

文部科学省プロジェクト 「多メディアWEB解析基盤の構築及び社会分析ソフトウェアの開発」での経験と 今後の情報分野の研究推進に向けて

国立情報学研究所

佐藤 真一

プロジェクトの概要

- 文部科学省 次世代IT基盤技術のための研究開発「Web社会分析基盤ソフトウェアの研究開発」
- 採択課題名:「多メディアWeb解析基盤の構築及び社会分析ソフトウェアの開発」
- 研究期間:平成21年～平成24年
- 参加機関:国立情報学研究所、東京大学、早稲田大学
- 予算:4.6億円

全体概要

目的

社会学、言語学、リスク管理、マーケティング等多様な社会分析ニーズに応じるために、膨大な多メディアWeb情報を収集、蓄積し多様な解析を可能とする多メディアWeb情報解析基盤の構築と社会分析ソフトウェアの研究開発並びに実証を行う

背景

Web情報は人類社会の観測・調査・解析において新価値創出のために必要不可欠な情報源

- 多メディア化が急速に進むと同時に、実世界情報と相互に及ぼし合う影響も拡大
- 放送映像との密接な相互作用

課題

多メディアWeb情報の**収集・蓄積**、多メディア**内容解析**、高並列計算環境上での大容量・高スループット**解析基盤**、有効な**社会分析ソフトウェア**の実現が必須



イラン抗議デモ・チュニジア政変における
twitter / facebook / YouTubeの役割

twitter

jkrums

<http://twitpic.com/135xa> - There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy.



「ハドソン川の奇跡」は
twitterの投稿写真で話題に

単一メディアではない
複数メディアの
有機的な統合による解析は
世界初

研究開発の概要

多メディアWeb解析基盤の構築
及び社会分析ソフトウェアの開発



Web情報



放送映像

多メディアWeb基盤技術
多メディアWebアーカイブ データメンテナンス スケジューリング技術

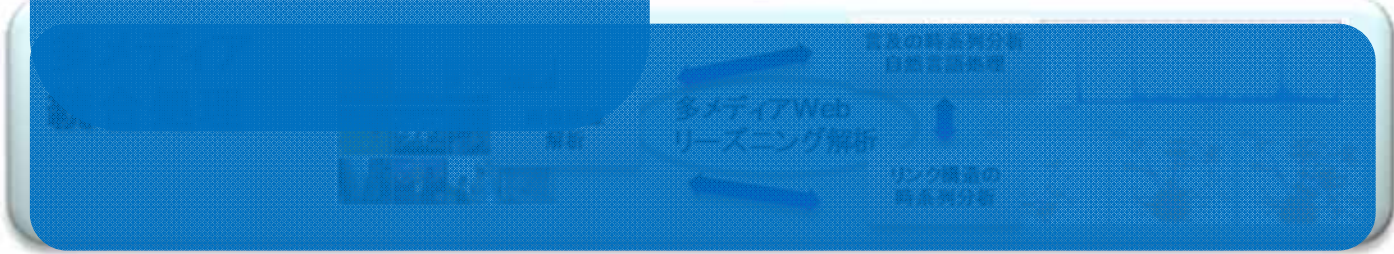
データ処理基盤
アーカイビング
クラウド

データ配信の最適化
索引更新処理

多メディアWeb解析要素技術

メディア解析技術
画像・映像解析
自然言語処理

- 画像解析
- ワーキングブア問題
- 画像解析
- リンク
- 多次元
- 高速
- 多メディアWeb
- 分析可視化技術



多メディアWeb解析の実証評価

ワーキングブア問題

社会起業の広がり

応用

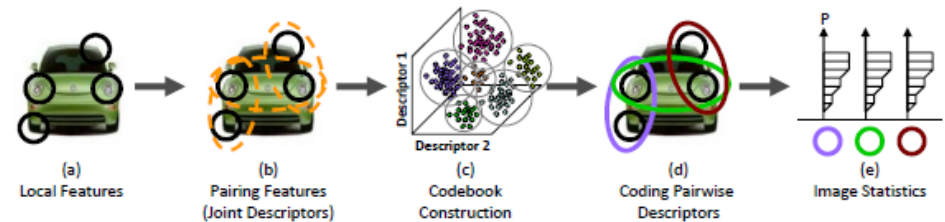
画像・映像キーワード抽出技術 (画像・映像意味分類技術)

多メディアWeb解析基盤の構築
及び社会分析ソフトウェアの開発

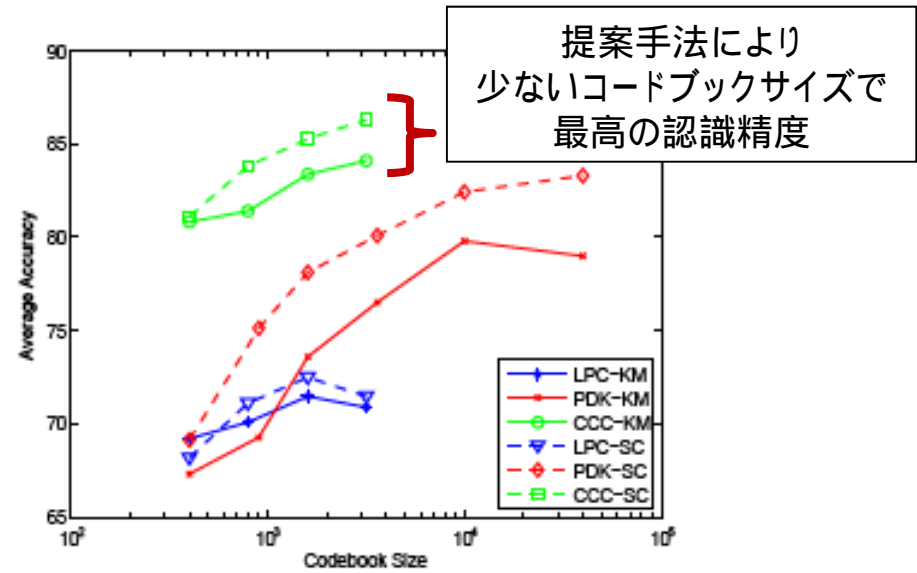
- 与えられたショットに対し、写っている物体種別、情景種別、画像・映像の種別などに基づいて、自動的に概念レベルの意味分類を行う技術(車、建物、スポーツなど)
- 正解データ付きの学習用映像(数百時間規模)で意味分類器を学習
- インターネット映像でも高性能となることを確認済み



キーワード付与結果例



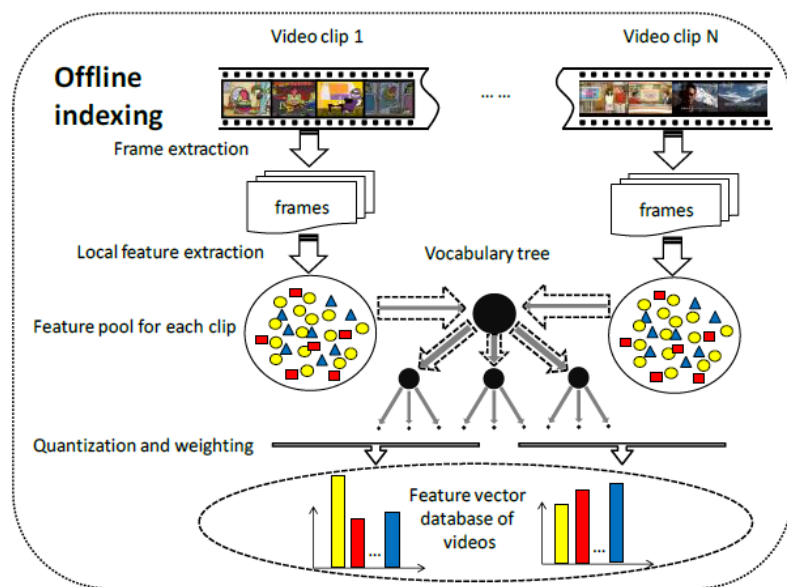
局所特徴量のペアを使うことにより
飛躍的に認識精度を向上



コンピュータビジョントップ国際会議ECCV, ICCV
機械学習トップ国際会議NIPS

画像・映像リンケージ技術 (同一物体検索)

- 画像間で共起する物体を検出する技術
- 物体による画像・映像検索
- Webと放送映像とで共起する物体の検出



TRECVID2011インスタンスサーチタスク
検索精度**世界第一位**

多メディアWeb解析基盤の構築
及び社会分析ソフトウェアの開発

問い合わせ
画像



任意の画像で
コマーシャル映像
数万本を検索

検索結果



多メディアWeb分析・可視化技術 (高速固有表現抽出アルゴリズム)

多メディアWeb解析基盤の構築
及び社会分析ソフトウェアの開発

- 大規模ウェブデータを社会分析に利用可能とする
人物名や製品名などの固有表現の超高速抽出技術

人物名

健康上の理由により療養休暇中の**アップル**のCEO,
スティーブ・ジョブズ氏は次期**iPad**や**iPhone**の開発に
関与していると明かしました

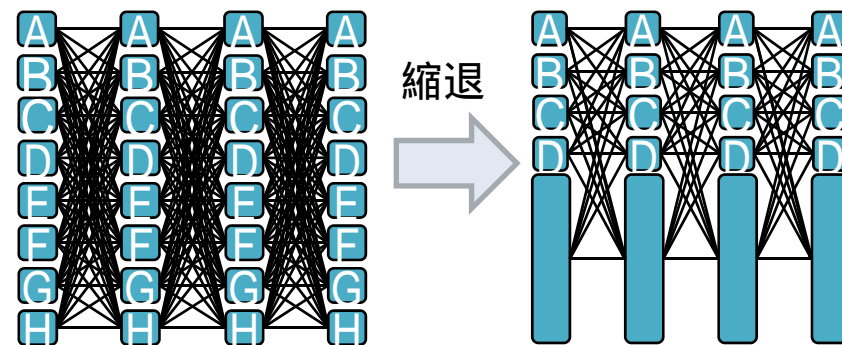
製品名

会社名

- メモリ効率を保持しながら従来より最大**300倍の高速化**を実現
 - 縮退ラティスによる探索空間の削減法を提案

解析速度(文数/秒)

アルゴリズム	結合タグ	Supertag
Viterbi	77	1.1
CarpeDiem	51	0.26
提案手法	1600	300



自然言語処理における**最高峰の国際会議ACL2010**に採択

Kaji et al. Efficient Staggered Decoding for Sequence Labeling

多メディアWeb分析・可視化技術 (組合せ素性に基づく分類器の学習)

多メディアWeb解析基盤の構築
及び社会分析ソフトウェアの開発

- 係り受け解析の頑健化には**多数の素性の組合せ**を考慮し、**大量の訓練例**を用いた学習が必要

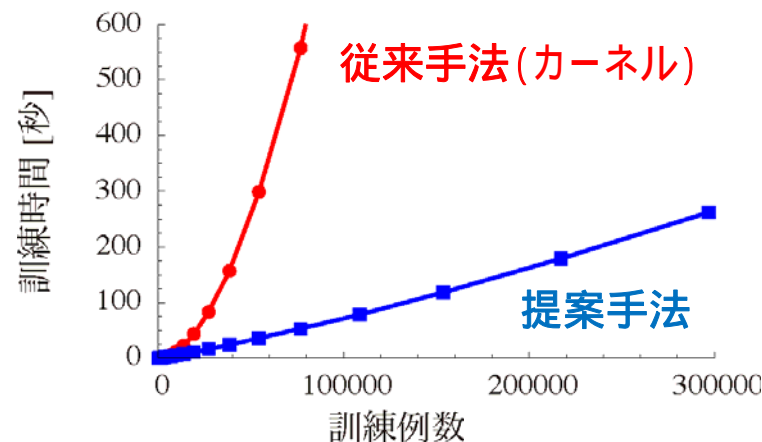
例) 係り受け解析



基本素性: 品詞(細分類), 活用, 距離
組合せ素性: 品詞×活用, 品詞×品詞細分類,
品詞細分類×活用, 品詞×距離

- 組合せ素性を用いた大規模学習のための逐次学習法を提案

- 低頻度素性に関する組合せを多項式カーネルで効率的に計算
学習の時空間効率を制御
- 素性の組合せを頻度を考慮して再分割し部分計算結果を再利用
学習の規模耐性を向上



空間効率を保ちつつ学習を最大250倍高速化

トップ会議 COLING 2010 採択 高速学習器 opal をオープンソースで公開

多メディアWeb統合処理 (QueueLinker)

多メディアWeb解析基盤の構築
及び社会分析ソフトウェアの開発

- リアルタイム分散解析の容易な実現を可能とする「リアルタイム分散解析ミドルウェア」

応用例: 世界最高速・並列分散Webクローラ

特徴

- モジュール実装と接続関係をProducer - Consumerモデルにより記述するのみ (データストリーム処理の容易な実現)
- 並列数・分散数を自由にユーザ制御可
- モジュール内の並列制御はQueueLinker側で制御

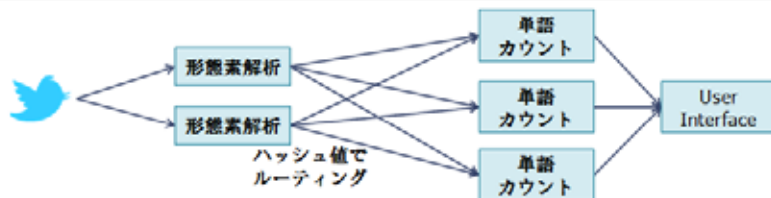
モジュール接続の記述例と実行方法

```

LogicalGraph graph = new LogicalGraph();
LogicalVertex twitter = graph.addLogicalVertex(TwitterDataSource.class);
LogicalVertex morphAnalyzer = graph.addLogicalVertex(MorphAnalyzer.class, 2);
LogicalVertex wordCount = graph.addLogicalVertex(WordCount.class, 3, PMode.Hash);
LogicalVertex ui = graph.addLogicalVertex(UI.class);

graph.addLogicalEdge(twitter, morphAnalyzer);
graph.addLogicalEdge(morphAnalyzer, wordCount);
graph.addLogicalEdge(wordCount, ui);

QueueLinkerClient client = QueueLinkerClientFactory.getClient();
QueueLinkerJob job = new QueueLinkerJob(graph);
JobHandle handle = client.startJob(job);
    
```

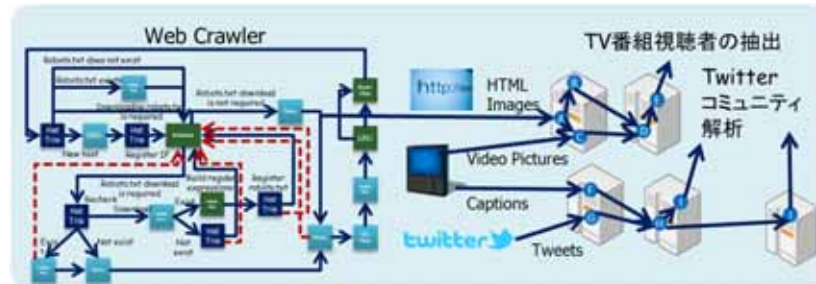


- モジュールのデータ並列方式を指定
- Full: 指定なし (ラウンドロビン)
 - Hash: ハッシュ分割
 - Customize: ユーザ制御

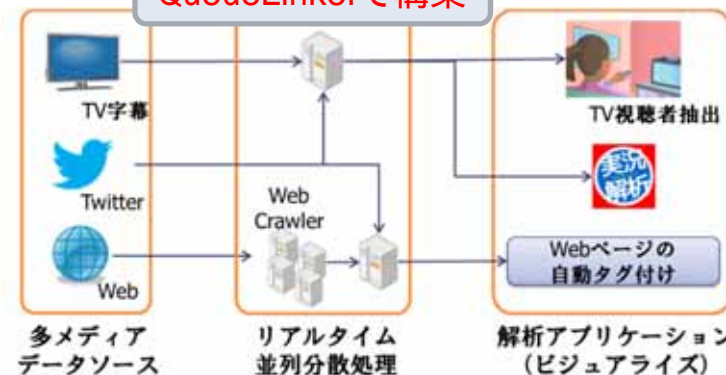
「キュー」を「接続」
⇒ QueueLinker

QueueLinker提供の基底クラス

名称	用途	ブロック処理の可否
PushModule	フィルタリング・算術演算を含む、ノンブロッキング処理一般	×
PullModule	ファイル I/O を含む、ブロッキング処理一般	○
SourceModule	外部データの取得	○
SinkModule	処理データの保存・表示	×



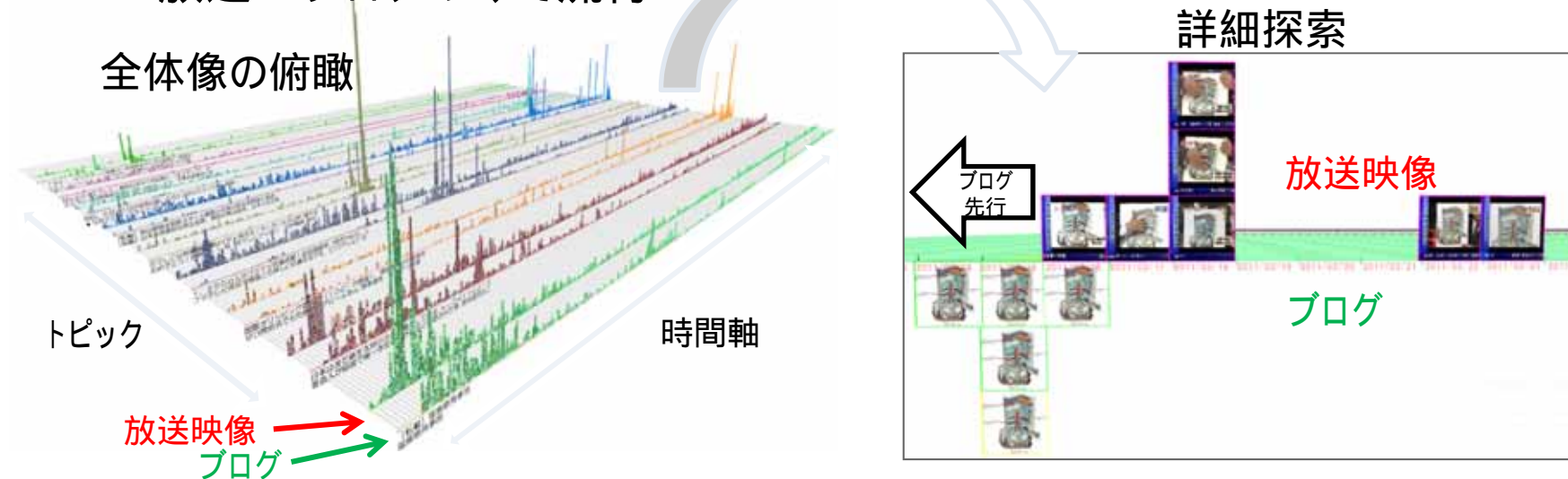
QueueLinkerで構築



オープンソース公開

放送・ブログを用いた 相補的なイベント抽出・反応差分分析

- 放送・ブログから類似画像クラスタ群を抽出、3次元空間に時系列可視化
- メディア間の相違、話題間の相違を探索
 - 放送先行で流行
 - ブログ先行で流行
 - 同時に流行
 - 相互に影響
 - 放送 / ブログのみで流行

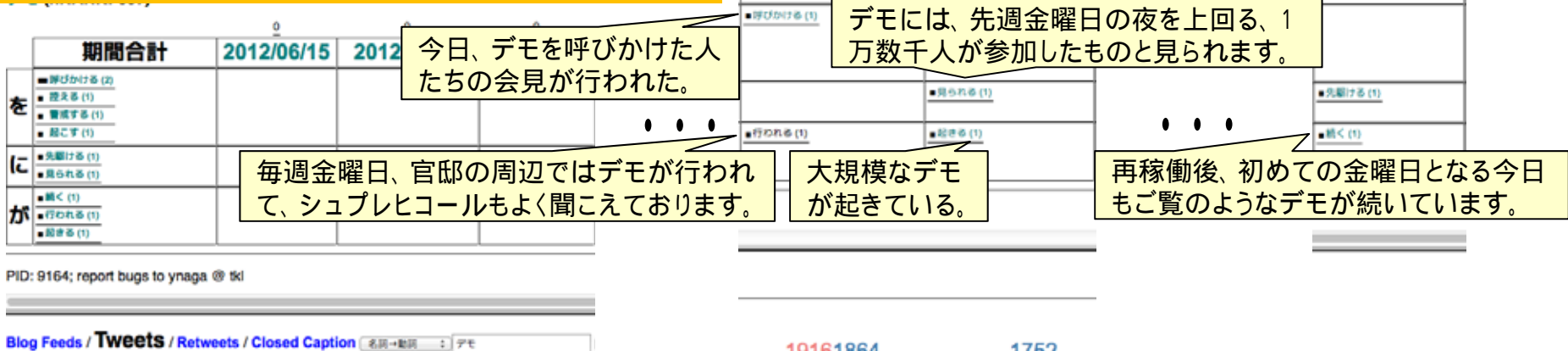


イベント・感情追跡システム

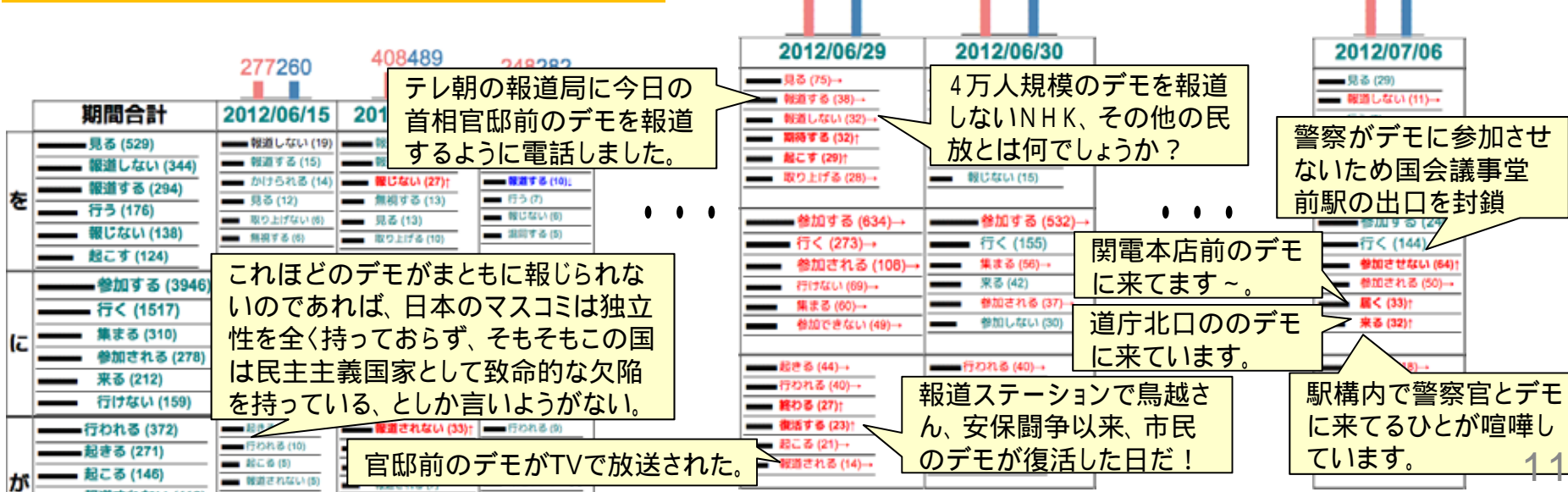
～ 頑健・高速な依存構造解析器J.DepPによる
(ソーシャル)メディア解析～

多メディアWeb解析基盤の構築
及び社会分析ソフトウェアの開発

ニュース字幕におけるデモに関する言及



ツイッターにおけるデモに関する言及



TV・Twitter解析

- TVに連動するtweetの自動抽出, 及びTV番組の盛り上がり部分の自動検出を実現

TV・Twitter解析



DEIM発表・サービス公開 (2013.2)

多メディアWeb解析基盤の構築 及び社会分析ソフトウェアの開発

TV連動tweet自動抽出

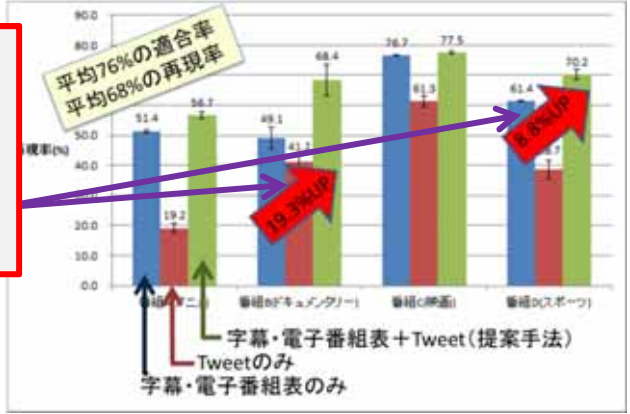
- TV番組視聴ユーザのtweetの特徴語 + 字幕 + 電子番組表利用

TV番組視聴ユーザの内、ハッシュタグを利用するユーザは一部(13.8%)



- 1) 番組公式の特徴語 (字幕テキスト、電子番組表)
- 2) ハッシュタグ付でtweetされたtweetそのもののテキスト (クエリ拡張的に利用)
- 3) 1)と2)の併合

ドキュメンタリー、スポーツ系番組での再現率向上顕著 (8-19%)

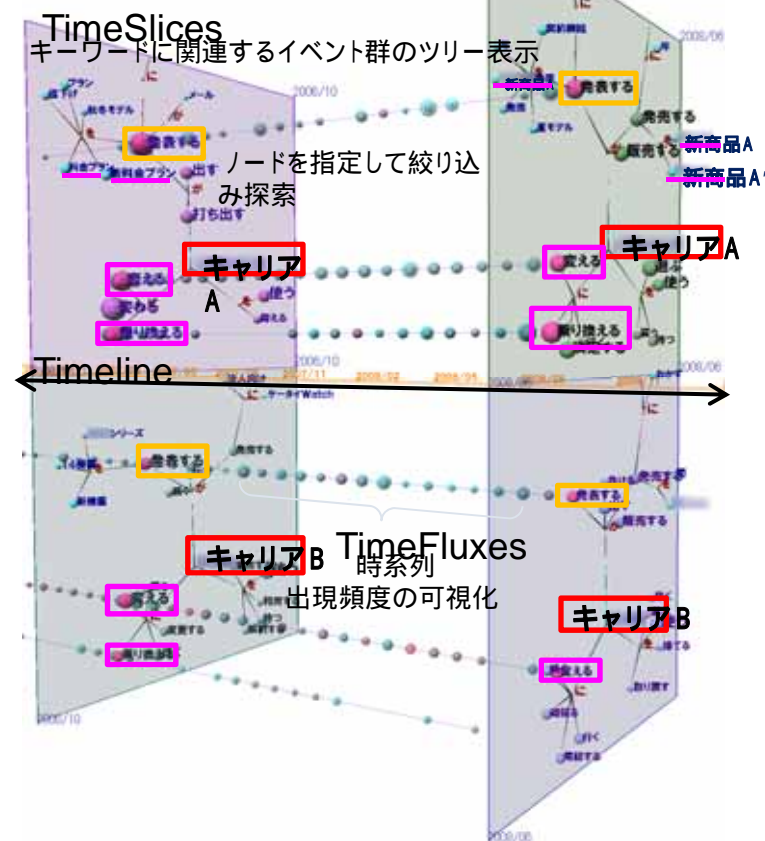


多メディア話題追跡システム

- Webグラフ及び係り受け関係の時系列変化を可視化し、インフルエンサー、人々の行動・興味の推移を追跡探索
- **メディア間、話題間の比較分析が可能**
 - メディアによるインフルエンサー、書き込み内容の差
 - 商品間、人物間の差



係り受け解析を用いた 話題追跡システム



国際会議IV2010採択、第72回情報処理学会全国大会 **大会優秀賞**
国際会議PVis2012採択、DEIM2011 **優秀論文賞**

学術的貢献

- 伊藤 正彦、豊田 正史、喜連川 優: “時系列情報をもつウェブグラフ発達過程の3次元可視化,” 情報処理学会第72回全国大会, 2010. (大会優秀賞受賞)
- 吉永直樹, 喜連川優: “組み合わせ素性に基づく分類器の効率的学習法,” 情報処理学会第73回全国大会, 2010. (大会優秀賞受賞)
- Nobuyuki Morioka and Shin'ichi Satoh, “Building Compact Local Pairwise Codebook with Joint Feature Space Clustering,” Proc. of European Conference on Computer Vision (ECCV2010), 2010.
- Zhenglu Yang, Jianjun Yu, Masaru Kitsuregawa: “Fast Algorithms for Top-k Approximate String Matching,” Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2010), 2010.
- Nobuhiro Kaji, Yasuhiro Fujiwara, Naoki Yoshinaga and Masaru Kitsuregawa: “Efficient Staggered Decoding for Sequence Labeling,” Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), 2010.
- Naoki Yoshinaga and Masaru Kitsuregawa: “Kernel Slicing: Scalable Online Training with Conjunctive Features,” The 23rd International Conference on Computational Linguistics (COLING 2010), 2010.
- 第3回データ工学と情報マネジメントに関するフォーラム (DEIM2011) 優秀論文賞, 伊藤正彦, 吉永直樹, 豊田正史, 喜連川優, 2011年5月
- Nobuyuki Morioka, Shin'ichi Satoh, “Compact Correlation Coding for Visual Object Categorization,” International Conference on Computer Vision (ICCV2011), Nov., 2011.
- Nobuyuki Morioka and Shin'ichi Satoh, “Generalized Lasso based Approximation of Sparse Coding for Visual Recognition.” Proc. of Neural Information Processing Systems (NIPS), 2011.
- Nobuhiro Kaji, Masaru Kitsuregawa, “Splitting Noun Compounds via Monolingual and Bilingual Paraphrasing: A Study on Japanese Katakana Words,” Conference on Empirical Methods in Natural Language Processing (EMNLP2011), pp. 959-969, 2011.
- 情報処理学会コンピュータサイエンス領域奨励賞, 上田高德 (2012年11月)
- 第4回データ工学と情報マネジメントに関するフォーラムDEIM2012 (第4回データ工学と情報マネジメントに関するフォーラム) 最優秀論文賞, 上田高德, 秋岡明香, 山名早人 (2012年3月)
- 第4回データ工学と情報マネジメントに関するフォーラムDEIM2012 学生奨励賞, 上田高德 (2012年3月)
- Xiaomeng Wu and Shin'ichi Satoh, “Ultra-high-Speed TV Commercial Detection, Extraction, and Matching,” IEEE Trans. on Circuits and Systems for Video Technology, Mar. 2013.
- Duy-Dinh Le and Shin'ichi Satoh, “Auto Face Re-Ranking By Mining the Web and Video Archives,” Proc. of Computer Vision and Pattern Recognition (CVPR), 2012.
- Yohei Takaku, Nobuhiro Kaji, Naoki Yoshinaga, and Masashi Toyoda, “Identifying Constant and Unique Relations by using Time-Series Text,” Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing (EMNLP 2012), pp. 883-892, 2012.
- 言語処理学会2012年最優秀論文賞、鍛冶伸裕、喜連川優、2013年3月
- 上田高德, 秋岡明香, 山名早人: “マルチコアCPU環境における低レイテンシデータストリーム処理”, 電子情報通信学会論文誌D, Vol.96-D, No.5 (2013.5), 論文賞
- 他、招待講演等多数

社会的貢献

- 日本語係り受け解析、高速固有表現抽出、画像意味分類、高レイテンシスケジューラなどをオープンソース提供
- 防災科学技術研究所、東大メディア研究者らと連携し、震災時の報道の解析を実施
- Twitterのトレンド解析「Realtime trend analytics」が2015年9月より角川アスキー総研より提供開始
- ビデオリサーチ社との共同研究による視聴率との統合解析

後継プロジェクト

- 科学研究費補助金 基盤研究(A)「大規模映像アーカイブにおける事物マイニングによる社会センシング基盤技術」(H26-28)(研究代表:佐藤真一) 4,238万円
- 創造的研究推進事業CREST「ビッグデータ統合利用のためのセキュアなコンテンツ共有・流通基盤の構築」(H27-32)(研究代表:山名早人) 2.75億円

国際連携の進行

Int J Comput Vis (2015) 111:98–136
DOI 10.1007/s11263-014-0733-5

IJCV

ACM Multimedia

Fast and Accurate Content-based Semantic Search in 100M Internet Videos

The PASCAL Visual Object Classes Challenge: *f*

Mark Everingham¹, S. M. Ali Eslami¹, Luc Van Gool²,
Christopher Sweeney³, Andrew Zisserman¹
英・スイス・ベルギー

Lu Jiang¹, Shou-I Yu¹, Deyu Meng¹,
Yi Yang¹, Tsing-Ming Hu¹, G. Hauptmann¹
米・中・オーストラリア
¹ Carnegie Mellon University, Pittsburgh, PA, USA
(lujiang, iyu, teruko, alex)@cs.cmu.edu, dymeng@mail.xjtu.edu.cn

- 日本の研究者らの国際連携の立ち遅れが感じられる
- 研究に必要な技術が多様化しており、総合力が求められる
- 欧州のような予算面での連携の推奨も有効か

ACM Multimedia

Deep Face Beautification

Jianshu Li^{1*}, Chao Xiong²
jianshu@u.nus.edu superx21@gmail.com
Luqi Liu³, Xiangbo Shu^{4*}, Shuichang Zhou⁵
llq667@gmail.com shu_xb04@mails.jnu.edu.cn shuichang@tencent.com
eleyans@tencent.com
¹ School of Computing, National University of Singapore, Singapore
² Department of Electrical and Electronic Engineering, Imperial College London
³ School of Computer Science and Engineering, Nanjing University of Science and Technology
⁴ Department of Electrical and Computer Engineering, National University of Singapore
⁵ SAP Research and Innovation Singapore, SAP Asia Pte Ltd

ベンチマークとタスク設定

- PASCAL VOC
 - タスク:画像の意味分類、物体検出等
 - 欧州のコンピュータビジョン研究者らが策定
- ImageNet
 - タスク:大規模な画像の意味分類(数万カテゴリ、一千万画像)
 - 米国のコンピュータビジョン研究者
- TRECVID
 - タスク:映像解析・検索
 - 2003-2004 米国のニュース映像、2005-2006 英・中・アラビア語
 - 米国標準技術局により運営
- 与えられたタスクに諾々と従っているだけでは主導権が取れない
- 日本の技術的等の特性を生かし、かつ技術の進展の結果が日本の国益につながり、加えて世界の研究者らの興味を惹けるようなタスクは設定できないものか

研究資源

CVPR2014

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman

Ming Yang

Marc' Aurelio Ranzato

Lior Wolf

Facebook AI Research
Menlo Park, CA, USA

{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University
Tel Aviv, Israel

wolf@cs.tau.ac.il



97.35%

人間: 97.53%

400万顔画像で学習

ing, rather than the standard convolutional layers. Thus we trained it on the largest facial dataset to-date, an identity labeled dataset of four million facial images belonging to more than 4,000 identities. The

4000個人

CVPR2015

FaceNet: A Unified Embedding for Face Recognition and Clustering

Florian Schroff

fshroff@google.com

Google Inc.

Dmitry Kalenichenko

dkalenichenko@google.com

Google Inc.

James Philbin

jphilbin@google.com

Google Inc.



99.63%

1-2億顔画像で学習

5. Experi

If not mentioned otherwise we use between 100M-200M training face thumbnails consisting of about 8M different identities. A face detector is run on each image and a tight bounding box around each face is generated. These face thumbnails are resized to the input size of the respective network. Input sizes range from 128 to 224 pixels in our experiments.

800万個人