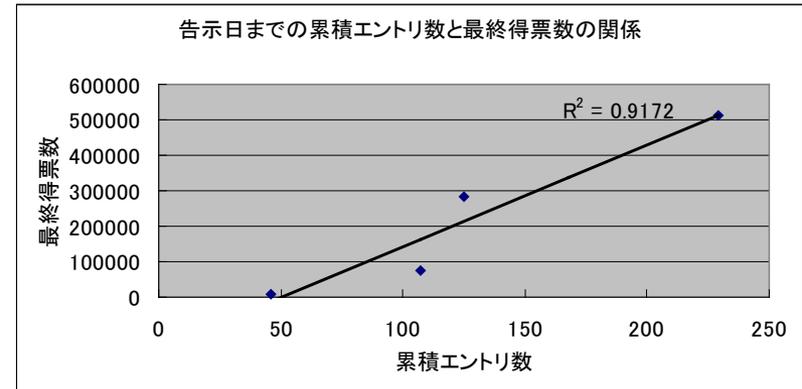


名古屋市長選挙



告示日: 2009年4月12日
投票日: 2009年4月26日

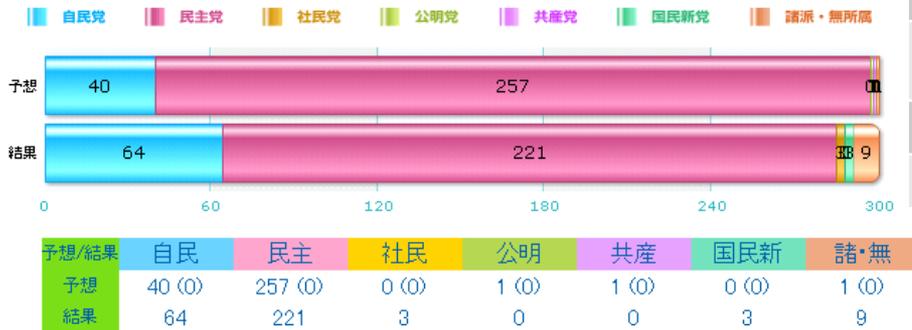
投票日前日までのブログエントリー数と最終得票数の間の
相関係数: 0.95
当選者の河村氏は話題性が高く、告示前から0.9以上の
相関で推移



結果

- ◆ 8月30日時点の予想と総選挙の開票結果は、300選挙区中241選挙区で当選者を的中(80.33%の的中率)
- ◆ 混戦を許容すると、新聞社の予想とほぼ同じ的中率。

予想的中 241 / 300 (80.33%)

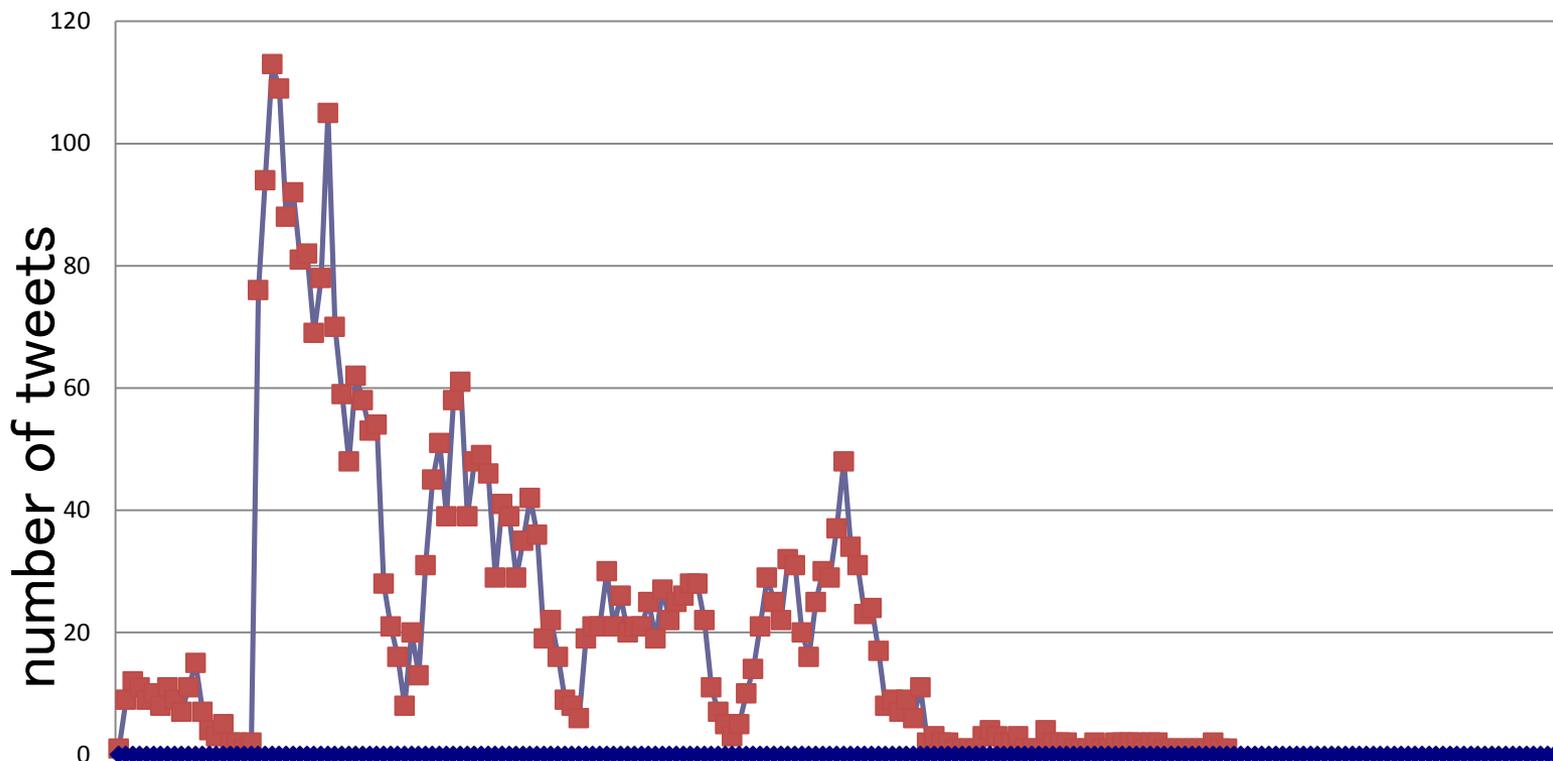


	予測数	正解数	誤り数	混戦	的中率
プロブ分析	300	256	43	0	85.67%
ブログ分析 (混戦あり)	254	241	39	36	94.88%
朝日新聞	260	245	15	40	94.23%
日経新聞	288	264	24	12	91.67%



(クチコミ総選挙, 2009, <http://senkyo.kakaricho.jp/>より引用)

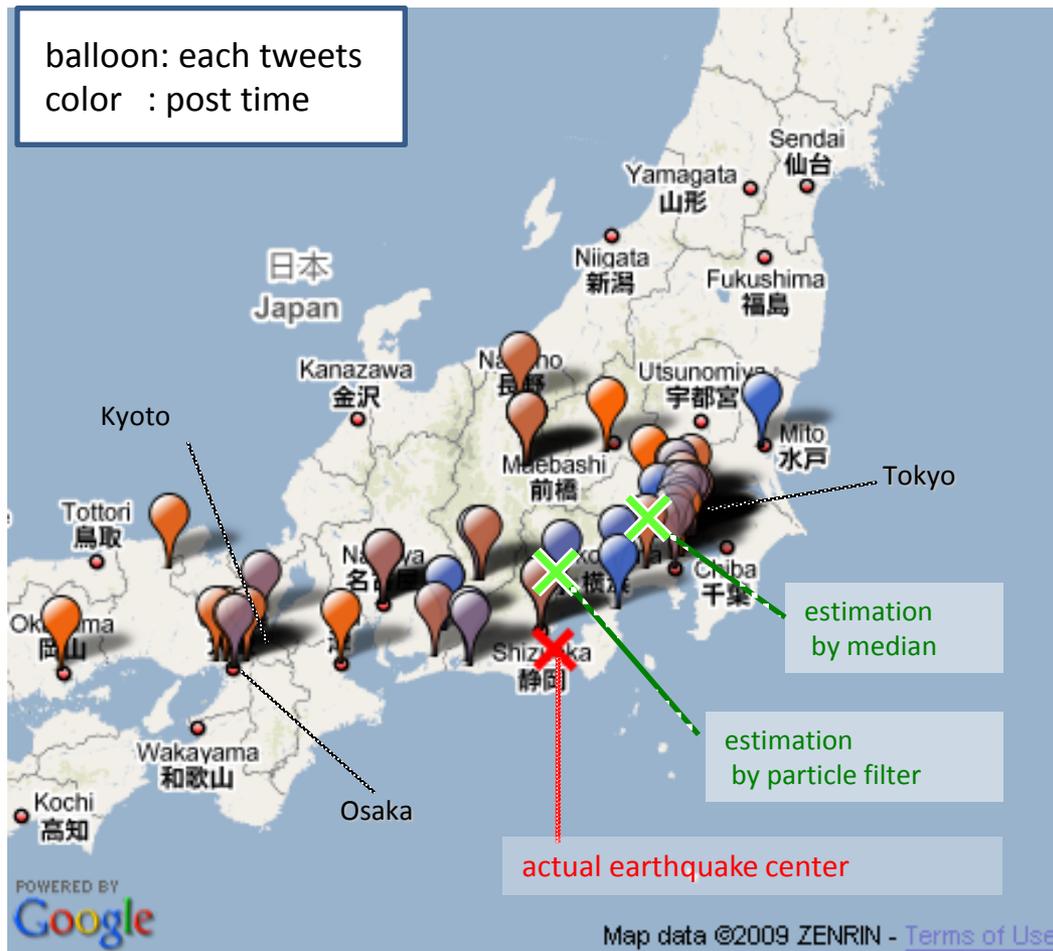
「地震！」「揺れた？」



- 今起こっている地震について述べているつぶやき
- △そうではないつぶやき

それぞれのつぶやきに対して、7割程度の精度で判定できる。

瞬間的なつぶやきの分布 (ソーシャルセンサ)



震度3の地震の96%がつぶやきだけから検知できる。80%は1分以内に検知できる。
Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, Proc. 18th International World Wide Web Conference (WWW2010), April 2010. (370+ citation)



News Front Page



Africa

Americas

Asia-Pacific

Europe

Middle East

South Asia

UK

Business

Health

Science & Environment

Technology

Entertainment

Also in the news

Video and Audio

Programmes

Have Your Say

In Pictures

Country Profiles

Special Reports

Related BBC sites

Sport

Weather

On This Day

Editors' Blog

BBC World Service

Page last updated at 03:04 GMT, Tuesday, 15 December 2009

✉ E-mail this to a friend

🖨️ Printable version

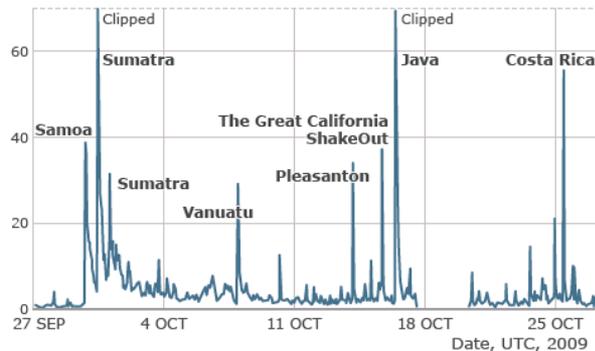
OMG. Did you just feel a quake?

By Jonathan Amos

Science correspondent, BBC News, San Francisco

Spikes in tweet traffic at times of earthquakes

Tweets per minute, 30 minute moving average



Source: US Geological Survey

The Great California Shakeout was a drill and not a real quake event

Tweets are being used by the US Geological Survey (USGS) to get instant public reaction to earthquakes.

The agency is trawling the messages to find out what people felt during a tremor - whether there was a lot of shaking in their area or not.

There are big spikes in Twitter traffic immediately following a quake and the USGS believes emergency responders might find the information useful.

It could help them assess very quickly the severity of a particular event.

However, the survey stressed that the social networking tool would only ever supplement the existing scientific reporting systems which determine shake effects.

ADVERTISEMENT

SEE ALSO

- ▶ Major quakes could be aftershocks
04 Nov 09 | Science & Environment
- ▶ China feels the quake online
16 May 08 | Asia-Pacific
- ▶ Earthquakes weaken distant faults
30 Sep 09 | Science & Environment
- ▶ Typhoons trigger slow earthquakes
11 Jun 09 | Science & Environment

RELATED INTERNET LINKS

- ▶ USGS
- ▶ Twitter
- ▶ AGU Fall Meeting 2009

The BBC is not responsible for the content of external internet sites

TOP SCIENCE & ENVIRONMENT STORIES

- ▶ Asian giants hail Copenhagen deal
- ▶ Data to expose 'ghost mountains'
- ▶ The first glimpse of dark matter?

2009/12/15 BBCニュース 米国地質調査所の研究



カゼミル、エスエス製薬、<http://kazemiru.jp/>より引用

Flu detection using twitter



Traffic event detection
using twitter

経済産業省との取り組み

- マンガやアニメなどのコンテンツを海外に発信する「クール・ジャパン」戦略
- ビッグデータに基づいて、日本製品のプレマーケティングを行うことができないか。
- ひいては、「消費者として優れている」日本の国民性を競争力にできないか
- マンガ・アニメ 5000件についてのアジアでのトレンドを分析、サイトを構築

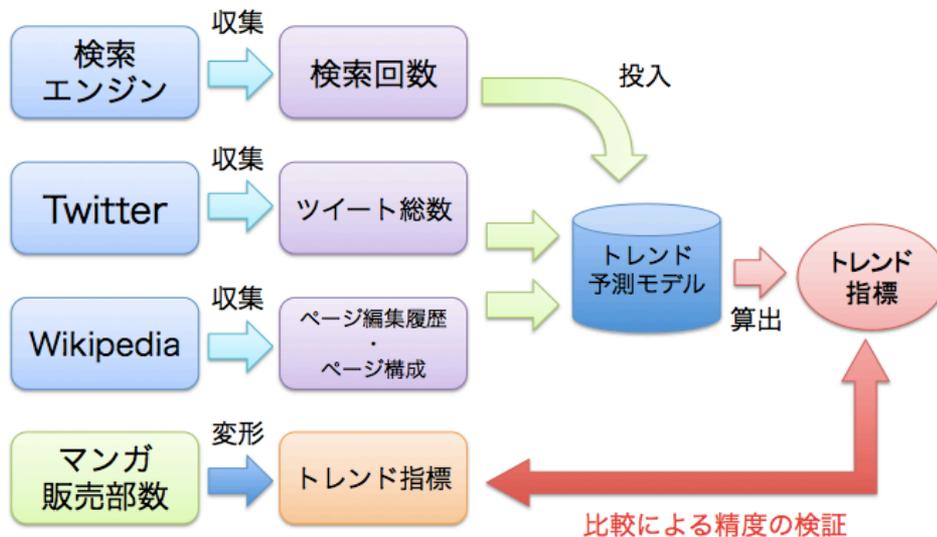


図:トレンド予測モデルの精度評価方法



ASIA TREND MAP, <http://www.asiatrendmap.jp/>より引用

経済産業省との取り組み

- 商務情報政策局、および西山審議官を中心としたメンバー
- マンガやアニメなどの日本製コンテンツを多く海外に輸出していく「クール・ジャパン」戦略が推進されている。
- ビッグデータに基づいて、日本製品のプレマーケティングを行うことができないか。
- ひいては、「消費者として優れている」日本の国民性を競争力にできないか



(出典: <http://www.nagoyanavi.jp/topix1108012.htm>)



(出典: http://news.livedoor.com/article/image_detail/7046139/)

消費トレンドの必要性とアジア消費トレンドマップ

- 日本企業は、アジア諸国をはじめとして、海外展開を急速に進めている
- 各市場における消費性向やトレンドは刻々と変化するため、それをタイムリーに把握し、各国の消費者ニーズにあった商品の開発・販売が十分にできているとは言い難い。

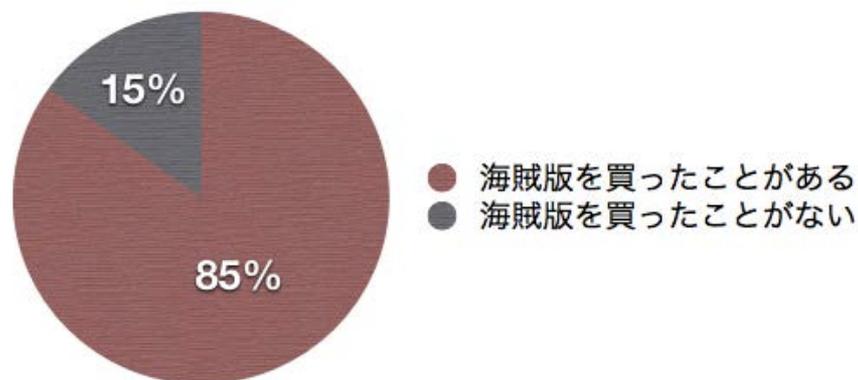


図: 海賊版に対する中国人消費者への意識調査結果

(出典: http://www.nikkeibp.co.jp/style/biz/china/chizai/061107_3rd/index1.html)

- 本研究では、世界各国で利用されるWebサービスをマイニングし、世界における日本製コンテンツの消費トレンドを統一的に予測するシステムを設計する。
- アジア各国における日本製コンテンツの消費トレンド情報を提供するWebサービス「アジア消費トレンドマップ」を提供する。

用いるデータ

- 検索エンジンのキーワード数(どの漫画タイトルが何回検索されたか): 国別
- Twitterのつぶやき回数(どの漫画タイトルが何回つぶやかれたか): 言語別
- Wikipediaの編集回数(どの漫画タイトルが何回編集されたか): 言語別
- これらから、販売部数を予測するトレンド指標を作る

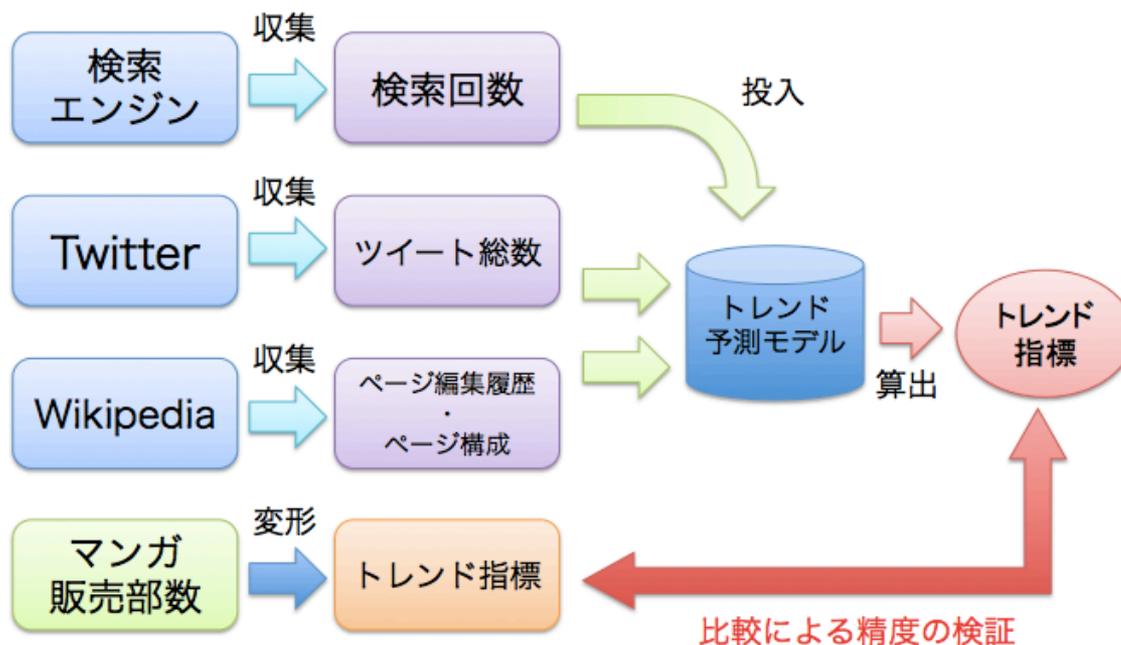


図: トレンド予測モデルの精度評価方法

売上データの加工

- マンガ販売部数はパルス状の変化をするので、売上部数を売上指数に変換し、それからトレンド指標を作成する。

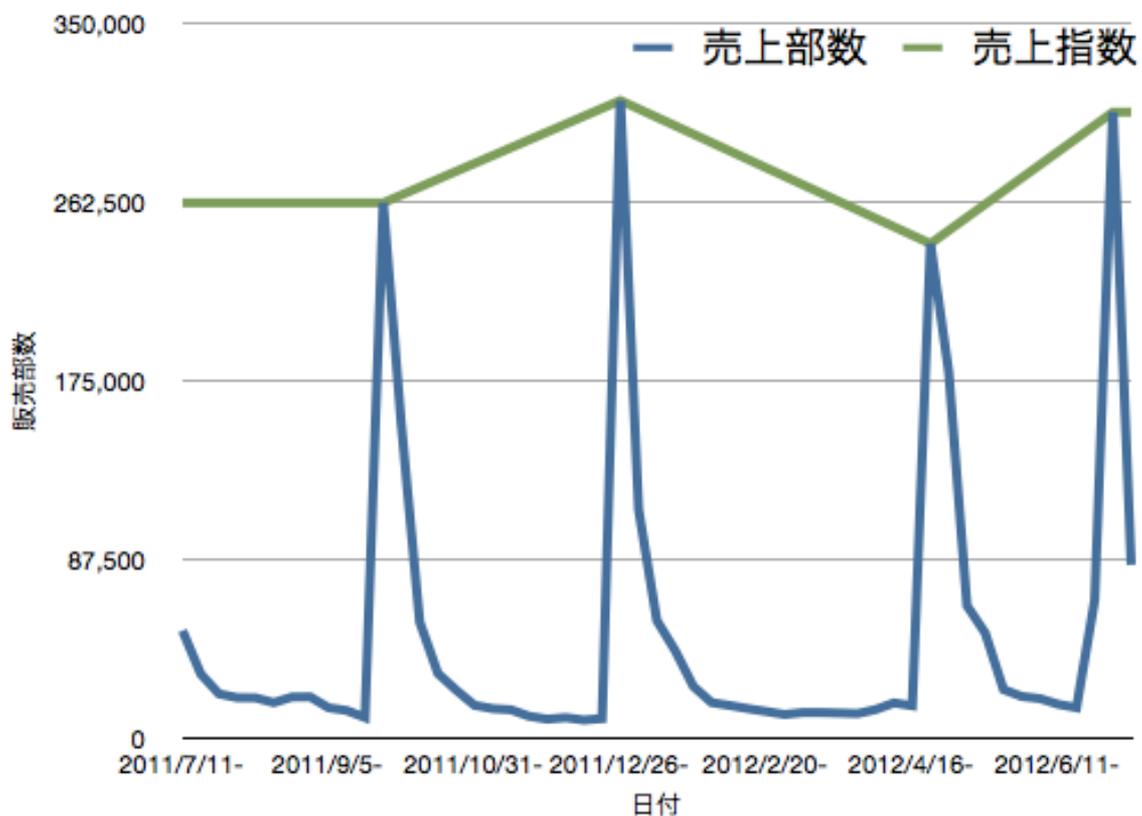


図: 売上部数から売上指数への変換例

予測モデルの精度

- 複数の情報源を組み合わせると、高い精度が得られる。
- 特にWikipediaの編集回数は、トレンドの先行指標となっている。

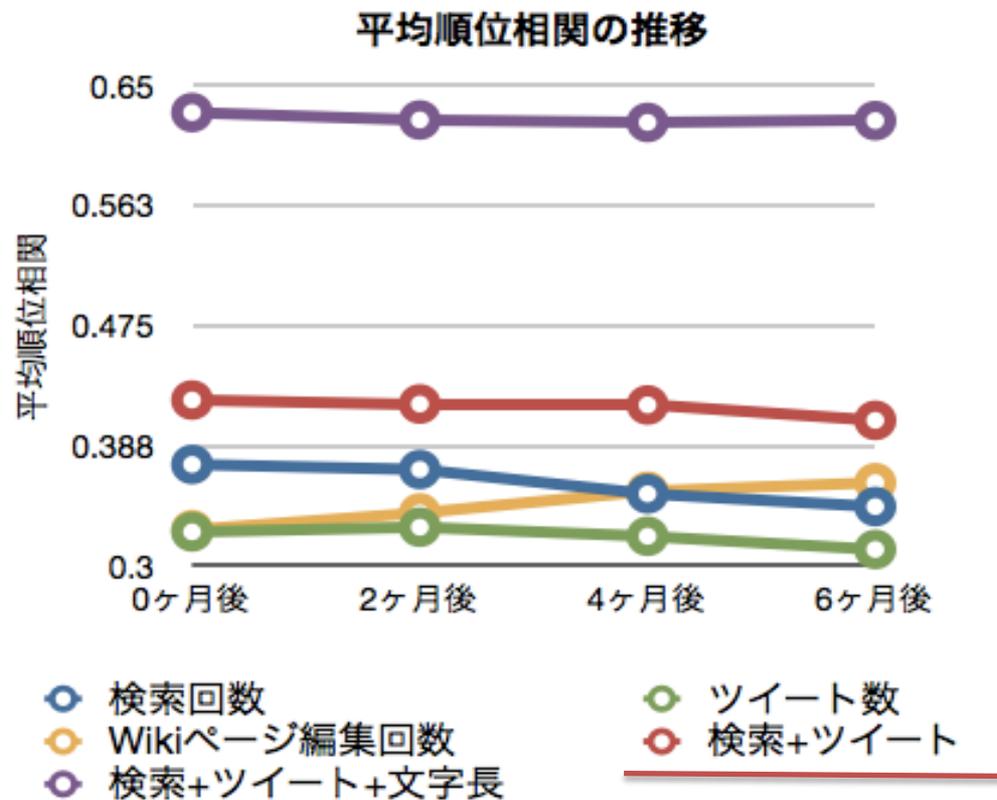


図:使用した素性の組み合わせによる精度の変化

アジア消費トレンドマップ

- 「アジア消費トレンドマップ」を、経済産業省と株式会社チームラボと共同で開発した。
- アジアの14言語圏における日本製コンテンツの消費トレンド指標と、その推移状況を調べることができる。



ASIA TREND MAP, , <http://www.asiatrendmap.jp/>より引用



中国 コンテンツ市場予想

海外コンテンツの輸入には様々な規制が設けられているほか、海賊版も多い。海外コンテンツ企業の進出には困難がつきまとう。

中国の基本情報

人口	GDP	コンテンツ市場規模	コンテンツGDP比較
13億37.2万人	\$5,871,798	5兆5,079億円	1.0%

中国のコンテンツ市場規模うちわけ

書籍	雑誌（広告+販売）	映画興行市場	テレビ放送市場	映像ソフト市場	ゲーム	その他
8,503億円	3,452億円	1,630億円	15,592億円	317億円	5,828億円	317億円

2012/11/14 ヒューマンメディア制作

詳細レポートをダウンロード



中国のHIT予想コンテンツ一覧 コンテンツ数: 295件

一覧データをダウンロード

全て 295件	マンガ 292件	アニメ 129件	ゲーム 16件
------------	-------------	-------------	------------

1 2 3 4 5 ... 6 次の50件 >

コンテンツ	予想指数	現在の人気	過去半年間の人気推移	ジャンル	登録日	更新日
ONE PIECE		32		アニメ マンガ	2012/11/01	2012/12/01
NARUTO -ナルト-		42		アニメ マンガ	2012/11/01	2012/12/01



ASIA TREND MAP

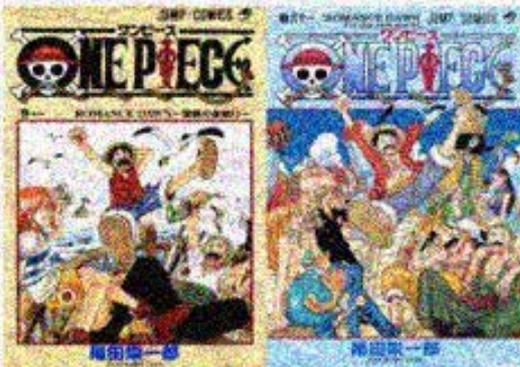
調査国一覧

例: ドラゴンボール

検索

ONE PIECE コンテンツ市場予想

6カ国でONE PIECEの人気が観測されました。中国、ラオス、マレーシア、ロシア、台湾、世界では、ますますONE PIECEがヒットしていきそうです。先月に対してもっともONE PIECEの人気が上がった国は世界、韓国、日本です。先月に対してもONE PIECEの人気下がった国はロシア、台湾、インド、中国、タイ、バングラデシュ、ラオス、ベトナム、マレーシア、フィリピン、インドネシアです。



orangesorangs.jp



『ONE PIECE』（ワンピース）は、尾田栄一郎による日本の少年漫画。および、これを原作としたテレビアニメ、映画、ゲームといったメディアミックス作品。『週刊少年ジャンプ』（集英社）にて1997年34号から連載されている。略称は「OP」「ワンピ」。

提供: Wikipedia

更新日: 2012/12/01

最終日: 2012/11/01



アイコンについて

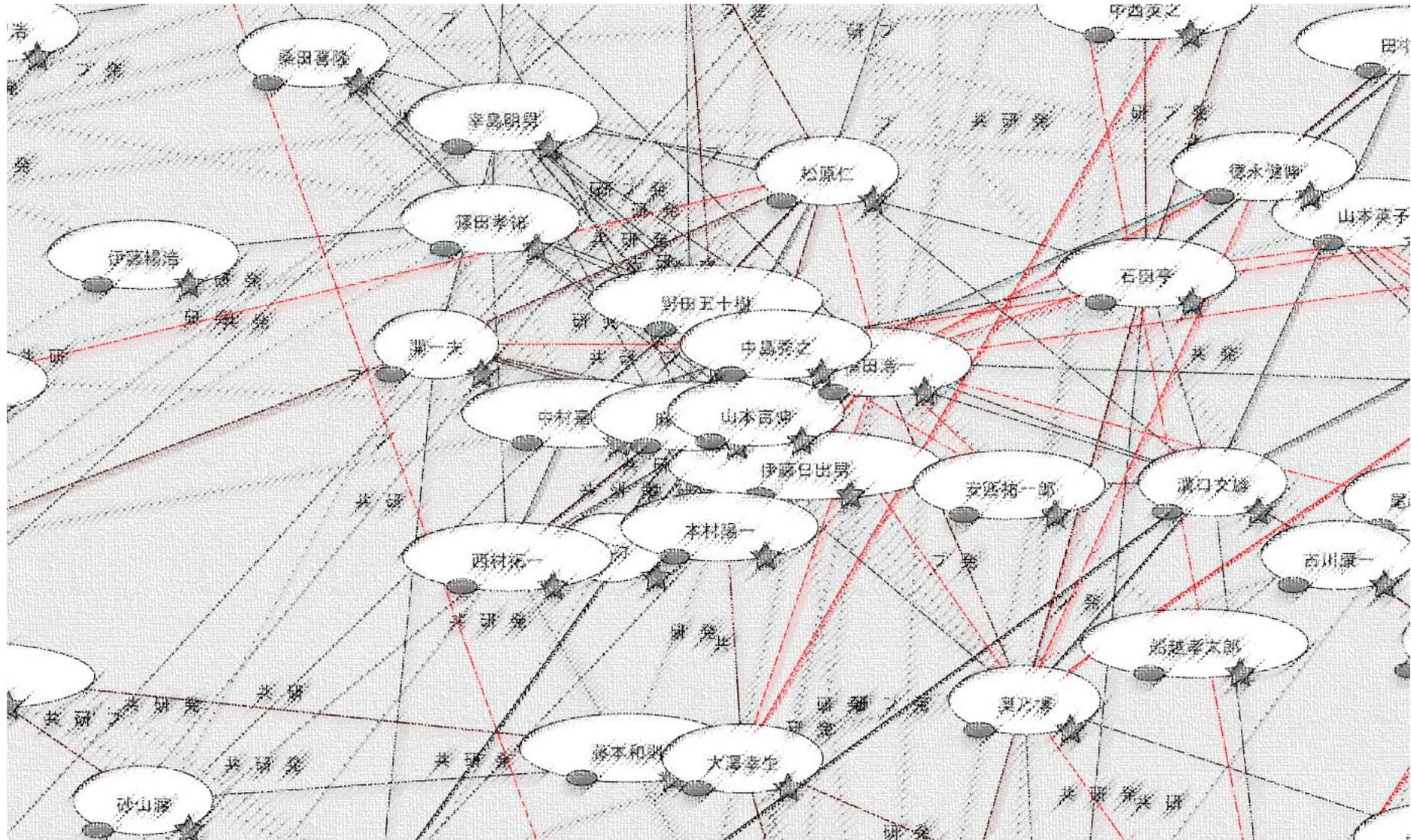
📍 人気急上昇
 📍 人気上昇
 📍 変動無し
 📍 人気下降
 📍 人気急下降

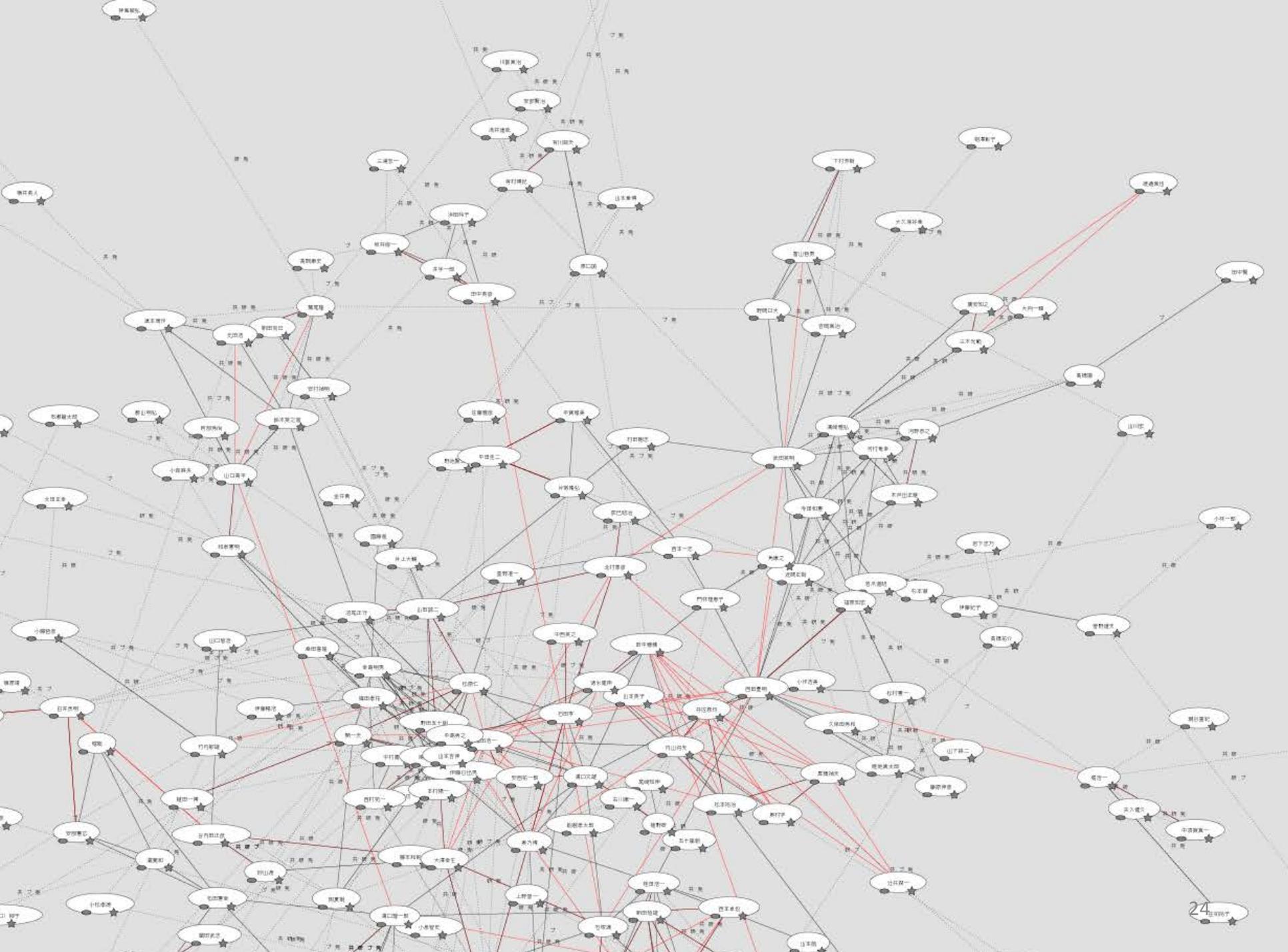
○ 中の数字は前月の人気（1～100）を表しています

国を選択

中国

Web上の情報からの研究者ネットワーク抽出





関係の強さの取得

	YM	IS	MI	MN	MK	...
YM	---	114	3030	1300	441	
IS	114	---	6	50	17	
MI	3030	6	---	281	448	
MN	1300	50	281	---	13900	
MK	441	17	448	13900	---	
...						

Google ウェブ イメージ グループ ディレクトリ

松尾豊 石塚満 Google 検索 検索オプション

ウェブ全体から検索 日本語のページを検索

ウェブ 松尾豊 石塚満 の検索結果 約 124 件中 1 - 10 件目 (0.17 秒)

主要邦文論文 -- 石塚研究室

... 松尾豊, 石塚満: 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム, 人工知能学会論文誌, Vol.17, No.3, pp.217-223 (2002.3) [人工知能学会論文賞受賞] [PDF]. 松尾豊, 石塚満 ...
www.miv.t.u-tokyo.ac.jp/j-pub.html - 16k - 2004年4月24日 - キャッシュ - 関連ページ

松尾豊の研究内容

... 松尾豊, 石塚満: コストに基づく仮説推論の2種の連続値最適化問題への置換法とその協調による推論法, 人工知能学会誌, Vol.16, No.4(B), pp.400-407, 2001 [PDF file] 仮説推論の数学的な制約式への置き換え法が2種類あるよ ...
www.miv.t.u-tokyo.ac.jp/~matsuo/study.html - 30k - キャッシュ - 関連ページ
[他, www.miv.t.u-tokyo.ac.jp内のページ]

松尾豊

... 発表リスト. 6月25日(水) 10:20~10:40 F 小会議室1+2 題目, 1F1-02 Webからの人間関係ネットワークの抽出と情報支援. 著者, 松尾豊 友部 博教 橋田 浩一 石塚満, メモ, Webマイニング 2番目の発表です. 論文, PDFファイル ...
www.carc.aist.go.jp/ESISP/JSAI2003/PROGRAMS/PE00297.HTM - 3k - キャッシュ - 関連ページ

石塚満

石塚満 氏名, 石塚満 ふりがな, いしづか みつる. 所属, 東京大学. 発表リスト. 6月25日(水) 10:20~10:40 F 小会議室1+2 ...
www.carc.aist.go.jp/ESISP/JSAI2003/PROGRAMS/PE00182.HTM - 4k - キャッシュ - 関連ページ
[他, www.carc.aist.go.jp内のページ]

石塚満

石塚満 氏名, 石塚満 ふりがな, いしづか みつる. 所属, 東京大学. 発表リスト. 6月25日(水) 10:20~10:40 F 小会議室1+2 ...
www.kasm.nii.ac.jp/jsai2003/programs/person-182.html - 4k - キャッシュ - 関連ページ

松尾豊

... 題目, 1F1-02 Webからの人間関係ネットワークの抽出と情報支援. 著者, 松尾豊 友部 博教 橋田 浩一 石塚満, メモ, Webマイニ

例)124件

- ✦ 検索エンジンを使って、共起関係の強さを測る。
 - ✦ Jaccard係数、相互情報量などさまざまな尺度がある。

- ✦ 検索されたページから、特徴量(素性)を抽出し関係を把握。
 - ✦ 共著、研究室、プロジェクト、発表

機械学習

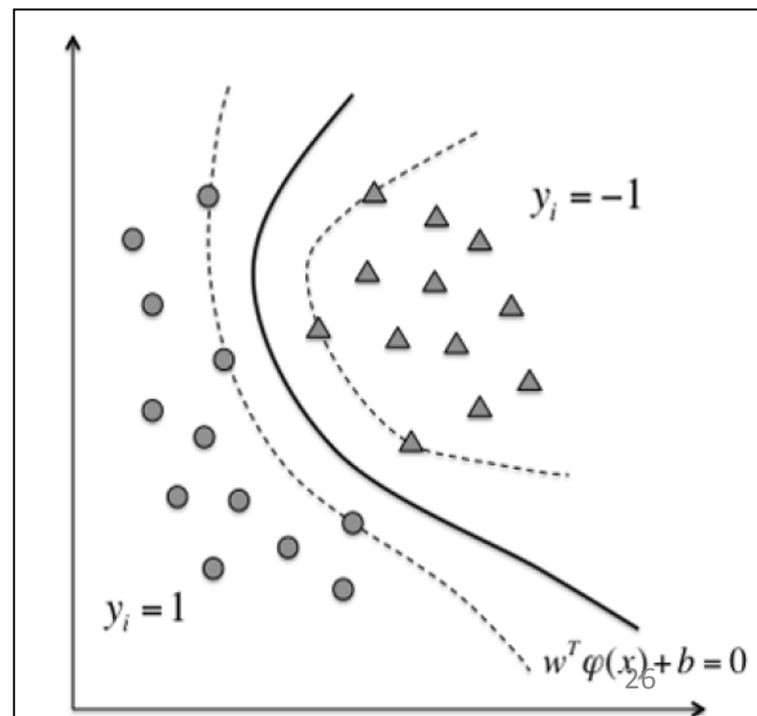
素性(そせい、feature): 特徴を表す量 クラス

事例(example)

X1	X2	X3	X4	X5	...	class
10	2	1.1	1.3	0		○
1	1	-2.1	1.1	0		○
18	4	3.9	5.5	1	1	△
20	13	2.8	1.5	0		△
2	1	-1.0	3.2	1		○
23	2	10.2	2.0	0	1	△
...						

2つのクラスを分離する曲面を見つける。(=学習させる)

いったん学習させれば、新しいデータに対し、
曲面のどちら側に存在するかで、クラスを予測できる



氏名・キーワードで人物を検索

検索

詳細検索



話題のひと 西方凌 水室京介 佐久間象山 渡辺麻友 もっと見る

松尾豊

ツイートする 0

いいね!



ウェブ上からみた人物像

プロフィールをもっと見る・変更する

松尾豊 (准教授) @ymatsuo 研究内容 人工知能とウェブ、ビジネスモデルに関する研究を行っています。ソーシャルメディアのデータ分析や知の構造化、ウェブが社会や経済にどういった影響を与えるかといったテーマを研究しています。自己紹介 1997年 東京大学工学部電子情報工学科卒業、2002年 同大学院博士課程修了、博士(工学)、同年より、産業技術総合研究所研究員、2005年10月よりスタンフォード大学客員研究員、2007年10月より、東京大学大学院工学系研究科総合研究機構 / ...

もっと見る

<http://weblab.t.u-tokyo.ac.jp/members...> より抜粋

航空宇宙工学で米国大進学

アメリカの大学へ高校から直接進学 5/14航空 創価学会の会員様が購入できる墓地 東京エリ

宇宙セミナー開催・NCN

創価学会の会員様が購入できる墓地 東京エリ

ア墓地の資料請求OK=無料



Ads by Google

*画像について



画像をもっとみる

入会金、教材費、iChina.jp

月額基本料すべて無料!



こんな人と関係があります



READYFOR?

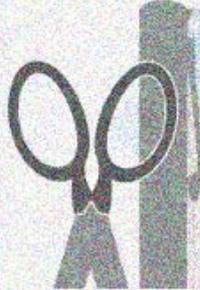
マイページ | ログアウト

プロジェクトを
始める

プロジェクトを
探す

READYFOR?を
もっと知る

2009～ チアスパ
2011～ READYFOR
クラウドファンディング
と言われるようになる前



READYFOR? は、“実行者”を支援する
日本初のクラウドファンディングサービスです。

あなたが実行したいと思ったアイデアを
このサイト上でプレゼンテーションすることで、
多くの人からの支援のチャンスを得ることができます。



- 誰もがやりたいことを実行できる、そんな世界へ -

プロジェクト



プロジェクト名: **Bangladesh 15歳のための雑誌創刊プロジェクト!**
by 三好大助

概要: Bangladesh 15歳たちの「ワクワクするビジョン」を育む雑誌を、現地大学生とともに創刊する。

達成率	達成金額	終了まで
53%	62,000円	64日



プロジェクト名: **ひらのりよのDVD作品集ほしいひとー!お金ちょーだい!!**
by ひらのりよ

概要: Apple主催学生デジタル作品コンテストグランプリに輝いたアニメーターひらのりよの作品を詰め込んだDVDを製作する。

達成率	達成金額	終了まで
83%	96,000円	33日



プロジェクト名: **14番目の偉人はキミ!学校に「DISTANCE」を贈ろう**
by イマライノワエ

概要: ナイチンゲールやエジソンなどをモチーフとした作品「DISTANCE」を学校に届けることで、子供たちにチャレンジの大切さを伝える。

達成率	達成金額	終了まで
53%	161,000円	17日

メディア掲載

2011年 6月 2日
R25 に掲載されました

2011年 5月 29日
TechCrunch Japan に掲載されました

2011年 5月 28日
TechCrunch に掲載されました

2011年 4月 25日

Facebookもチェック facebook

R? Ready for?
「いいね!」あなたが「いいね!」と言っています。

R? Ready for?
「いいね!」あなたが「いいね!」と言っています。

READYFOR?(レディーフォー)

“実行者”を支援する日本初のクラウドファンディングサービスです。
音楽、映画、アート、テクノロジーなどのクリエイティブな活動はもちろん、
夢を持つすべての“実行者”がアイデアをサイト上でプレゼンテーションすることで、
多くの人から少額の支援金を集めることができます。



原発と放射能の见えない恐怖と知っ
ておくべき本当の話

by わたなべりんたろう

参议院行政監視委員会でも証言を行っ
た、京都大学小出裕章助教のインタ
ビューを収録した、原発と放射能につい
てのドキュメンタリー映画を制作する。

プロジェクトが成立しました!

達成率	達成金額	終了日
189%	474,000円	6月30日

放射能の恐ろしさに関する 映画製作の費用を集めるプロジェクト

【引換券】

1,500円:上映会チケット1枚

10,000円:エンドロールに支援者のお名前を掲載

25万円の目標金額に対し、14日間で
73人のスポンサーから、47万円の支援金が集まりました。

途上国で教育事業を行なうプロジェクト

【引換券】

1,000円:子供たちからのサンクスレター

10,000円:実行者による事業報告会への招待券

20万円の目標金額に対し、86日間で
86人のスポンサーから、39万円の支援金が集まりました。



アジア最貧国ドラゴン桜

by 佐野真佳

留校早稲田大生によるバングラデシュ
農村部の高校生に海闊グッガ大学受験へ
のチャンスを与える教育事業コンデンツ
ルを制作するプロジェクト。

プロジェクトが成立しました!

達成率	達成金額	終了日
137%	394,000円	6月29日

世界で広がるクラウドファンディングの流れ

海外では、200以上のクラウドファンディングサービスがあります

Kickstarterは、クリエイティブな活動の資金を集める購入型サービスで、開始2年あまりで、月間取引額が10億円、月間130%で売り上げが伸び、大きな注目を集めています。READYFOR?は、日本版Kickstarterを目指しています。



寄付型: 見返りが無い

投資型: 見返りがお金



MUSIC

There is a lifetime of classical music out there, legally in the public domain, that has yet to be recorded and released to the public. Musopen is out to change that. This project will use your donations to purchase and release music for free use.

代表サイト:
Global giving(途上国への寄付)

Grow VC(エンジェル投資前の資金調達)

購入型: 見返りがお金以外

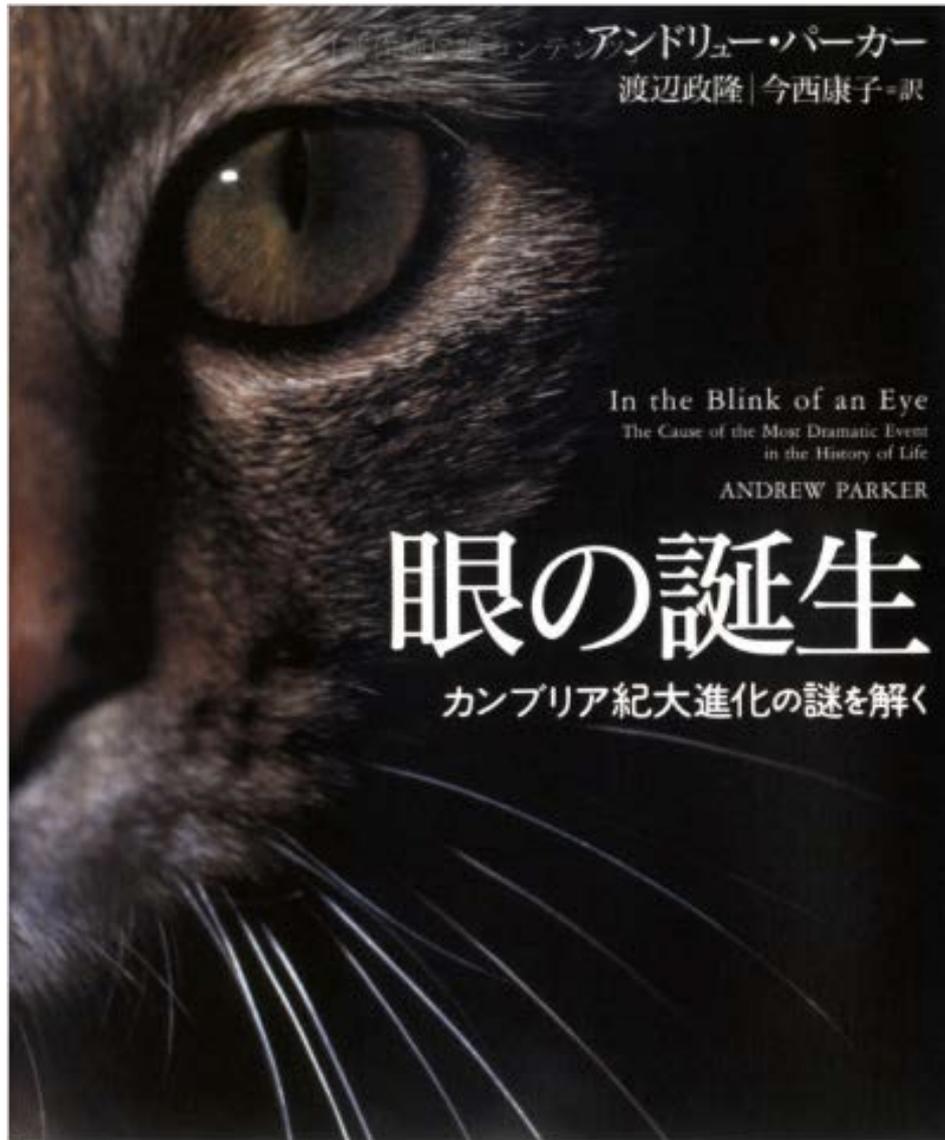


FOOD

"Economy Bites" is an online cooking show that teaches you how to make delicious meals on a budget. Each short episode focuses on a single dish that can be made for less than \$30.

Indie GOGO
Kickstarter(実行資金調達)

Kickstarter



(眼の誕生, 草思社, 2006, 表紙より引用)

カンブリア爆発

- 5億4200万年前から5億3000万年前の間に突如として今日見られる動物の「門(ボディプラン、生物の体制)」が出そろった現象



眼の機能

- 機会と脅威を発見する
 - 食べられるもの、敵
- 常時、情報を取得しているが、ほとんどの情報が無意味
 - \leftrightarrow 嗅覚、触覚
 - その中に一部、価値のある情報が含まれている
 - それを「遠くから」見つけることができる
 - 見つけたものに応じて、アクションを変える

ビッグデータの機能

- 機会と脅威を発見する
 - 食べられるもの、敵
- 常時、情報を取得しているが、ほとんどの情報が無意味
 - \leftrightarrow 嗅覚、触覚
 - その中に一部、価値のある情報が含まれている
 - それを「遠くから」見つけることができる
 - 見つけたものに応じて、アクションを変える

Deep Learningの取り組み

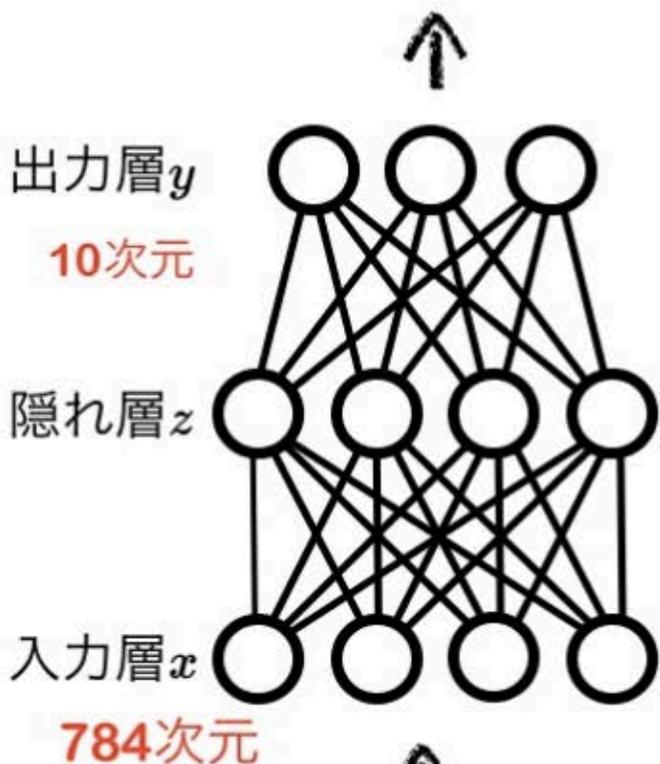
- 人工知能の50年来の画期的なブレークスルー
- 多段のニューラルネットワーク
- 素性を作り出す。「気づき」

普通の手書き文字認識

[0.05, 0.05, 0.05, 0.40, 0.05, 0.05, 0.15, 0.05, 0.15, 0.05]

10次元の確率分布

(左から、入力画像が、
0である確率、
1である確率
...
9である確率)

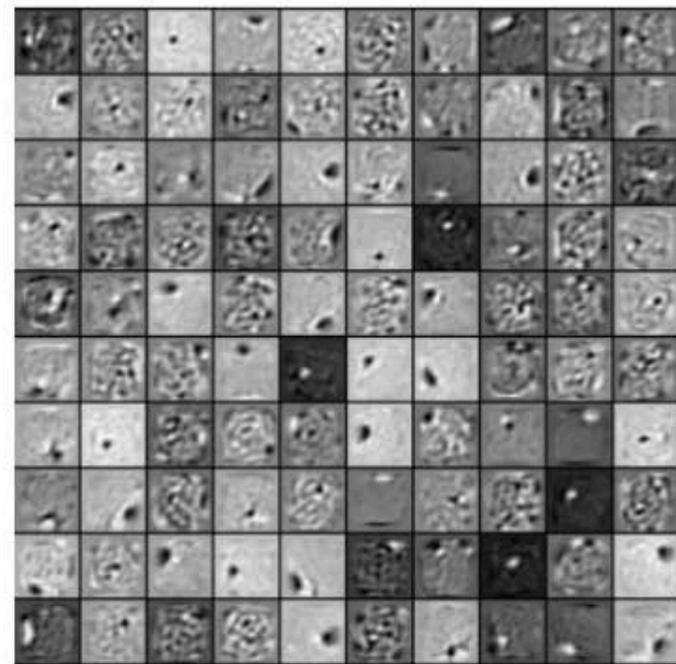
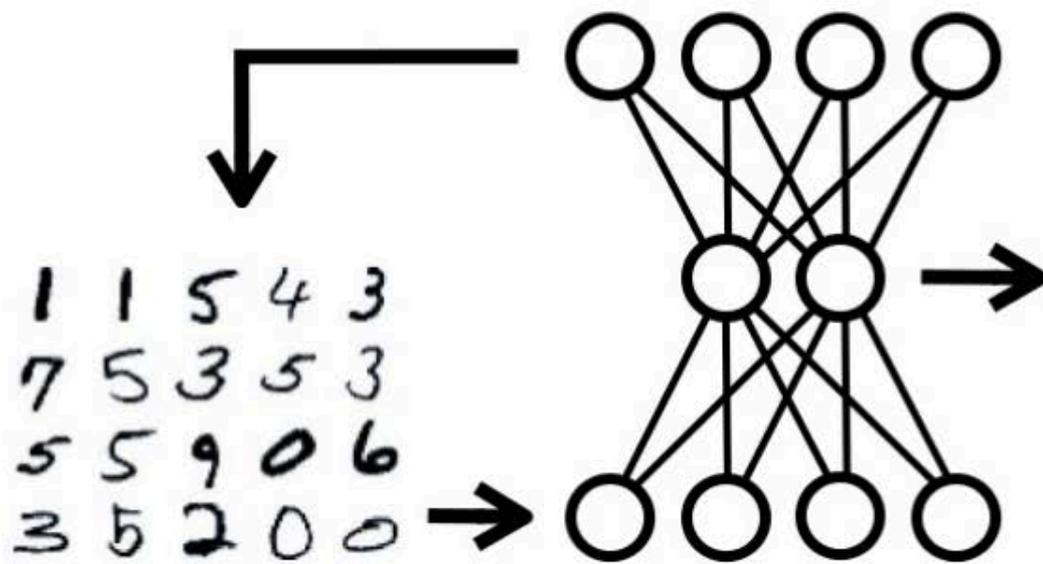


MNIST (28*28の画像)

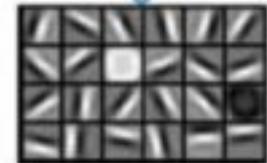
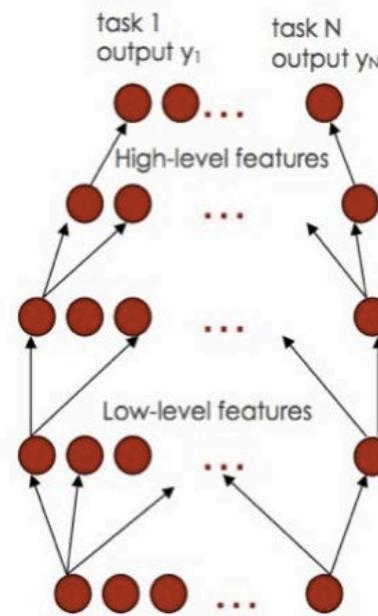
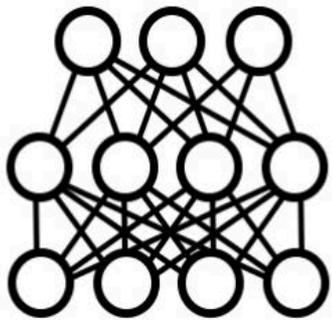
1	1	5	4	3
7	5	3	5	3
5	5	9	0	6
3	5	2	0	0

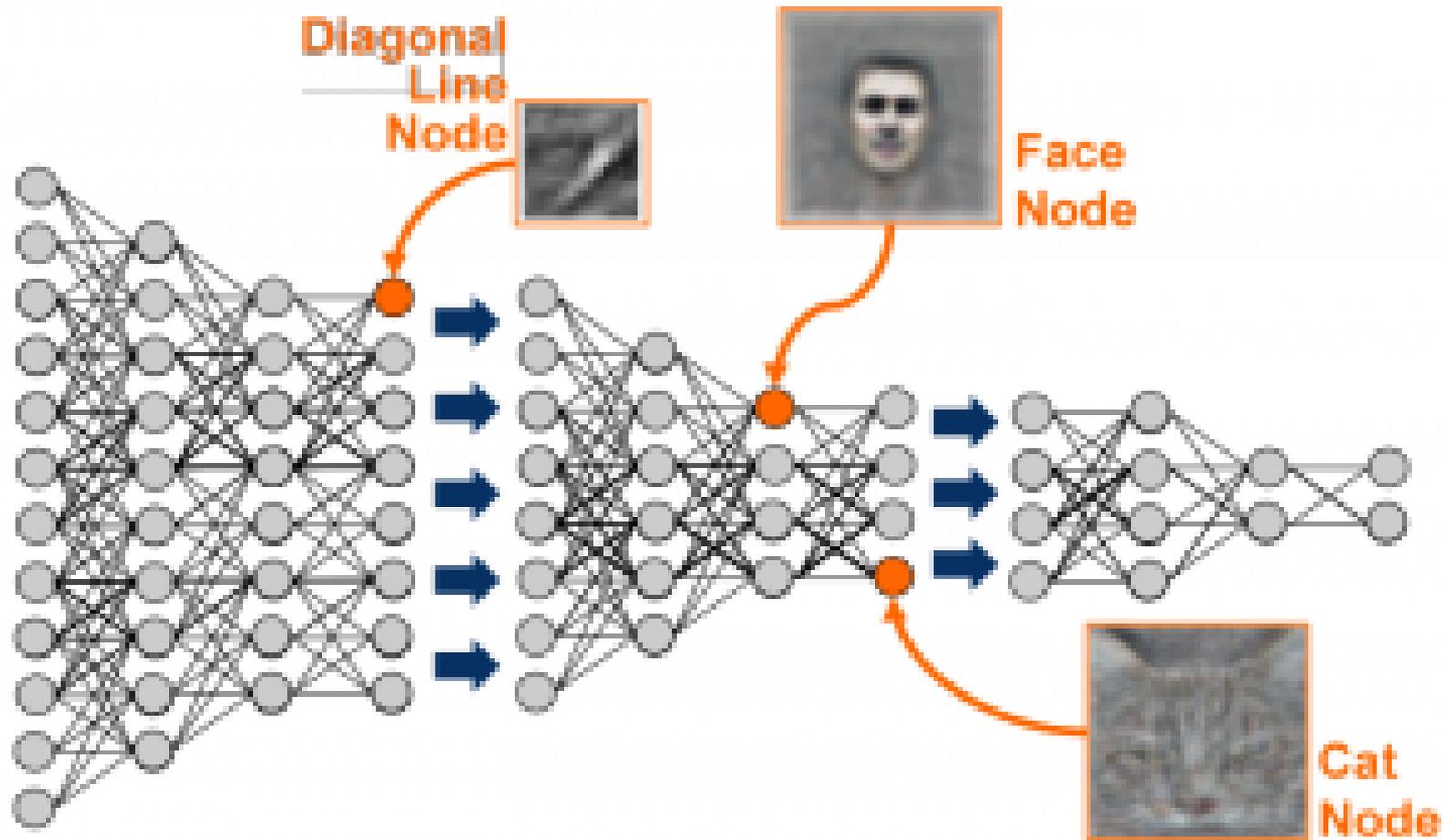
28*28=
784次元の数値ベクトル

Auto-encoderで得られる表現



“Deep”にした場合





これまでの人工知能の壁≡表現獲得の壁

- 難しい問題1:機械学習における素性生成
 - 素性をどう作るの？
- 難しい問題2:フレーム問題
 - ロボットが動くとどうなるかを、どう考えればいいか
 - = if-thenルールで書いた場合の素性をどう作るの？
- 難しい問題3:シンボルグラウンディング問題
 - シマウマがシマのある馬だと、どう理解すればいいか？
 - = 素性をどう作って、それに名前(シンボル)を与えるの？



結局、難しい問題は全部同じことを指しており、
素性(=表現)を、データをもとにいかに作るかという問題。

DL関連の国際会議

- ICLR: International Conference on Learning Representations (昨年から)
- ICML: International Conference on Machine Learning
- NIPS: Neural Information Processing Systems Conference



DL関連のニュース

- Google Brain: Hinton先生と学生をGoogleが買収(2013)
- BaiduがシリコンバレーにDeep Learningの研究所を作る(2013)
- FBに人工知能研究所設立: NY大のYann LeCun教授(2013)
- Yahoo!が画像のDeep LearningのためにLookFlowという会社を買収(2013)
- Googleが、Deep Learningの会社 Deep Mindを500億円で買収(2014)

10 BREAKTHROUGH TECHNOLOGIES 2013

[Introduction](#)[The 10 Technologies](#)[Past Years](#)

ウェブ工学技術

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart. →

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous. →

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child? →

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts. →

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people. →

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss. →

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket. →

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible. →

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases. →

Supergrids

A new high-power circuit breaker could finally make highly efficient DC power grids practical. →

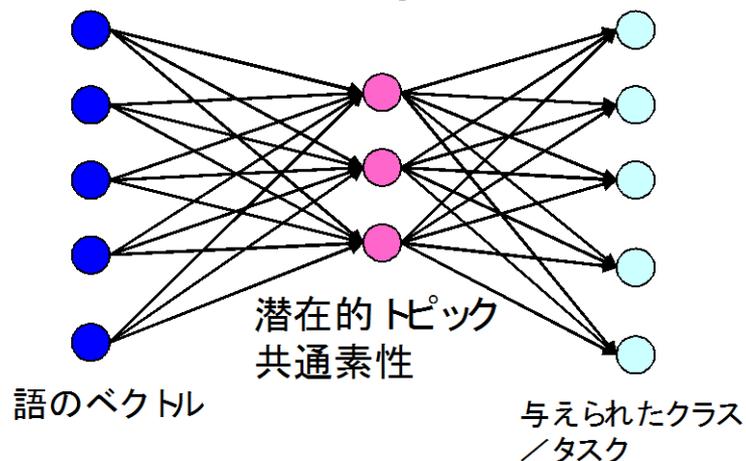
ウェブ工学技術

ディープラーニング
人工知能

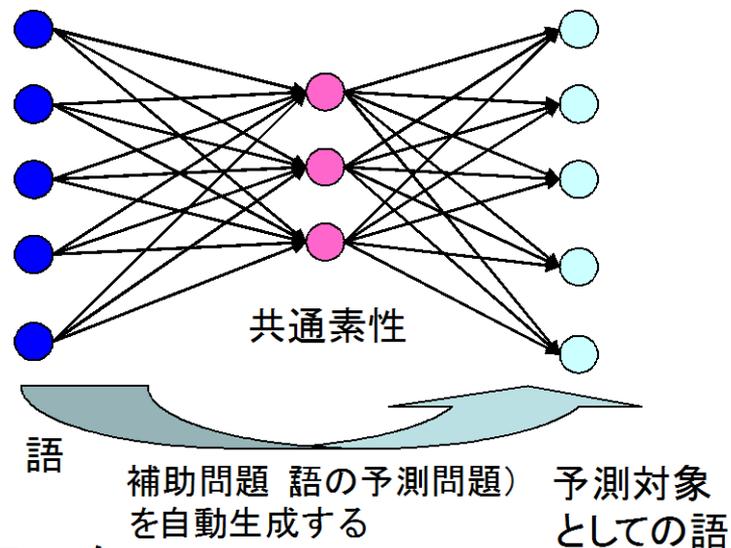
さきがけ提案書 2008

従来手法

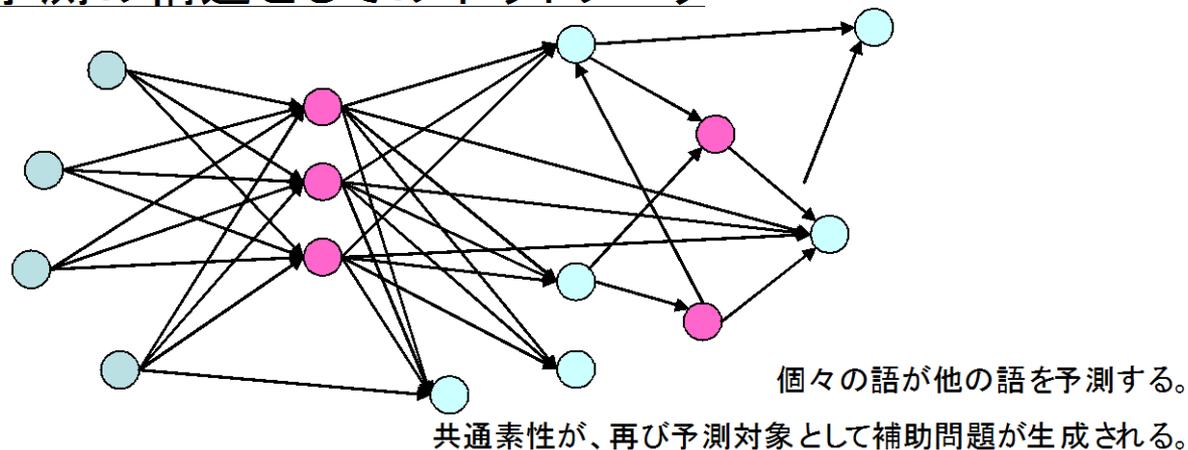
Multi-task learning (Amit et al. ICML'07)
Latent Semantic Indexing等



Andoらの方法 (ICML07, NIPS06等)



提案手法 : 予測の構造としてのネットワーク



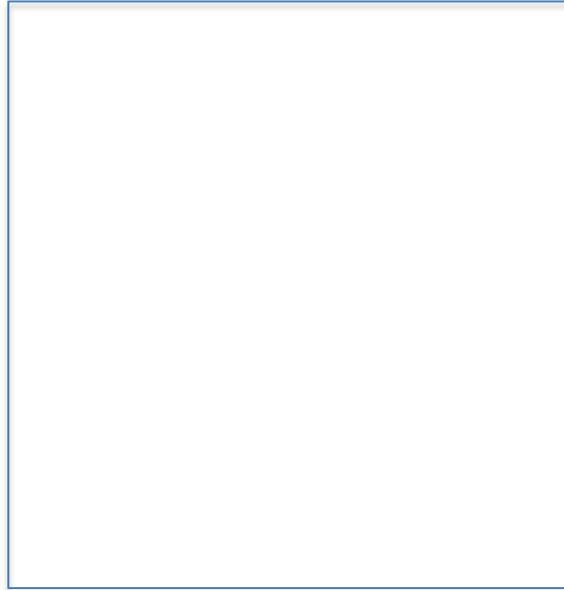
博士論文2002

また、同著ではこのようにも述べられている。

...ところで、内的表現を選ぶということ自体がひとつの型の問題—そして最も扱いにくいもの—なのであるから、問題還元の技法を逆にしてその上に重ねてみようとするかも知れない! そうするためには、抽象的空間のおびただしい変種を表現する方法がなければならないだろうが、これははずばぬけて複雑な企てである。(中略) いずれにしても、人工知能に痛切に欠けているのは、「一歩退いて」何が進行しているのかを見わたすことができ、この展望に立って目前の仕事に対するとり組み方を修正できるようなプログラムなのである。(p.602)

これが書かれたのは20年以上前であるが、これは今でも、人工知能研究の目指すべき大きな空白領域であろう。

この問題に対するアプローチのひとつとして、人の気づきを支援すること、データのどこに着目すればいいかという手がかりをつかむことが考えられる。それは、人間にとっても重要であろうし、コンピュータにより知的な処理をさせるという点からも非常に重要なことであろう。まだ、研究は始まったばかりだが、なぜ人間が森羅万象のデータの洪水の中からあることに「気づく」ことができるのか、興味を持って研究を続けたいと考えている。



5. 高次特徴の言語による
バインディング

シンボルグラウンディング

4. 行為を介しての抽象化

→ 名詞だけでなく動詞

(その様態としての形容詞や副詞)

推論・オントロジー

1. 画像 → 画像特徴の抽象化

2. 観測したデータ(画像+音声+圧力センサー+...) → マルチモーダルな抽象化

3. 自分の行動に関するデータ + 観測したデータ → 行為と帰結の抽象化 ^{プランニング}

身体性

6. バインディングされた言語データの大量の入力 → さらに抽象化、言語理解、自動翻訳 ^{言語}

超知能の実現に向けて

- 何がこの先にあるか
 - 行動と観測のデータ入力
 - 概念と言語ラベルのグラウンディング
 - 大量の言語表現の入力
 - ...
- 人工知能で唯一難しかった問題(表現の獲得)に、突破の糸口が見えた
- ディープラーニング: 大脳新皮質
 - 脳の各構造をどのように作っていけばよいか?
- 20XX年?に、人間を超える? 社会とAIの関わり
- 国内では実は人材が多い。大きな動きを作っていければ。

ソーシャルメディアからの社会観測→高度なAIによる予測技術