

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 26 日現在

機関番号：14301

研究種目：若手研究(A)

研究期間：2013～2016

課題番号：25700033

研究課題名(和文) エンティティマイニングに基づく情報補完機構に関する研究

研究課題名(英文) A Study on Information Complementation Based on Entity Mining

研究代表者

馬 強 (MA, QIANG)

京都大学・情報学研究科・准教授

研究者番号：30415856

交付決定額(研究期間全体)：(直接経費) 12,600,000円

研究成果の概要(和文)：本研究は、ユーザの多様な要求に応えて様々な情報・知を提供する情報補完システムを実現するため、インターネット上に射影されている人物・組織・地域・商品などの実世界エンティティに関する知のマイニング技術とそのニュース理解支援(メディアリテラシー)、投資及び観光における応用について研究開発を行いました。

本研究の成果は、人工知能のトップ国際会議IJCAIをはじめ、国内外の学術論文誌論文6本、査読付き国際会議論文17本など計43本の論文で公開しております。

研究成果の概要(英文)：In this research, toward the realization of information implementation systems which satisfy users' diverse information needs by providing various kinds of information and knowledge, we study on the fundamental technologies of entity mining to discovery knowledge from the information (related on persons, organizations, locations and products) available on the Internet. We also develop application systems in various domains, such as news understanding (media literacy), investments, and sightseeing.

Based on the research results, we have published 43 papers including 6 journal papers and 17 peer-reviewed international conference papers(IJCAI, ASONAM, SIGSPATIAL, ICME, DEXA, and so on.) .

研究分野：社会情報学

キーワード：エンティティマイニング 情報補完 関係マイニング 観光情報学 投資情報学 ニュース メディアリテラシー

1. 研究開始当初の背景

インターネット上の情報を知の創出に活用するためには、そこに射影されているエンティティ (人物・組織・地域・商品、etc.) に関する情報の発見、分析および整理が極めて重要である。エンティティの属性や関係に着目したエンティティ検索が Web 検索における重要な研究課題となりつつある。既存研究は、関連する(類似、etc.)エンティティの発見とそのランキングに関するものが多く、その精度や効率に関する課題が多く残っている。一方、インターネットでは、様々な観点や見方の情報が発信されているが、情報の不足や発信者の意図で偏った情報しか発信されない場合も多い。そのうえ、ユーザは自分の興味のある情報にしかアクセスしない傾向が強いため、情報の偏食や偏りが深刻な問題となり、反対意見など多様な視点からの情報を補完的に提示して、効率よくかつ偏りなく情報獲得を支援することが必要である。また、分かりやすいが表面的な説明しかしない Web コンテンツも多く、多様な素材を用いてより詳しく説明しているページを補完的に提示することがユーザの学習や理解支援には有効である。

2. 研究の目的

本研究は、インターネット上に射影されている人物・地域・組織などの実世界エンティティに関する知のマイニング技術および、それに基づく情報補完分析の基盤技術の確立を目的とする。これらの技術を用いて、例えば、米大統領選挙でオバマ寄りのニュースに対して、オバマに対する批判的な意見や対抗馬のロムニー寄りの記事など別の観点からの情報や、それぞれの意見について多様な事実を列挙したより詳しい情報を提供するような、情報の偏りを補正し詳しくかつバランスよく情報を提供できる情報補完システムの実現を目指す。

3. 研究の方法

本研究では、エンティティマイニングの基盤技術とそれを用いた応用システムの研究開発を同時に遂行して、エンティティマイニングの基盤技術の確立を目指す。本研究では、エンティティを、ニュースなどで報道される人・組織、SNSなどで情報を発信するユーザ、投資活動における投資者と投資商品、および上場企業を対象として研究開発を進める。

具体的に、エンティティマイニングの基盤技術の開発では、エンティティの情報の抽出・分析技術を開発し、それに基づいてエンティティに関する知のマイニング技術について開発する。

- エンティティ指向の差異分析：ニュース記事における人物などのエンティティに関する記述を抽出・分析して、記事や記述間の差異を定量的に分析手法について研究開発を行う。

- エンティティの特徴分析：SNS (Social Network Services) において絶えず生成されるコンテンツを用いて、地域の特徴および人々の行動パターンを分析する技術について研究開発を行う。さらに、株データなどと併用して企業関係を分析する手法について開発を行う。

- エンティティの要因分析：投資商品の価額などに影響を及ぼす要因を定量的に分析する手法について開発する。

さらに、情報補完システムの実現に向けて、ユーザのニュース理解支援やメディアリテラシー向上のための応用システムや、投資支援や地域・観光などにおける応用についても研究開発を行っている。

4. 研究成果

4.1 エンティティ指向の差異分析と応用

4.1.1 エンティティ指向の差異の定量分析

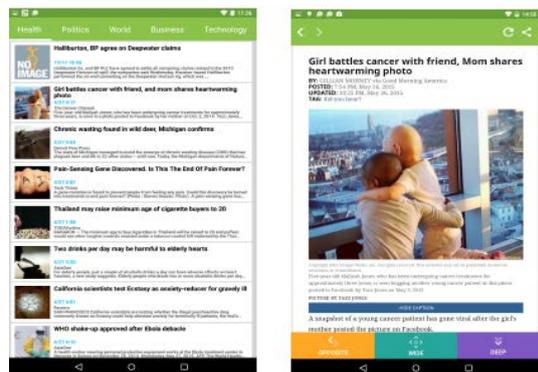
ニュース報道では、同じ話題でも報道機関によって内容は様々である。論争のある話題に対して、様々な視点の報道にアクセスすることが、偏りを補正してバランスよく情報を獲得するには重要である。

そこで、本研究では、話題に対する多様な視点から情報獲得と情報の差の発見を支援するため、ニュース報道の差異分析とその結果によるランキング手法を開発している。

本研究では、十人の大学 (院) 生を対象に実験を行い、関連ニュース記事を閲覧する際に着目する違いの要因を分析し、ニュース記事の差異を計る尺度として、視点の差 (DC)、カバレッジの差 (DC)、詳細の差 (DD) を提案している。

我々は、これらの尺度の計算手法に関する評価実験をクラウドソーシングを利用して行った。提案手法のランキングの精度 nDCG を計算した。トップ 3 記事での nDCG の値が 0.875 以上、トップ 6 記事での nDCG の値が 0.900 以上となり、提案手法が差分の定量評価に有効であることが分かった。

4.1.2 情報補完指向のニュースアプリ NewsSalad



Top News View Mode

Article Details View Mode

図1 NewsSaladの実行画面例

エンティティ指向の差異の定量分析手法を用いて、ユーザのニュースの理解を支援するための多様性指向のニュースアプリケーション News Salad を開発して Google Play に公開した。News Salad はユーザの閲覧中のニュースイベントに対して最も差のある関連記事を提示してニュースイベントを様々な角度から理解することを支援するニュースアプリである。図1は News Salad の実行画面である。

News Salad の有用性を、一般的なシステムの評価視点であるアプリへの満足度と、ユーザの心理に踏み込んだ多様な情報取得の効率性とメディアリテラシーの向上の3視点から、2週間以上のユーザ実験に基づいて確認した。

4.1.3 エンティティ指向の事象の詳細・概要関係分析

我々は、クラウドソーシングを利用して、事象の抽象的記述に対応する具体的記述を提示すれば、具体的に誰が何をしたのかを読者が容易にイメージでき、理解支援に効果があることを確認できた。これに踏まえて、本研究では、事象の抽象的記述に対して具体的記述をもって理解を支援する手法について開発を行った。我々は、まず、文章に記述される事象を Who, Whom, What, When, Where の5W要素で表現する5Wモデルを提案して、係り受け構造の解析や文章の潜在トピックによるクラスタリングを利用した事象の5Wモデルの構築手法を開発した。そして、5Wモデルを用いて、事象記述の5W要素を抽出する難易度及びそれぞれの要素の意味的包含関係に基づいて事象記述の抽象度を推定する手法、及び、5W要素ごとの階層構造に基づいて事象間の類似度を計算し、抽象度によって類似する事象間の抽象-具体の対応関係を明らかにする手法を開発した。

開発手法の評価実験では、5Wモデルを利用したランダムフォレストベースの事象記述の共参照分析手法が既存手法と比べてF値が改善されることを確認した。また、抽象度の評価実験では、単語数と抽象度を併用した手法が事象記述の抽象・具体の2値分類により結果を得られ、ベースラインより平均7%を改善してことを確認した。

4.2 エンティティ特徴分析

4.2.1 ユーザ特徴分析

・地域におけるユーザの活動範囲の推定

人々が日常生活において住んでいる、あるいは活動している範囲を知ることは行政的にも商業的にも重要である。我々は、複数の人々が日常生活において活動している領域を地域生活圏としてツイートをを用いて検出する手法を提案している。

まず、ユーザ個人が日常生活において活動している領域（個人領域）を投稿から推定する。そして、推定した個人領域を統合する

ことで、地域生活圏を生成する。

地域生活圏は地理的、あるいは心理的な境界線で区切られることが考えられ、単純な物理的距離をクラスタリングに用いるのは適当ではないため、我々は、個人領域間をユーザが移動する頻度に着目した共訪問を考慮した認知距離という新たな概念を導入し、生活圏の生成に用いる。評価実験では、この認知距離の有効性を確認した。

・ユーザの地域への熟知度の推定

ユーザの多様なニーズを把握するため、地域・スポットに関する知識の有無・多寡によるユーザ分類手法がある。地域についてよく知っている住民と知識のあまりない観光客の振る舞いの違いに着目して、ユーザの地域熟知度の推定とそれに基づく分類手法として、フォロワーなどの友人関係に着目したソーシャルネットワークモデルと、地域への訪問回数・時間に基づく時間モデルと、訪問場所に基づく空間モデルが開発した。

データの性質によってこの三つのモデルのパフォーマンスが異なるため、利用可能なデータの特徴に応じて使い分けするか、アンサンブル学習手法を適用して統合モデルを作成する手法を開発した。北京、サンフランシスコと京都の約1000名Flickerユーザデータを用いた評価実験では、アンサンブル学習手法によるユーザ分類の精度が最もよい結果となった。

・投資活動におけるユーザの特徴分析

ソーシャルトレーディングは、投資家が意見や投資情報を共有できる新しい投資手法である。ソーシャルトレーディングサービスでの利益額は、フォローするトレーダのパフォーマンスに依存する。従って公開されているトレーダの情報を考慮して適切なトレーダを選ぶことが重要である。しかしながら、サービスが提供している評価指標は過去の一定期間の結果に基づいて算出されるため、日々の変動が激しい。安定したランキングではないため、ユーザに推薦されるトレーダの一貫性がない。

そこで、本研究では、トレーダの取引履歴を分析して、performance, risk と consistency の三つの評価指標を提案し、トレーダの安定的なランキングの実現を試み

W2F (Whom to Follow)

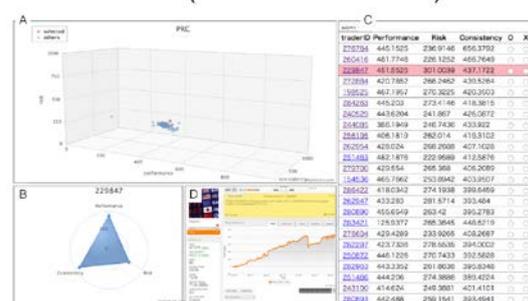


図2 W2Fの実行例

た. この三つの尺度は基本的に、ユーザの過去の履歴との比較、およびその他のユーザの比較で算出される. この三つの尺度を用いたトレーダ検索・ランキングシステム W2F (Whom to Follow)を開発した. フィードバック機能を導入して、インタラクションしながらユーザがフローするトレーダを選べるようになっている. 図2はW2Fの実行例である.

100名のトレーダのデータを用いた評価実験では、この三つの尺度を併用した手法が最もよいランキング結果 (nDCG) を達成した. また、W2Fで上位ランキングされたk名のトレーダ (k=5, 10, 15, 20) を用いたシミュレーションでは、既存手法より収益が大きいことも確認した. さらに、1か月以上の履歴データを用いれば、W2Fが安定したランキングを得られることも確認できた.

4.2.2 地域特徴分析

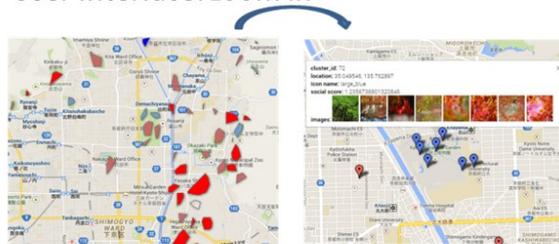
本研究では、まだ知られていないが潜在的に観光価値の高い穴場などの多様な観光スポットを発見するため、観光スポットを、1) 人気度 (ユーザ分布) と2) “質” (観光価値) の二つの軸から分析するスポットの分析フレームワークを提案している.

本研究では、ユーザの地域への熟知度の分析手法を利用して、ユーザ分布を考慮した人気度分析手法を開発し、異なった視点からの観光資源の再発見に試みる. また、スポットの観光価値を、Flickrにアップしたジオタグ付きの写真データを用いた評価手法を開発している.

- ユーザ行動ベースの間接評価: 写真をアップしたユーザと閲覧したユーザの行動を分析してスポットの価値を評価する. 景観のよいスポットを訪ねたユーザは、そのスポットの全体と部分の両方の写真を撮ることが多い. この観察に基づいて、スポットの写真を遠景と近景に分類して、その割合を調べることで観光価値を評価する手法を開発している. 同時に、アップした写真の閲覧者の行動 (評判) を利用した分析手法も開発している.
- 景観ベースの直接評価: スポットで撮った写真を対象に、環境心理学の知見を活用して、写真に写されている景観物の色、配置、種類などを分析して、スポットの自然や人文的景観価値を直接推定する手法を開発している.

京都・蘇州などを対象とした評価実験では、地域・スポットを訪ねたユーザの熟知度を分析すれば、観光客に知られていない地元住民がよく行くスポット、また、その逆のスポットの発見が可能となったことが分かった. また、画像の質が景観の質とは関連あるものの、イコールではないことが分かった. 図3は我々が開発した穴場探索システムの実行例である.

User Interface: zoom in



Zoom out

図3 穴場探索システム

4.2.3 企業分析

株価が企業の業績を反映することが多い. また、業績に影響を及ぼす関係が存在すれば、両企業の株価に相関の高い動きが現れることが多いと考えられる. したがって、株価の動きの関連性を調査することで、企業間の関係性を明らかにすることが可能である.

我々は、企業自身の要因に影響される株価を実価、業種の影響度合いを業種指数、市場の影響度合いを市場指数として、企業の株価はこの三つの要素からなる株価の合成モデルを構築する. 株価の合成モデルを用いて求められた実価の相関を分析して企業間の関係性を発見する.

評価実験では、合成モデルを用いて算出された実価を用いた手法が企業間の関係をより正確にとらえることが分かった. また、ニュースイベントを用いたほうが、計算量を削減できるほか、関係の検出精度の向上にも効果があったことが確認できた.

4.3 エンティティの要因分析

投資信託商品は、専門知識がなくても投資が出来ることから、金融商品として大変人気がある. しかし、現状では、適切な情報利用もできていないと言える. 我々は状態空間モデルを適用して商品理解を深めるための要因の定量分析手法を開発している. 要因の定量分析は基本的に状態空間モデルを用いて行うが、状態空間モデルを構築する際、要因の選別および要因の有効時間 (影響範囲) の推定も行う点が独創である.

評価実験では、2016年1月から10月の期間の国内株式を対象とした投資信託商品を

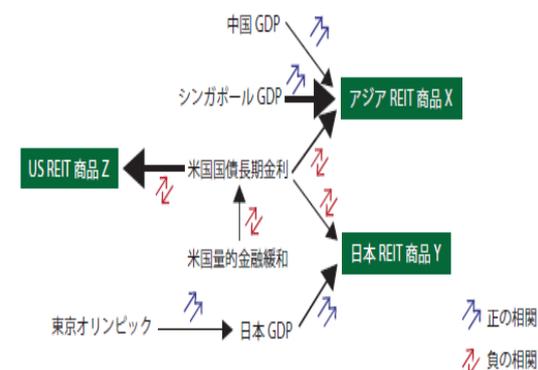


図4 商品の要因比較

用いた。実験結果から要因の抽出および分類を高い精度で行えることを確認できた。また、トレンドごとに状態空間モデルを構築した状態空間モデルの誤差が、トレンドを考慮しない場合より小さいことを確認できた。

さらに、図5のような投資商品を比較するインタフェースも開発している。開発したユーザインタフェースを利用した場合、既存サービスの半分の時間で商品の特徴を理解可能となったことをユーザ実験の結果から分かった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 6 件)

1. Yizhu Shen, Min Ge, Chenyi Zhuang, Qiang Ma: Sightseeing Value Estimation by Analyzing Geosocial Images. International Journal of Big Data Intelligence (to appear). 査読有
2. 馬 強: 投資支援のための関係マイニング技術の開発。技術情報誌 Telecom Frontier 95:1-6, 2017. 査読無, <http://www.scit.or.jp/frontier/frontier95/ma.pdf>, 招待論文
3. 馬 強: 観光情報の最前線-観光の分散化と個人化を促進する集合知活用情報技術-. 情報処理 58:220-226, 2017. 査読無, https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=177477&file_id=1&file_no=1, 解説記事
4. Chenyi Zhuang, Qiang Ma, Masatoshi Yoshikawa: SNS user classification and its application to obscure POI discovery. Multimedia Tools Appl. 76(4): 5461-5487 (2017). 査読有, DOI 10.1007/s11042-016-4034-6
5. Keisuke Kiritoshi, Qiang Ma: Named Entity Oriented Difference Analysis of News Articles and Its Application. IEICE Transactions 99-D(4): 906-917 (2016). 査読有, <http://hdl.handle.net/2433/210129>
6. Jiyi Li, Qiang Ma, Yasuhito Asano, Masatoshi Yoshikawa: Tag Quality Improvement for Social Image Hosting Website. 情報処理学会論文誌: データベース 6(3):177-186 (2013). 査読有, DOI 10.2197/ipsjtrans.6.75

[学会発表] (計 37 件)

1. Chenyi Zhuang, Nicholas Jing Yuan, Ruihua Song, Xing Xie, Qiang Ma: Understanding People Lifestyles: Construction of Urban Movement Knowledge Graph from GPS Trajectory, IJCAI 2017 (to appear), 2017. 8. 24, オーストラリア・メルボルン

2. Yizhu Shen, Chenyi Zhuang, and Qiang Ma: Element-oriented Method of Assessing Landscape of Sightseeing Spots by using Social Images, APWeb-WAIM Joint Conference on Web and Big Data 2017 (to appear), 2017. 7. 8, 中国・北京
3. 馬 強: 観光情報学 2.0: 観光の分散化と個人化を促進する集合知活用情報技術, IT コンソーシアム京都「知的情報処理技術が変える観光情報サービスと地域情報の活用」, 2017. 2. 22, 京都府・京都市, 招待講演.
4. 中村玄貴, 馬 強: コンテキストウェア敵対的生成ネットワークによる画像生成, DEIM2017, 2017. 3. 7, 岐阜県・高山市
5. 馬場慧, 馬 強: 株価とニュースの統合分析のためのヘテロトピックモデル, DEIM2017, 2017. 3. 7, 岐阜県・高山市
6. 大西恒彰, 馬 強: 状態空間モデルの拡張による投資信託商品の要因分析, DEIM2017, 2017. 3. 7, 岐阜県・高山市
7. 竹田創, 庄晨熠, 馬 強: ソーシャルトレーディングサービスにおけるトレーダーの特徴分析, DEIM2017, 2017. 3. 7, 岐阜県・高山市
8. Yizhu Shen, Min Ge, Chenyi Zhuang, Qiang Ma: Sightseeing Value Estimation by Analyzing Geosocial Images. BigMM 2016: 117-124, 2016. 4. 20, 台北
9. Satoshi Baba, Qiang Ma: Analyzing Relationships of Listed Companies with Stock Prices and News Articles. DEXA (2) 2016: 27-34, 2016. 9. 6, ポルトガル・ポルト
10. Shintaro Horie, Keisuke Kiritoshi, Qiang Ma: Abstract-Concrete Relationship Analysis of News Events Based on a 5W Representation Model. DEXA (2) 2016: 102-117, 2016. 9. 7, ポルトガル・ポルト
11. 大西 恒彰, 馬 強: 月次報告書と基準価格を用いた投資信託商品の要因分析, DEIM2016, 2016. 2. 29, 福岡県福岡市
12. 馬場 慧, 馬 強: 株価とニュース報道を用いた上場企業の暗黙関係の発見, DEIM2016, 2016. 3. 1, 福岡県福岡市
13. 堀江 伸太朗, 切通 恵介, 馬 強: ニュース事象の 5W モデルとそれに基づく事象関係分析, DEIM2016, 2016. 3. 1, 福岡県福岡市
14. 切通 恵介, 楠見 孝, 堀江 伸太朗, 馬 強: 多様性指向のニュースアプリの開発とその有用性評価, DEIM2016, 2016. 2. 29, 福岡県福岡市
15. Chenyi Zhuang, Qiang Ma, Xuefeng Liang, Masatoshi Yoshikawa: Discovering Obscure Sightseeing Spots

- by Analysis of Geo-tagged Social Images. ASONAM 2015: 590-595, 2015. 8. 27, フランス・パリ
- 1 6. Keisuke Kiritoshi, Qiang Ma: A Diversity-Seeking Mobile News App Based on Difference Analysis of News Articles. DEXA (2) 2015: 73-81, 2015. 9. 1, スペイン・バレンシア
- 1 7. Chenyi Zhuang, Qiang Ma, Masatoshi Yoshikawa: Location familiarity based flickr photographer classification for POI mining. SIGSPATIAL/GIS 2015: 84:1-84:4, 2015. 11. 6, アメリカ・シアトル
- 1 8. Woonyeol Lee, Qiang Ma: Whom to Follow on social trading services? A system to support discovering expert traders. ICDIM 2015: 188-193, 2015. 10. 21, 韓国・済州
- 1 9. 堀江 伸太郎, 切通 恵介, 馬 強 関係事象の組織化によるニュース記事の理解支援 電子情報通信学会技術研究報告 115(177):85-90, 2015. 7. 29, 奈良県・奈良市
- 2 0. 田原 琢士, 馬 強 : Twitter ユーザの活動するローカル地域の発見, 情報処理学会第 77 回全国大会, 2015. 3. 18, 京都府京都市
- 2 1. Lee Woonyeol, 馬 強 : ソーシャルトレーディングにおけるコピートレーダーの推薦方法の提案, 情報処理学会第 77 回全国大会, 2015. 3. 18, 京都府・京都市
- 2 2. Ling Xu, Keisuke Kiritoshi, Qiang Ma and Masatoshi Yoshikawa : Entity-oriented Materials and Opinions Extraction from News Articles, 情報処理学会第 77 回全国大会, 2015. 3. 18, 京都府・京都市
- 2 3. 切通恵介, 許冷, 馬 強 : 報道の多様性を考慮したニュース閲覧システムの提案 情報処理学会第 77 回全国大会, 2015. 3. 18, 京都府・京都市
- 2 4. Keisuke Kiritoshi, Qiang Ma: Named Entity Oriented Related News Ranking. DEXA (2) 2014: 82-96, 2014. 9. 1, ドイツ・ミュンヘン
- 2 5. Takuji Tahara, Qiang Ma: Searching for Local Twitter Users by Extracting Regional Terms. DEXA (1) 2014: 89-96, 2014. 9. 1, ドイツ・ミュンヘン
- 2 6. Chenyi Zhuang, Qiang Ma, Xuefeng Liang, Masatoshi Yoshikawa: Anaba: An obscure sightseeing spots discovering system. ICME 2014: 1-6, 2014. 7. 14, 中国・成都
- 2 7. Yuki Awano, Qiang Ma, Masatoshi Yoshikawa: Causal Analysis for Supporting Users' Understanding of Investment Trusts. iiWAS 2014: 524-528, 2014. 12. 4, ベトナム・ハノイ
- 2 8. Qiang Ma, Keisuke Hasegawa: Organizing Sightseeing Tweets Based on Content Relatedness and Sharability. WAIM 2014: 510-521, 2014. 6. 18, 中国・マカオ
- 2 9. Yuya Hirano, Qiang Ma, Masatoshi Yoshikawa: Linkage of compound objects for supporting maintenance of large-scale web sites. ICUIMC 2014: 75:1-75:8, 2014. 1. 10, カンボジア・シェムリアップ
- 3 0. 切通恵介, 馬 強 Named Entity に対する記述の差異分析に基づく関連ニュースランキング手法, DEIM 2014, 2014. 3. 5, 兵庫県・淡路市
- 3 1. 田原琢士, 馬 強: Twitter から有益な日常情報を発見するための 特徴語による地域ユーザの検索, DEIM 2014, 2014. 3. 5, 兵庫県・淡路市
- 3 2. 長谷川 馨亮, 馬 強, 吉川 正俊: Twitter からの地域特徴語辞書の構築とその観光情報検索への応用, DEIM 2014, 2014. 3. 5, 兵庫県・淡路市
- 3 3. 栗納 裕貴, 馬 強, 吉川 正俊: 因果関係分析に基づく投資関連商品の比較手法の検討, DEIM 2014, 2014. 3. 5, 兵庫県・淡路市
- 3 4. Jiyi Li, Qiang Ma, Yasuhito Asano, Masatoshi Yoshikawa: Potential Semantics in Multi-modal Relevance Feedback Information for Image Retrieval. COMPSAC 2013: 830-831, 2013. 7. 22, 京都府京都市
- 3 5. 平野雄也, 馬 強, 吉川正俊: Web サイト管理支援のための複合オブジェクトのリンク手法の提案, WebDB2013, 2013. 11. 28, 京都府京都市
- 3 6. Liang Yan, Qiang Ma, Masatoshi Yoshikawa: Classifying Twitter Users Based on User Profile and Followers Distribution. DEXA (1) 2013: 396-403, 2013. 8. 26, チェコ・プラハ
- 3 7. Chenyi Zhuang, Qiang Ma, Xuefeng Liang, Masatoshi Yoshikawa: Discover "Anaba" Sightseeing Spots Using Social Images. 第 157 回 DBS 研究会, 2013. 7. 15, 北海道・札幌
- [その他]
<https://www.db.soc.i.kyoto-u.ac.jp/soc/>
<https://www.db.soc.i.kyoto-u.ac.jp/~qiang/topic-j.html>
 NewsSalad: goo.gl/X3CIja
6. 研究組織
 (1) 研究代表者
 馬 強 (Qiang Ma)
 京都大学大学院情報学研究科 准教授
 研究者番号: 30415856