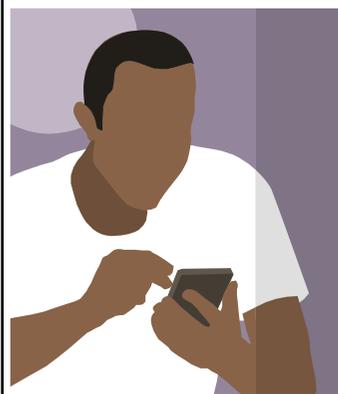
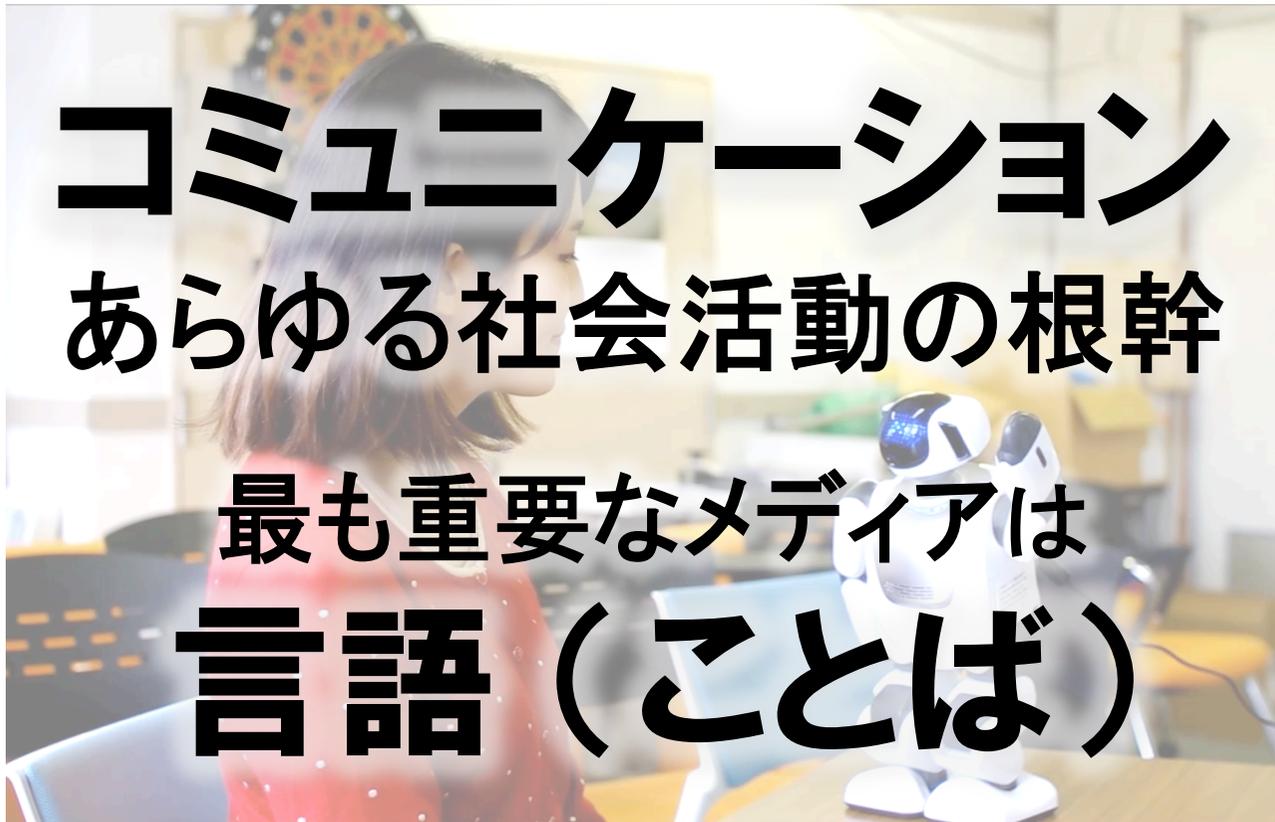


言葉がわかるコンピュータを 創るには

～言葉の不思議と自然言語処理の最前線～

乾 健太郎

情報科学研究科
情報知能システム総合学科
(電気情報物理工学科に改称予定)



自然言語処理

人工知能

言葉がわかる賢いコンピュータを作る

言葉 (= 自然言語) で伝達される情報を
理解・検索・抽出・翻訳・整理・分析し、

地球規模のコミュニケーションを**支援**するソフトウェア技術



Research - Tohoku Universit x

www.cl.ecei.tohoku.ac.jp/en/index.php?Research

これは 英語 のページです。翻訳しますか? いいえ 翻訳 オプション x

AND OR Search

Communication
Science
Laboratory

Tohoku University
Graduate School of Information
Sciences
System Information Sciences
Intelligent Information Science

English | Japanese

Members
Address
Research
Publications
Albums

New / Edit / Diff / Rename / Upload
Backup / Freeze / List of pages

Top / Research

Research Area ⁺

The most important means for communication are the languages that we use everyday, like Japanese and English. In this lab, we conduct research in the following areas: (i) theoretical research to clarify and model the mechanism of communication, namely, what it means to understand language and the conditions that make communication possible (ii) Natural Language Processing research on the development of software that automatically processes the information and knowledge that is represented and transmitted in language; and (iii) applied research supporting intellegent communication or information analysis for the benefit of mankind. We approach an understanding of human intellegence from the perspective of natural language.

Mathematical Models of Language and Communication ⁺

Constructing statistical models that capture the properties of language and the mechanism of communication can be an effective way to incorporate semantic language analysis in applications such as advanced language understanding and information analysis.

Foundational Technology for Natural Language Processing ⁺

recent(10)

- 2013-02-15**
◦ [Address](#)
- 2013-02-11**
◦ [研究発表一覧](#)
- 2012-10-11**
◦ [FrontPage](#)
- 2012-09-09**
◦ [Members/FormerMemb](#)
- 2012-05-29**
◦ [Members](#)
- 2011-10-12**
◦ [MenuBar](#)
- 2011-02-19**
◦ [RecentDeleted](#)
- 2011-02-17**
◦ [Publications](#)
- 2010-08-03**
◦ [MoinMoin](#)
- 2010-08-03**
◦ [Research](#)



Communication
Science
Laboratory

東北大学

情報科学研究科
システム情報科学
知能情報科学

英語 | 日本語

- メンバー
- アドレス
- 研究
- 出版物
- アルバム

新しい / 編集 / 差分 / リネーム / アップロード
バックアップ / フリーズ / ページのリスト

トップ / 研究

研究領域 +

コミュニケーションのための最も重要な手段は、日本語と英語のように、我々が日常使用している言語です。このラボでは、以下の分野で研究を行う者で、(i) 理論的な研究を、それが上の言語とコミュニケーションが可能になる条件 (ii) の自然言語処理の研究を理解することが何を意味するのか、すなわち、コミュニケーションのメカニズムを解明し、モデル化するおよび (iii) 応用研究は、人類の利益のためにintelligent通信や情報分析を支援する。自動的に言語で表現されて伝送される情報や知識を処理するソフトウェアの開発。私たちは、自然言語の観点から人間Intelligenceドキュメントの理解に近づく。

言語とコミュニケーションの数理モデル +

言語のプロパティを取得し、通信のメカニズムは、高度な言語理解と情報分析などのアプリケーションで、セマンティック言語解析を組み込むために有効な方法であることができる統計モデルを構築する。

自然言語処理のための基礎技術 +

言語情報を処理するために、我々は自動的に精度と丈夫さ、文章の統語構造（文章中の単語間の文法的関係、すなわち）とその意味内容（つまり、誰が誰に何をやった）で分析することができなければなりません。それは、文書内の漏れを埋めるとその組織構造を認識することも必要である。我々は、統語的意味的な、と談話分析のこの種の大規模な調査を実施しています。ま

最近の (10)

- 2013年2月15日
 - アドレス
- 2013年2月11日
 - 研究発表一覧
- 2012年10月11日
 - FrontPageの
- 2012年9月9日
 - メンバー/
FormerMembers
 - メンバー
- 2012年5月29日
 - MenuBarの
- 2011年10月12日
 - RecentDeleted
- 2011年2月19日
 - 出版物
- 2011年2月17日
 - MoinMoinの
- 2010年8月3日
 - 研究

加速する人工知能研究

日本経済新聞
2012年11月11日

様々な難題に臨む

2009年



サッカーをする
二足歩行ロボット

人工知能研究の主な歩み

1950年ごろ 世界で人工知能研究が始まる

50年代後半 チェスを指せるようになる

70年代 開発の難しさから「冬の時代」に

80年代 チェス専用コンピューターの
開発が盛んに

93年 国内研究者らが「2050年までに
サッカーで人間の世界王者に勝つ
ロボットを開発」と表明

97年 ■ チェス
世界王者に
勝ち越し

■ オセロの
世界王者に
勝利



GARRY
KASPAROV

DEEP
BLUE

ロイター共同

2010年

将棋のプロ棋士に勝利

2011年

米クイズ番組の歴代王者に勝利



2010

2011



\$2,600

\$3,400

\$1,200

KEN

WATSON

BRAD

国立情報学研などが2021年に
東大入試を目指す計画発表

2012年

はこだて未来大研究者らが
小説執筆の構想を発表

Watson 米クイズ番組で歴代王者に勝利



<http://www.youtube.com/watch?v=Wq0XnBYC3nQ>

Watson 米クイズ番組で歴代王者に勝利

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE
PRINCIPALITIES OF
WALLACHIA AND
MOLDAVIA" INSPIRED
THIS AUTHOR'S MOST
FAMOUS NOVEL

<http://www.youtube.com/watch?v=Wq0XnBYC3nQ>

ウィリアム・ウィルキンソンの『ヴィラキアとモルダビア公国の理由』に最も刺激を受けた19世紀の小説家は？



Google音声検索



Siri



VoiceTra

ロボットは東大に入れるか

Todai Ro

ム
ース
ント
活動
会
学
科
プロジェクトメンバー
語
語
携基盤
究成果
ター入試タスク
い合わせ

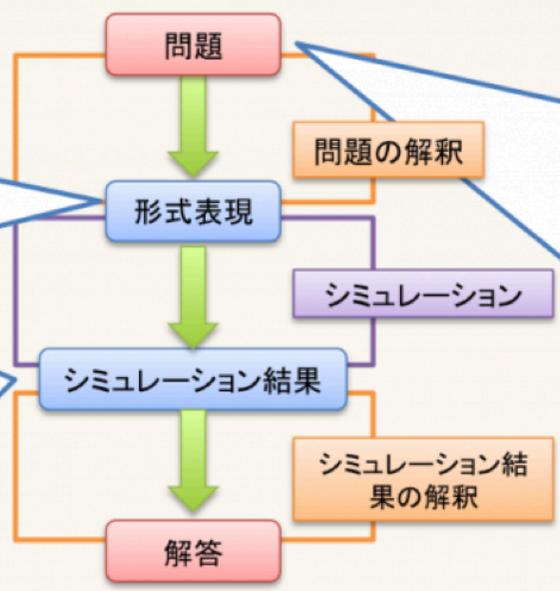
実世界へのグラウンディングに挑む理科問題への解答

理科，特に物理の問題を解くためには，自然言語で記述された物理的な状況を理解し，将来的にどのようなイベントが展開しながら，計算をしたり，理由を述べる必要があります。他の教科とは異なる戦略が求められます。鍵となる技術は大きく分けると，(1)自然言語から物理状態，および，時系列的に生じるイベントの論理式表現への変換，(2)物理シミュレータを用いた将来予測のシミュレーション，(3)シミュレーション結果の解釈の3つに分けられます。

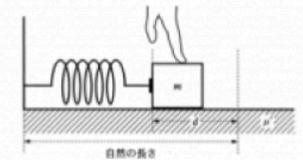
```

∃x(floor(o) ∧ mass(a) ∧ weight(a, m)
  ∧ fixpoint(p) ∧ spring(s)
  ∧ const_spring(s, k) ∧ connect(p, ep1(s), e1)
  ∧ hand(h) ∧ shorten(s, d, e1)
  ∧ distance(a, ep2(s), 0, e1) ∧ hold(a, h, e1)
  ∧ release(a, h, e2) ∧ move(a, e2) ∧ stop(a, e3)
  ∧ ga(g) ∧ dfc(o, a, u)
  ∧ distance(a, ep2(s), x, e3))
    
```

Time	x	Velocity of a
...
10.786	1.27535595	0.00088038
10.788	1.27535751	0.00068438
10.790	1.27535868	0.00048838
10.792	1.27535946	0.00029238
10.794	1.27535985	0.00009638
10.796	1.27535990	0.00000000
10.798	1.27535990	0.00000000
...



あらい水平面上での質量mの物体の運動を考える。図3のように，一端を固定したばねを自然の長さからdだけ縮め，他端に物体を置いて手で押さえた。次に，手を離すと物体は初速度0で動き始め，やがてばねから離れて，手を離れたときの位置からxだけ進んで止まった。xとして正しいものを，下の①～⑥のうちから一つ選べ。ただし，ばねの質量は無視できるものとする。また，ばね定数をk，面と物体の間の動摩擦係数を μ' ，重力加速度の大きさをgとする。



- ① $\frac{k}{2\mu' mg} d^2 + d$
- ② $\frac{k}{\mu' mg} d^2 + d$
- ③ $\frac{k}{2\mu' mg} d^2$
- ④ $\frac{k}{\mu' mg} d^2$
- ⑤ $\frac{2\mu' mg}{k}$
- ⑥ $\frac{\mu' mg}{k}$

(2009年度大学入試センター試験追試験 物理 I より引用)

人工頭脳「東ロボくん」、センター模試英語で「偏差値50.5」 国公立4校でA判定

人工頭脳プロジェクト「ロボットは東大に入れるか」で、人工知能「東ロボくん」が代ゼミセンター模試に挑戦したところ、英語で偏差値50.5と平均を超え、全体成績では国公立4校6学部でA判定だったという。

<http://www.itmedia.co.jp/news/articles/1410/30/news130.html>

[ITmedia]

印刷/PDF

チェック

チェック

8+1 12

8+1 12

通知

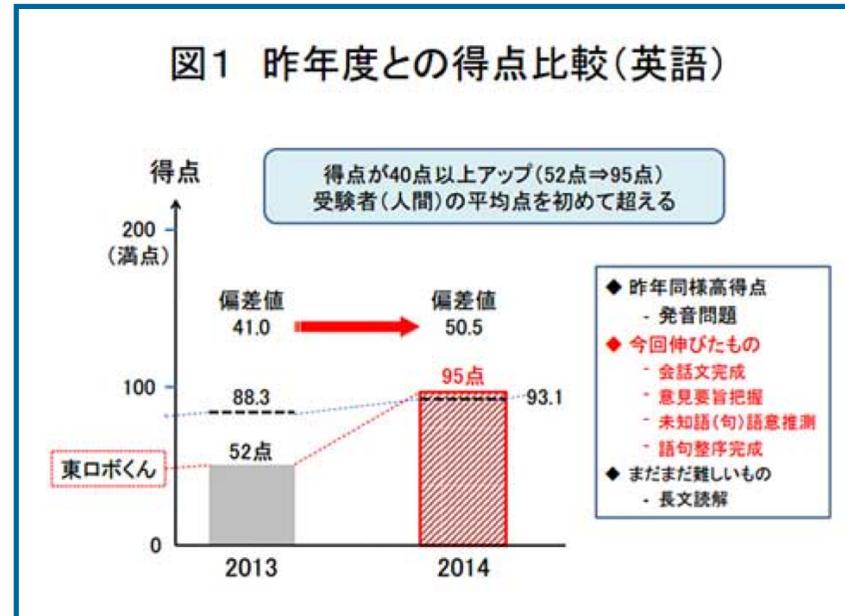
PR 11/12 情報セキュリティ最新動向がわかる一日：FOCUS JAPAN

PR 小室淑恵氏に聞く、生産性を上げ、結果を出す仕事術

国立情報学研究所（NII）とNTT（持ち株会社）は10月30日、人工頭脳プロジェクト「ロボットは東大に入れるか」で、人工知能「東ロボくん」が代々木ゼミナールの大学入試センター試験模試の英語に挑戦した結果、偏差値が50.5と平均を超える好成績を達成したと発表した。

英語以外の教科のチームとの共同による全体成績では国公立大学4校6学部について「A判定」（合格可能性80%以上）に到達したという。

図1 昨年度との得点比較(英語)



英語成績が躍進 = ニュースリリースより

TPP交渉

IBM、人工知能「ワトソン」事業に1000億円投資

2014/1/10 1:43

[小](#) [中](#) [大](#) [保存](#) [印刷](#) [リプリント](#)

【ニューヨーク=小川義也】米IBMは9日、人工知能を搭載した高性能コンピューター「ワトソン」の本格的な事業化に向け、10億ドル（約1050億円）を投資すると発表した。ワトソンの用途開発に特化した2000人規模の事業部門を新たに設立。「学習するコンピューター」と呼ばれるワトソンの普及を目指す。

IBMが2011年に開発したワトソンは、話し言葉による自然な文章を理解し、膨大なデータベースから瞬時に最適な答えを導き出す能力を持つ。米国の人気クイズ番組で、人間のチャンピオンに勝ったことで一躍有名になった。

新設する「IBMワトソングループ」はワトソンを開発したIBM研究所の研究者に加え、ソフト、IT（情報技術）サービス、ハードの各事業部門出身のコンサルタントや営業担当で構成。ソフトウェア・ソリューション部門を率いるマイク・ローディン上級副社長がトップに就任する。10億ドルの投資の

自然言語処理

言葉がわかる コンピュータ



「言葉がわかる」とは？ 単語を知っている、単語に分けられる

そのひとことで元気になった。

そのひとことで元気になった。

そのひと(人)こと(事)で元気になった。

そのひ(目)とことで元気になった。

どうすれば正しい読み方を

選べるようになるか？

そのひとことで元気になった。

「言葉がわかる」とは？

構文構造(修飾関係)がわかる

(名詞)と(名詞)に(動詞)

彼女 と 仙台 に 行きました

彼女と 仙台に 行きました

I visited Sendai with her

盛岡 と 仙台 に 行きました

盛岡 と 仙台 に 行きました

I visited Mori. and Sendai

どうやってコンピュータに 言葉を教えるか



基本戦略①:

答えつき例題から答えの傾向を学ばせる

例題と答え(教師データ)

言語 と は 、 コミュニケーション の た め の 記号 の 体系 で ある
名詞 助詞 助詞 記号 名詞 助詞 名詞 助詞 名詞 助詞 名詞 助動詞 助動詞

自然 言語 処理 は 、 人間 が 使っ て いる 言葉 を コンピュータ に
名詞 名詞 名詞 助詞 記号 名詞 助詞 動詞 助動詞 助動詞 名詞 助詞 名詞 助詞

落語 は 、 江戸 時代 の 日本 で 成立 し 、 現在 まで 伝承 さ れ
名詞 助詞 記号 名詞 名詞 助詞 名詞 助詞 サ変 動詞 記号 名詞 助詞 サ変 動詞 助動詞

都市 に 人口 が 集積 す る こ と に よっ て 芸能 と し て 成立 し た
名詞 助詞 名詞 助詞 サ変 動詞 名詞 助詞 動詞 助詞 サ変 助詞 動詞 助詞 サ変 動詞 助詞



統計解析(機械学習)

規則性(答えの傾向)

- 名詞の後には助詞が一番続きやすく、つぎに名詞が続きやすい
- 助詞の後には名詞が一番続きやすく、つぎに動詞が続きやすい
- 「が」「を」「に」「で」という文字は助詞になりやすい
- 長い単語が候補にあるなら、短い単語に細切れにするよりも良い

基本戦略②

大量の生データから得られる特徴を利用

彼女と 仙台に 行きました



「彼女と行く」はよく言う(560万回)

盛岡と 仙台に 行きました



「盛岡と行く」は言わない(23回)

彼女 と 仙台 に 行きました



「彼女や仙台」は言わない(4回)

盛岡 と 仙台 に 行きました



「盛岡や仙台」はよく言う(20万回)

実際の解析例

米航空宇宙局は太陽系外に存在する観測史上最小の惑星を発見した。

* 0 5D 3/4 3.287031

米	名詞,固有名称,地域,国,*,*,米,ベイ,ベイ,,	B-ORGANIZATION
航空	名詞,一般,*,*,*,航空,コウクウ,コークー,,	I-ORGANIZATION
宇宙	名詞,一般,*,*,*,宇宙,ウチュウ,ウチュー,,	I-ORGANIZATION
局	名詞,接尾,一般,*,*,*,局,キョク,キョク,,	I-ORGANIZATION

固有表現抽出

は 助詞,係助詞,*,*,*,は,ハ,ワ,, 0

* 1 2D 1/2 1.773937

太陽系	名詞,一般,*,*,*,太陽系,タイヨウケイ,タイヨーケイ,,	0
外	名詞,接尾,一般,*,*,*,外,ガイ,ガイ,,	0
に	助詞,格助詞,一般,*,*,*,に,ニ,ニ,,	0

項

ガ

項

ニ

述語

(存在する)

述語-項
構造解析

ガ

項

ヲ

述語

(発見する)

係り受け
解析

* 2 3D 1/1 0.000000

存在	名詞,サ変接続,*,*,*,存在,ソンザイ,ソンザイ,,	0
する	動詞,自立,*,*,サ変・スル,基本形,する,スル,スル,,	0

* 3 4D 2/3 0.000000

観測	名詞,サ変接続,*,*,*,観測,カンソク,カンソク,,	0
史上	名詞,一般,*,*,*,史上,シジョウ,シジョー,,	0
最小	名詞,一般,*,*,*,最小,サイショウ,サイショー,,	0
の	助詞,連体化,*,*,*,の,ノ,ノ,,	0

* 4 5D 0/1 0.000000

惑星	名詞,一般,*,*,*,惑星,ワクセイ,ワクセイ,,	0
を	助詞,格助詞,一般,*,*,*,を,ヲ,ヲ,,	0

* 5 -1D 1/2 0.000000

発見	名詞,サ変接続,*,*,*,発見,ハツケン,ハツケン,,	0
した	動詞,自立,*,*,サ変・スル,連用形,する,シ,シ,,	0
た	助動詞,*,*,*,特殊・タ,基本形,た,タ,タ,,	0
。	記号,句点,*,*,*,。,,	0

EOS



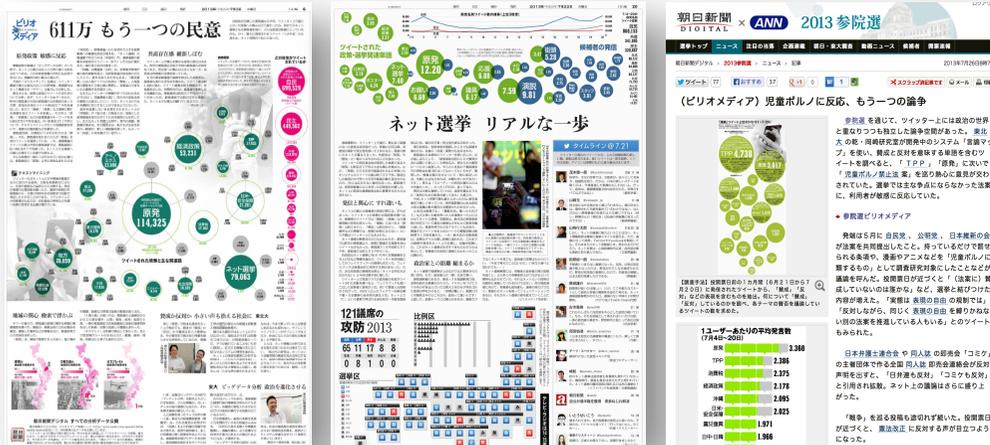
大規模コミュニケーション の支援へ

自然言語処理最前線



ニーズは社会のどこにもある

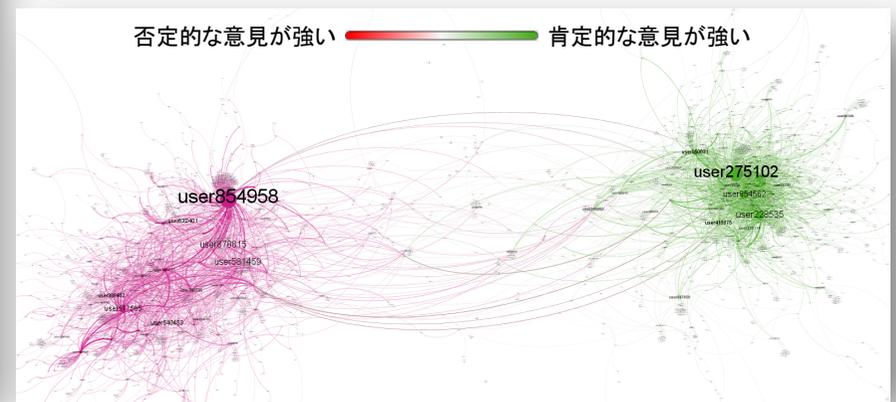
参院選の世論分析(朝日新聞)



誤情報の検出と検証の支援

トピック	スコア	命題
田尻智さん	18.7	海外で、ポケモンの生みの親の田尻智さんが亡くなった
コスモ石油千葉製油所	11.8	@市原市のコスモ石油千葉製油所LPGタンクの爆発により、千葉県、近隣圏に在住の方に有害物質が雨などと一緒に飛散する
女性暴行	10.4	@「阪神大震災の際には女性暴行が増えた」
有毒物質	8.6	コンビナート火災に関し『有毒物質が発生し、雨に混じって降ってくるので肌をさらさないように』
尾田栄一郎先生	5.3	尾田栄一郎先生が15億円を寄付
円資産	5.2	池田経済産業副大臣は、「日本企業が円資産を買い戻している
空中投下	3.0	『物資の空中投下が日本でも認められていない』

風評被害の分析(NHK)



2014 ワールドカップ

トップ ニュース 日本代表 出場国 ランキング 日程 写真・動画 特集 PDF号外 タイムライン

グループ▶ A B C D E F G H

スタジアムガイド Vamos! ブラジル入門 里奈の広報見聞録 ドン小西

ツイ フォロー

ツイート { g+1 m

メール

朝日新聞デジタル > 2014ワールドカップ > ワールドカップ 日本戦ツイート分析

メニュー ▾

日本対コートジボワール 6月15日(日) 10:00 キックオフ

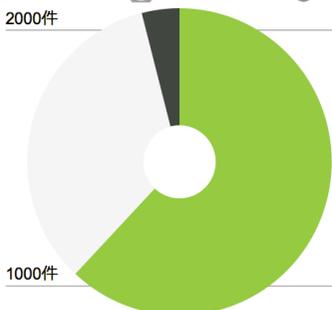
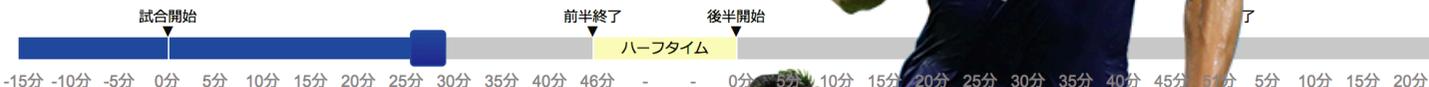
ツイッター上で選手名や監督名が含まれているツイートを抽出して、そのツイート件数を選手名や監督名を含むツイートが多いことを表しています。画像をクリックすると詳しい数字

自動再生 停止



日本 コートジボワール

前半23分



【日本代表全体のポジ・ネガの割合】

ネガ: 4% ポジ: 62% ニュートラル: 34%

0件



[PR]

※選手や監督の名前を含むツイートを収集、その中に多く出てくる言葉を分析・抽出しています。 ※ツイッターのデータは「NTTデータ」の協力で収集。東北大学大学院の乾・岡崎研究室の支援を受けて、ツイート件数、関連語、ポジ・ネガを分析しました。関連語では、意味のない文字など

震災直後に流れたデマ

 
【被爆予防方法】 イヅツを水に薄めて大量に飲んで下さい。特に子供に。自然ヨウ素で甲状腺を満たし、放射線ヨウ素の吸収を抑えます。 #jishin

3月13日  から

デマを否定する情報も



放射能対策でヨウ素剤の代わりにイソジン3滴という全くのデマが流れているので止めてください。イソジンだと14.3cc飲む必要が有りますが、こんなに飲んだら胃腸がただれます。 <http://bit.ly/> 

言論マップ



平時の一般のトピックでも

コラーゲンは肌に良い

質問：

同意 (459文) フェミダンプラスはコラーゲンとヒアルロン酸の働きで、
噛むだけでお肌つやつや！若返りと美肌を提供します log1 log2

対立 (76文) ただ、研究職の私から言わせてもらうと、コラーゲンは体
に入ったからといって、肌^に潤いをもたらすわけではあり
ません log1 log2

• コラーゲンは健康な肌ならきれいな網目構造になっており、まるでベッドのスプリング
のように表皮を支えて弾力のある肌を保っています log1 log2

• お肌ハリがなくなってきたり、たるんできているのかな、と思うときにコラーゲンやヒ
アルロン酸を摂取すると、お肌ハリがよくなるかな、と思う方もいると思います log1 log2

**フェミダンプラスはコラーゲンとヒアルロン酸の働きで、
噛むだけでお肌つやつや！若返りと美肌を提供します** log1 log2

• ただしコラーゲンは肌以外にも骨や血管・内臓にも使われるので不規則な生活で
体内がぼろぼろだと先に体内の改善に使われるので不規則な生活だといくら一生懸
命でも肌は綺麗にならない log1 log2

ニッピコラーゲンといったコラーゲンですが、肌か
ら浸透しないので、出来てしまったしわに対しての効
果は期待できません log1 log2

う事なので一緒にビタミンCやミネラルをとりましょう log1 log2

根拠 (10文) 美肌づくりに欠かせないコラーゲンですが、実はコラ
ーゲンは分子が大きいので、お肌に入りにくい！
という難点があります log1 log2

完了

同意・対立・根拠の認識

事実性(モダリティ)解析 + 構造的アライメント + 関係パターン認識

Collagen has beautiful skin effects.
コラーゲンには美肌効果がある

+ 存在
+ existence

- 存在
- existence

コラーゲンをサプリメントで飲んでも、
Even when taken as a supplement,

途中でアミノ酸に分解されてしまうので、
(collagen) ends up broken down into amino acids so,

(Φが)肌をきれいにする効果は期待できません。
we cannot expect (collagen) to be effective at making skin pretty.

事実性の推定

そのイベントは起こったのか起こっていないのか

流出しなかった（流出しなかった）

流出したではないか（流出した）

流出しただけでない（流出した）

流出を防いだ（流出しなかった）

流出を抑えた（流出した，でも少し）

流出を止めたい（流出した，今も流出）

同意・対立・根拠の認識

事実性(モダリティ)解析 + 構造的アライメント + 関係パターン認識

対立意見の根拠
evidence of
opposing opinion

Collagen has beautiful skin effects.

コラーゲンには美肌効果がある

美肌 = 肌を美しくする
beautiful skin = to beautify skin

肌を美しくする
to beautify skin

+ 存在
+ existence

美しい → きれい
beautiful → pretty

- 存在
- existence

対立
conflict

コラーゲンをサプリメントで飲んでも、
Even when taken as a supplement,

途中でアミノ酸に分解されてしまうので、
(collagen) ends up broken down into amino acids so,

根拠
evidence

(Φが) 肌をきれいにする効果は期待できません。
we cannot expect (collagen) to be effective at making skin pretty.

言論マップで利用している語彙知識

- 日本語WordNet (Bond+ 2009) : 56,741 synsets, 92,241 語
 - 過剰: 冗多, 過多, 冗長, 余った, 余計, 余分
 - 効果: 効目, 実効, 効, 利き目, 効き目, 利き, 効力, 効き, 機能, 効能
 - 病気 ⇔ 具合い, 丈夫, 健康, フィットネス, 健全, 多幸
- 事象間関係知識 (Matsuyoshi+ 2008) : 約140万対
 - <誰か>が<何か>を撮取する – <誰か>が<何か>を取る (上位語)
 - <何か>が普通だ – <何か>が異常だ (反義)
- Wikipediaから抽出した同義語: 113,401 synsets
 - HIDランプ – 高圧放電灯 – 高輝度放電ランプ
 - バイオ燃料 – エコ燃料 – バイオエネルギー
- Wikipediaから抽出した上位下位関係 (Sumida+ 2008) : 約350万対
 - イソフラボン – 女性ホルモン様成分
 - キシリトール – 甘味添加物
- 文脈類似度に基づく動詞間含意関係 (ALAGIN) : 約170万対
 - 引き起こす – もたらす, 予防する – 防げる, ケアする – 防止する

うがいはインフルエンザの予防に効果的だ

賛成意見の根拠

(33文)

反対意見の根拠

(6文)

あと、現在の豚インフルエンザの特徴は、日本でも冬に流行する、インフルエンザA型に似ていて、湿気・熱に弱いので、マスク・手洗い・うがいでかなりの予防はできるとのこと log1 log2

ウイルスは、粘膜に接着後、
わめて短時間で粘膜内に進入
効果は低いと考えられてい
防にはなりません log1 log2
耳より健康教室
でもほぼ落ち着き始めたので
B型の流行もまだ懸念される

入るということですから、
やうがいも有効でも、イン
ど効果はないともいわれて
漢方110腰痛・関節痛・

インフルエンザは、咳やく
ぶきによってヒトからヒトに
の着用、うがい・手洗いは
が、それにも増して、イン
接種が大切です log1 log2
健康長寿ネットーインフル

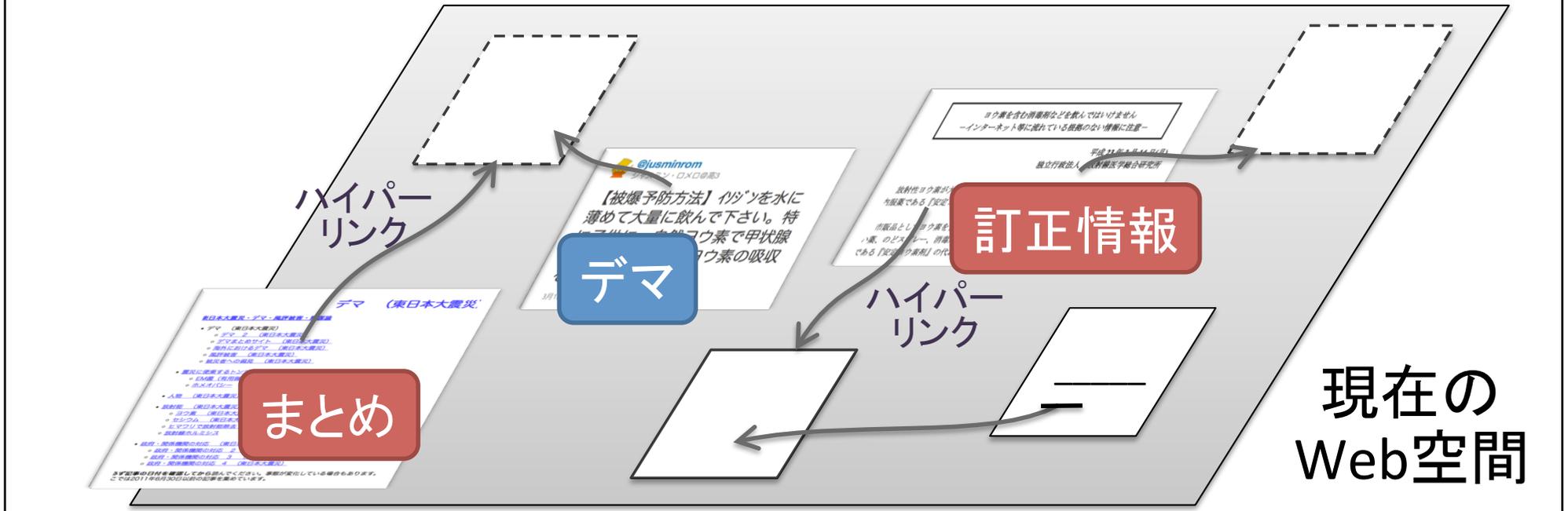
またインフルエンザは、イ
た人の咳やくしゃみに含まれ
ことによって感染するので、

よくうがい励行！と言われますが、インフルエンザウイルスは咽頭（のど）、眼、また鼻粘膜に付着すると約20分で細胞中に取り込まれてしまうので、外出後のうがいは予防には無効であり、外出時にマスクをすることをお勧めします log1 log2

インフルエンザウイルスは、粘膜に接着後、20～30分というきわめて短時間で粘膜内に進入しますので、うがいの効果は低いと考えられていますが、ある程度の予防にはなりません log1 log2

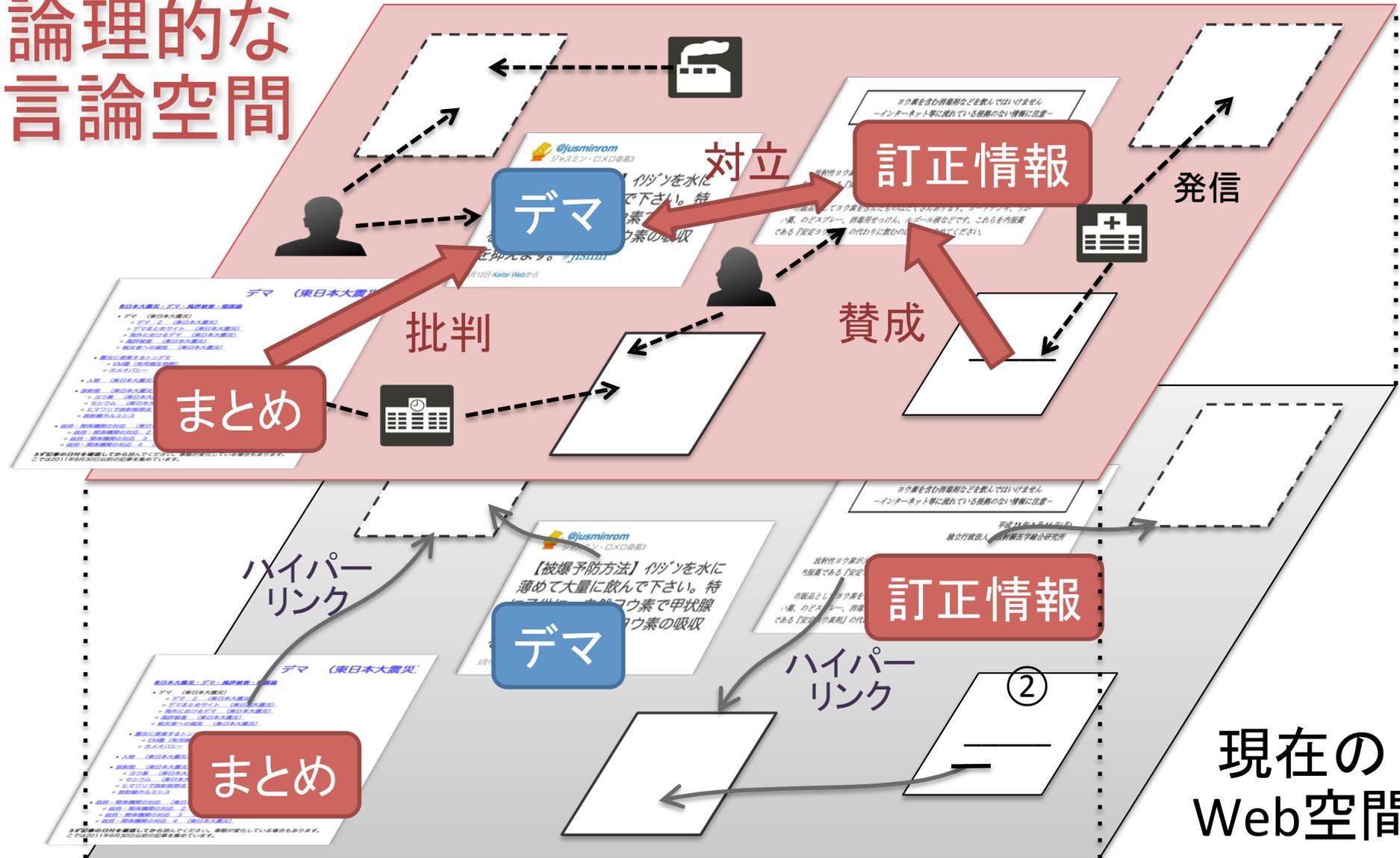
現在のウェブの問題

誤情報を見ている人に
訂正情報が伝わらない！



ウェブの情報の流れを変える

論理的な
言論空間



現在の
Web空間

「裏を取る」を支援

現状

対策

これは、おっ、



深海魚油
ソフトカプセル
Deep sea Fish Oil
Soft Capsules

〇〇社によって製造された深海魚油のソフトカプセルです。...

主な成分である ω -3脂肪酸は、血液のより良い循環を促進し、高コレステロールの改善、血栓の防止、高血圧の改善、心臓の健康維持をサポートします。また、双極性障害、ADHDの改善に有効です。

対立

根拠

あれ？ちがうのかあ

「 ω -3脂肪酸はADHDの改善に有効」の対立ページ

健康食品データベース

[独立行政法人 国立栄養研究所]

小児のADHDは血中DHAレベルの低さと関連があるが、DHAは小児の注意動性障害(ADHD)に対して、経口摂取で効果がないことが示唆されている。DHAを1日あたり345mg摂取しても改善はみられなかった...

魚油のオメガ3脂肪酸、頭を良くする効果ない=英研究

[ロイター] 英国で行われた研究によると、魚油に含まれるオメガ3脂肪酸は、心臓の健康のためには良いが、脳の機能を高める効果はないという。ロンドン大学の研究者らが専門誌「アメリカン・ジャーナル・オブ・クリニカル...

「 ω -3脂肪酸はADHDの改善に有効」の根拠ページ

生活習慣をサポートする『オメガ3脂肪酸 (後編)』

[サブスマートUSA] ... ADHD、統合失調症、アルツハイマー疾患のような精神・認知に関する疾患を患う人は、血中のオメガ3脂肪酸レベルが低い人が多いことから、このような精神・認知に関する疾患に対し、オメガ3脂肪酸は有効性が期待できます。...

その先の言語理解研究へ

「行間を読む」言語処理への挑戦



「行間を読む」

週末に母とやなぎに行ってみました。

うどん屋の名前

うどんを食べるために

私は評判のきつねうどんを頼んだの

私はうどんを食べた

やなぎで

ですが、期待通りの味に大満足。

うどんが美味しいと

きつねうどんの

「行間を読む」 常識に基づく推論

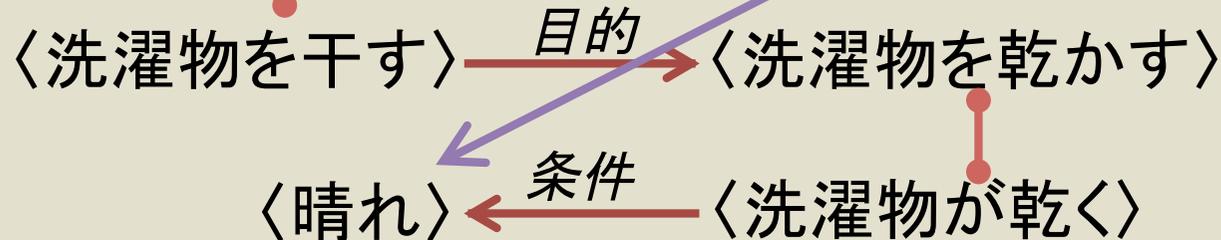


庭に洗濯物を干したとたんに雨が降ってきた

庭に洗濯物を干す

雨が降る

ガッカリ



大問題！

膨大な知識をどうやって集めるか

言葉の理解に必要な知識

- 「洗濯物を干す」目的は「洗濯物を乾かす」こと
- 「洗濯物が乾く」条件は「晴れている」こと
- 「うどん」は、「レストラン(うどん屋)で食べる」
- 「(料理を)注文する」ことを「頼む」とも言う
- 「レストラン」では、「料理を注文」して、しばらくすると料理が「出て」きて、それを「食べる」

常識もビッグデータから集める

天気が良いので、洗濯物がよく乾く
最近、晴れが続いているので、洗濯物がよく乾く
日が当たらないので、洗濯物が乾くのもままならず
お天気は晴れマークが多かったから、洗濯物が乾く
湿度はなかったから、洗濯物が乾く
今日は、空気が乾燥していて、洗濯物がよく乾く
梅雨空が続いたため、洗濯ものが乾かず

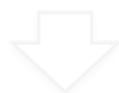


天気が良い(晴れる) → 洗濯物がよく乾く

ビッグデータから自動的に集める

天気が良いので、洗濯物がよく乾く
最近、晴れが続いているので、洗濯物がよく乾く
お天気が晴れ多く多かったから、洗濯物がよく乾く
湿度はなかったから、洗濯物がよく乾く
今日は、空気が乾燥しているから、洗濯物がよく乾く
梅雨空が続いたため、洗濯ものが乾かず

**言語処理から人工知能に
ブレークスルーをもたらす**



天気が良い(晴れる) → 洗濯物がよく乾く

談話処理(「行間を読む」解析)

多様な問題の複合

週末に母とやなぎに行ってみました。

固有表現認識

うどん屋の名前

談話関係解析

うどんを食べるために

私は評判のきつねうどんを頼んだの

??????

私はうどんを食べた

多義性解消

注文した

ですが、期待通りの味に大満足。

??????

うどんが美味しいと

照応解析

きつねうどんの

談話処理の難しさ

- 複合的問題
 - さまざまな部分問題が相互に依存
 - 語義、省略・照応、談話関係、意図、...
 - どの問題から解くかを予め指定するのが困難
- 知識の組み合わせ
 - かなり大規模な知識ベースが利用可能に
 - いつどの知識を組み合わせるべきかを予め指定するのが困難

大規模・多種類の知識を使いこなす
ための仕組みを研究する段階に

Ed shouted at Tim because he was angry.

X is angry → X shouts

Ed shouted at Tim because he crashed the car.

X crashes Y's car → Y shouts at X ?

仮説推論 (abduction)

観測に対する最良の仮説(説明)を求める推論

仮説



スプリンクラー消し忘れ?



夜のうちに雨が降った?

観測



自宅の庭が濡れている



隣りの庭も濡れている

「行間を読む」仮説推論

背景知識

「何かを食べる」には「飲食店に行って、それを注文」すればよい

$\text{食べる}(x, z) \rightarrow \text{飲食店}(w) \wedge \text{行く}(x, w) \wedge \text{注文する}(x, z, w)$

「(店で)注文する」ことを「頼む」とも言う

$\text{注文する}(x, z, w) \rightarrow \text{頼む}(x, z, w)$

「ある食べ物が評判」なら「それを食べたい」と思う

$\text{評判}(z) \rightarrow \text{食べる}(x, z)$

評判(z)

私はきつねうどんを食べた

食べる(x, z)

評判だから食べに行った

仮説

飲食店(w)

食べる目的
で行った

注文する(x, z, w)

やなぎ = 飲食店

y=w

「頼む」=
やなぎで注文する

w=v

観測

私(x) \wedge やなぎ(y) \wedge 行く(x, y) \wedge きつねうどん(z) \wedge 頼む(x, z, v) \wedge 評判(z)

母とやなぎに行った。 評判のきつねうどんを頼んだ。

乾・岡崎研の取り組み

背景知識

「何かを食べる」には「飲食店に行って、それを

① 仮説推論の高速化

7分 (MLN@ISI) → 3秒 (東北大)

COLING2012, PAIR2011

情報処理学会 山下記念賞 (2012)

言語処理学会 年次大会優秀賞 (2012)

③ 言語処理への応用

NLPAR2013

NLP若手の会 奨励賞 (2011)

言語処理学会 年次大会最優秀賞 (2014)



井之上
(研究員)

はきつねうどんを食べた

食べる(x, z)

評判(z)

評判だから食べに行った

山本(D2)



杉浦(D2)

仮説

飲食店(w)

食べる目的
で

注文する(y, w)

やなぎ = 飲食店

y=w

観測

私(x) ∧ やなぎ(y) ∧ 行く(z)

母とやなぎに行った

② 仮説の評価関数の学習

CICLing2013

NLP若手の会 奨励賞 (2012)

自然言語処理研究会 学生奨励賞 (2012)

IBIS 2012 Honorable Mention (2012)



判(z)

。

整数線形計画法による推論の高速化

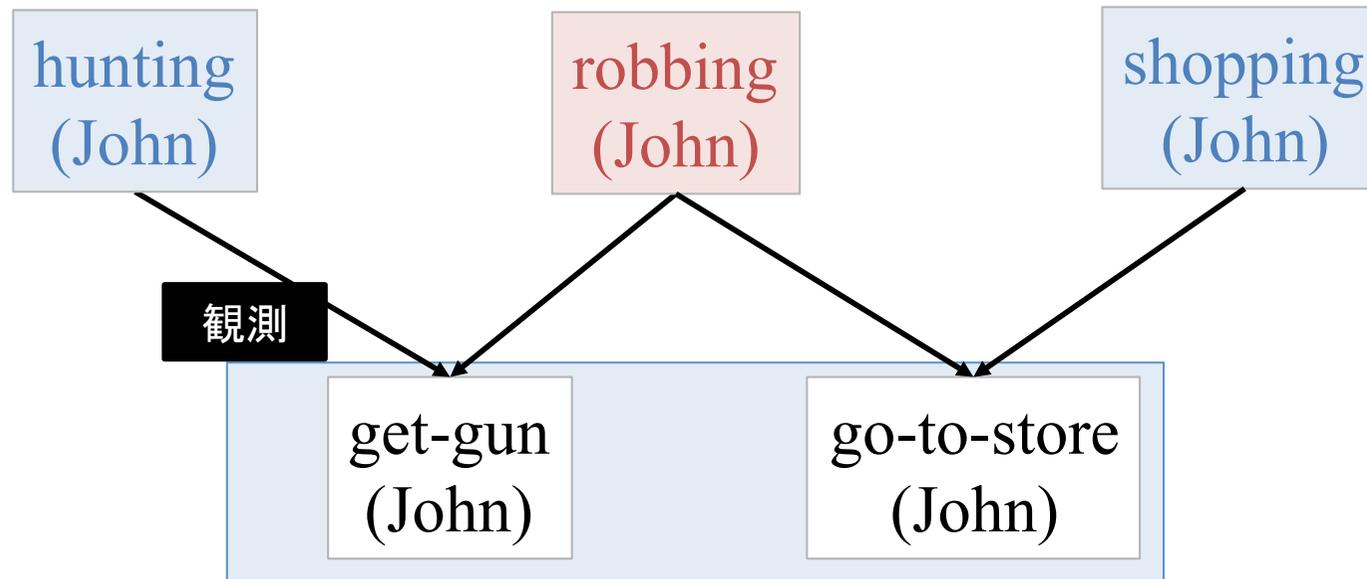
背景知識

hunting(x) \Rightarrow get-gun(x)

robbing(x) \Rightarrow get-gun(x)

shopping(x) \Rightarrow go-to-store(x)

robbing(x) \Rightarrow go-to-store(x)



John got a gun and went into the store.

整数線形計画法による推論の高速化

背景知識

$$S(x) \Rightarrow P(x)$$

$$S(x) \Rightarrow Q(x)$$

$$h_{x=y}$$

$$u_{S(x),S(y)=1} \Rightarrow h_{x=y}=1$$

$$u_{S(x),S(y)} \leq h_{x=y}$$

目的関数

$$\arg \max_{H} \text{score}(H) = \mathbf{w}^T \cdot [h_{P(x)}, h_{Q(y)}, \dots, u_{S(x),S(y)}]$$

$$= \mathbf{w}^T \cdot \Phi(h_{P(x)}, h_{Q(y)}, \dots, u_{S(x),S(y)})$$

$$h_{S(x)=1} \Rightarrow h_{S(x)}$$

$$h_{S(x)} \leq h_{S(x)=1}$$

$$h_{P(x)=1}$$

観測

$$u_{S(x),S(y)} \in \{0, 1\}$$

$P(x)$

$Q(y)$

$$h_{P(x)}$$

$$h_{Q(y)} \in \{0, 1\}$$

$$u_{S(x),S(y)=1} \Rightarrow h_{S(x)}=h_{S(y)}=1$$

$$u_{S(x),S(y)} \leq (h_{S(x)} + h_{S(y)})/2$$

推論効率の評価実験

- データセット
 - 背景知識
 - WordNet 3.0 [Fellbaum 98]
 - FrameNet 1.5 [Ruppenhofer+ 10]
→ 約37万の推論規則
 - 観測 (入力)
 - RTE-2 開発セット
 - RTE: 2つのテキストの対が含意関係にあるかを判別
 - 1,600 文章, 平均 20.3 リテラル
- ツール
 - 論理式への変換: Boxer [Bos 08]
 - ILP ソルバー: Gurobi Optimizer 4.6

評価実験の結果

後ろ向き推論の深さを変えたときの推論時間
(約37万の推論規則、平均20リテラルの入力)

深さ	潜在仮説集合の平均サイズ	2分以内に最適解と判断された問題	平均推論時間
1	89	99.8%	0.33 秒
2	479	98.4%	1.07 秒
3	1,171	95.8%	2.63 秒

比較参考値

- Mini-TACITUS [Mulkar-Mehta+ 07]: **30分以上**
- Markov Logic Networks [Richardson & Domingos 06] による実装 [Blythe+ 11]: **7分**

評価関数の学習

学習の目標

$$\mathbf{w}^T \cdot \Phi(H') + \Delta < \mathbf{w}^T \cdot \Phi(H)$$

出力された仮説 (H')

$e = h$ (**Ed** = **he**) ... 誤

$crash(e, u_2)$

$irritated(e)$

正解の仮説 (H)

$t = h$ (**Tim** = **he**) ... 正

$angry-with(e, t)$

=

$angry-with(u_3, h)$

教師が与えなくてもよい

$do-bad-thing(u_1, h)$

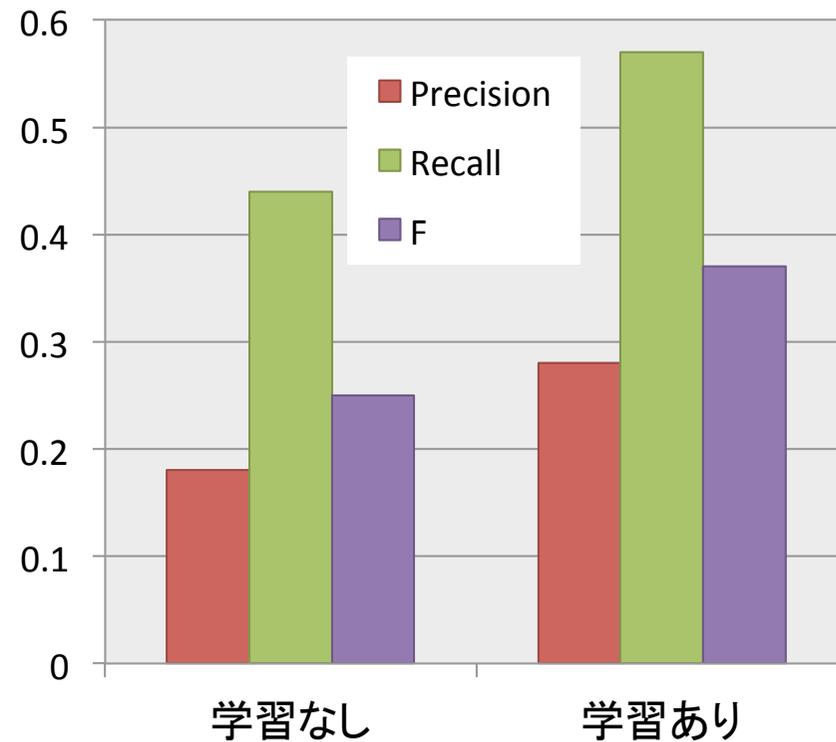
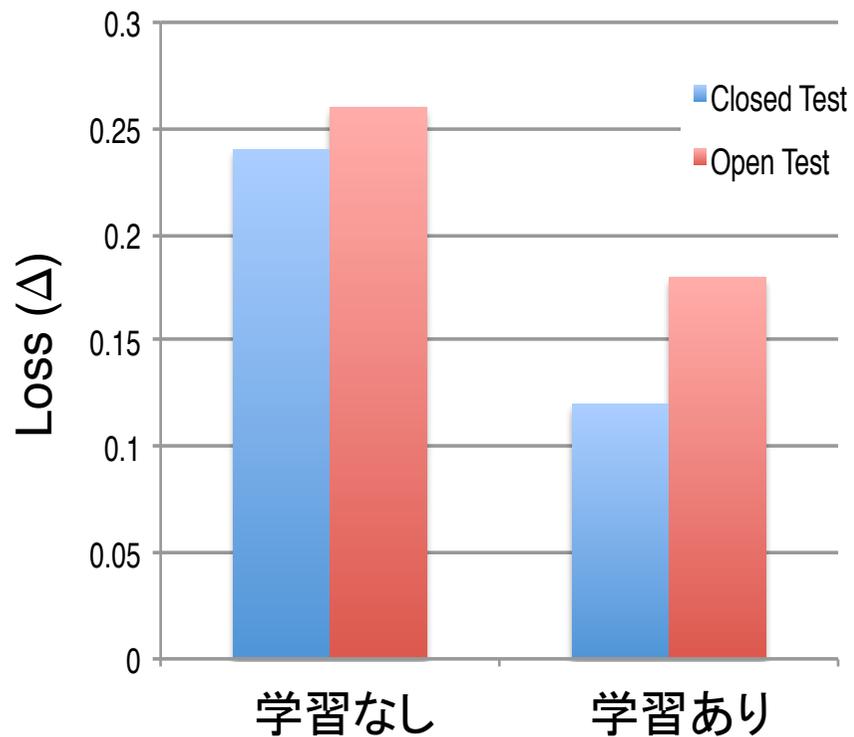
観測

$Ed(e) \wedge shout-at(e, t) \wedge Tim(t) \wedge male(h) \wedge crash(h, c) \wedge car(c)$

Ed shouted at **Tim** because **he** crashed the car.

学習の効果

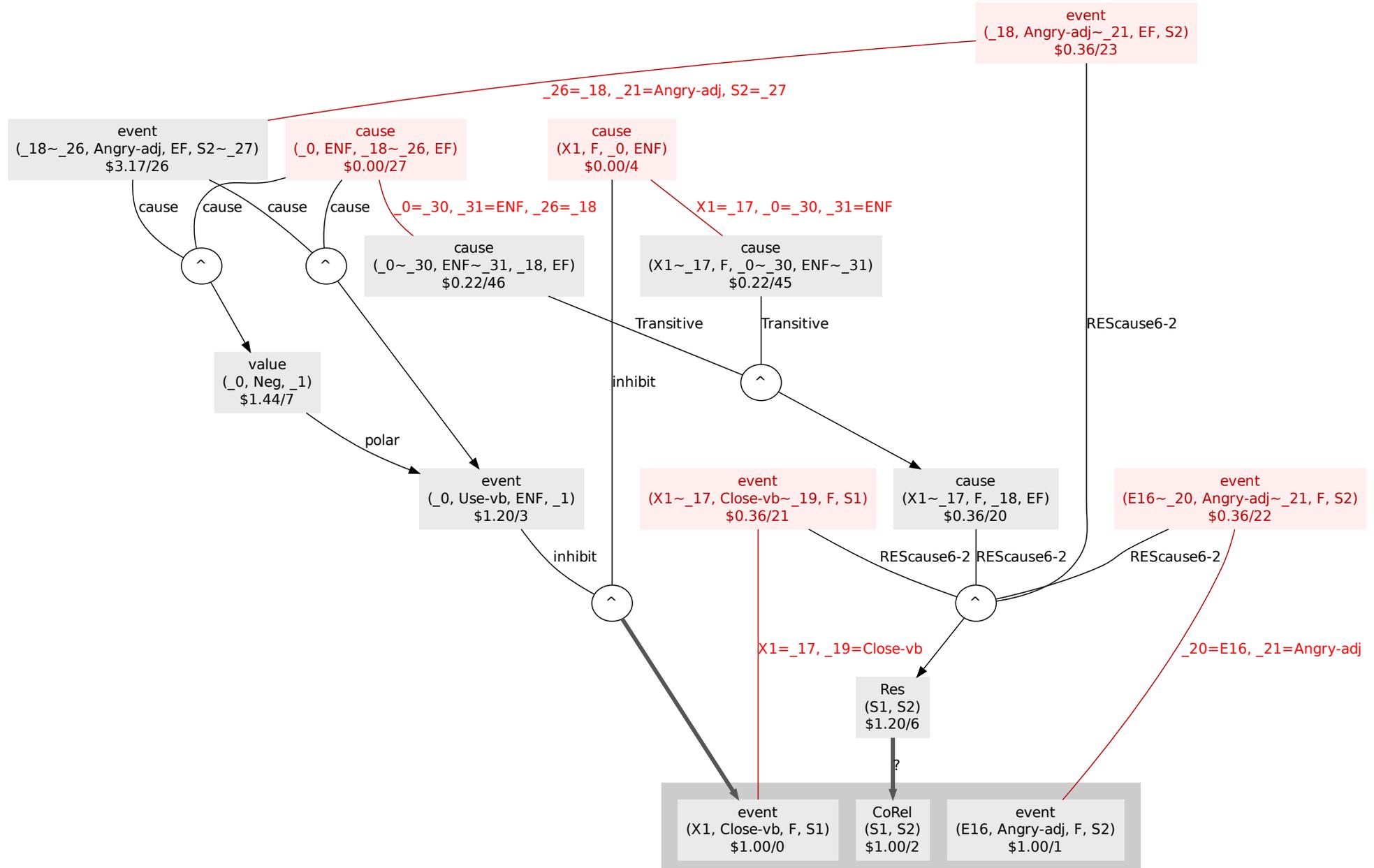
プラン認識課題 [Ng & Mooney 92]



(S1: A lot of traffic once used Folsom Dam Road.)

S2: Right now, the road is closed.

S3: Most of the people who used the road every day are angry.



言葉がわかるコンピュータは どこまでできたか

大規模コミュニケーションの支援が必要

カギは自然言語処理による情報編集

言葉がわかる^{コンピュータ}機械の実現は簡単でない

最大の技術課題は知識のボトルネック

ビッグデータからの知識獲得で技術革新

YouTube「東北大学サイエンスカフェ」

言葉がわかるコンピュータは どこまでできたか (2013年2月)

～言葉の不思議と自然言語処理の最前線～



乾・岡崎研究室「いつでも見学会」にもぜひお出てください。詳しくは研究室のHPを。