様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２５年３月３１日現在

機関番号：１２６０１
研究種目：基盤研究(A)
研究期間：2009～2012
課題番号：２１２４００１１

研究課題名（和文）　機械学習によるロングテール現象の解決方法に関する研究

研究課題名（英文）　A STUDY OF RESOLVING LONG TAIL PHENOMENA BY MACHINE LEARNING

研究代表者
中川　裕志　　(NAKAGAWA HIROSHI)
東京大学・情報基盤センター・教授
研究者番号：20134893

研究成果の概要（和文）：
2009年度は当初の予定通り、Webにおける人名検索結果を同姓同名であるが異なる人物ごとにまとめるクラスタリングシステムを開発し、実験的に評価した。2010年度は大規模データ処理のために非負の確率行列分解アルゴリズムを提案し、既存のLDAと同様な性能を得ることを実証し、並列化アルゴリズムにおいては変分ベイズ法をロングテールに対応するPitMan-Yoモデルに適用し高い性能を得た。2011年度は最近注目されているプライバシー保護データマイニングをネットワークデータに応用した。2012年度は、プライバシー保護データマイニングの応用手法と大規模データに適したオンライン学習で、新規な正則化手法を提案した。

研究成果の概要（英文）：
We developed a clustering system which makes clusters of web pages in response to a person name query in 2009 as planned, and evaluate it experimentally. In 2010, our contribution is a new non-negative probabilistic matrix decomposition algorithm and application of Variational Bayes method to Pitman-YO process. In 2011, our contribution for PPDM is a link analysis algorithm with public key encryption and specific protocol. In 2012, we developed a new online learning algorithm as well as new PPDM method.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	13,500,000	4,050,000	17,550,000
2010 年度	11,500,000	3,450,000	14,950,000
2011 年度	11,200,000	3,360,000	14,560,000
総　計	36,200,000	10,860,000	47,060,000

研究分野：情報学
科研費の分科・細目：知能情報学
キーワード：知識発見、データマイニング、機械学習、テキストマイニング、Ｗｅｂ

１．研究開始当初の背景

検索エンジンの検索結果上位にランクされないロングテールの情報の中に重要な情報が存在するにもかかわらずアクセス困難になっているというロングテール現象が顕在化している。また、最近では組織内情報が爆発的に増大しているにもかかわらず、目的に合致した情報を探しにくいという組織内ロングテール現象も顕著である。
このような状況に対して、対象文書群や応用目的、さらには個別言語に特化しない汎用性の高い解決策として、申請者は検索エンジンの検索結果、あるいは組織内データであればその全体を対象にして、申請者がこれまで行ってきた統計的機械学習による自然言語処理、特に文書クラスタリングと、同義語処理

を加味した文字列検索を中心とした技術を発展させた次の解決策を実現する。インターネットのデータについてはオンデマンド型の固有名曖昧性解消システムを開発してきている。これらの継続に加え、必要になる機械学習のアルゴリズム開発を行う。

２．研究の目的

このような状況に対して、対象文書群や応用目的、さらには個別言語に特化しない汎用性の高い解決策として、申請者は検索エンジンの検索結果、あるいは組織内データであればその全体を対象にして、申請者がこれまで行ってきた統計的機械学習による自然言語処理、特に文書クラスタリング技術を発展させた解決策を実現する。

３．研究の方法

(1)　ロングテール情報抽出のための機械学習：Webに存在するテキストにおいてロングテール化によってアクセスが困難な情報を抽出するためのアルゴリズムを開発する。初年度に名前参照の曖昧性解消のための実時間クラスタリングで成果をあげ、2年目は、テキストに現れる量的な情報を抽出するアルゴリズムを開発する。

　具体的には、名前参照の曖昧性解消の結果をWeb People Search Task という国際会議でのタスクに適用する。この内容を記載した論文が情報検索の最難関国際会議SIGIR’2010に投稿する。さらに、ロングテール現象の対応策のひとつである同義語抽出アルゴリズムを提案し、実験的に評価して査読論文として発表する。

(2)機械学習アルゴリズムの改善：申請者らは既にトピックモデルによるクラスタリングで成果をあげている。これを1)機械学習における探索範囲の拡大、2)並列化による高速化、の点で強化するための数理モデル化およびアルゴリズム開発を行う。基本的にはシミュレーテッドアニーリングを根本的に改造した量子アニーリングのモデル作成、LDAのようなトピックモデルにPitman-Yo過程を適用を行い、定量的に評価する。

(3) Wikipediaから検索で利用者にとって意外性の高い記事を特徴付けるモデルを提案し評価実験を行う。

(4)プライバシー保護データマイニング： 組織内のデータベースにおいてデータベースの内容自体は公開しないが、検索によって統計的処理結果だけ応答する場合、データベースの個別項目の変更を保護するプライバシー保護データマイニングのモデルの研究を行う。このモデル化によって、企業内のデータベースを公開する手段を提供し、企業内データにおけるロングテール情報を抽出するモデルを確立する。

　具体的には、複数の通信プロバイダが自己の顧客間および自己の顧客と他の通信プロバイダの間のアクセス情報は持つが、相手側の顧客間のアクセスは秘匿されたようなネットワーク構造のリンク解析を行う。さらにネットワークのリンク解析を行うＥＭアルゴリズムにおいて、プライバシー保護のためにデータは準同型性公開鍵暗号で暗号化し、暗号化したままで行える加算を基礎にしたアルゴリズムおよび参加者間のプロトコルを設計する。

(5)　オンライン学習アルゴリズム：ロングテール情報を喪失する傾向があることへの対策として、識別器が安定した重み持つ特徴は排除しない制約を加え、この問題への解決の有力な解決策を示す。

４．研究成果

(1)人名 Web 検索結果の曖昧性解消

人名の検索は、Web検索における主要なタスクの一つであるが、そこには常に曖昧性の問題が付いて回る。例えば、「吉田稔」という文字列だけでは、それがどの人を指しているのか特定することは難しい。何故なら、「吉田稔」という名前を持つ人は全国に多く存在するからである。このため、人名でのWeb_ 検索では、検索結果が目的の人のページなのかどうかについて常に気を配る必要がある。人名曖昧性解消とは、このような問題に対し、同じ人物のページを自動的にまとめあげ、検索結果を見易くするというタスクである。近年、人名曖昧性解消に関する国際ワークショップWePS(Web People Search Workshop)が開催されており、各国から様々なシステムが参加してその精度を競い合っている。WePSを通じて、特に人名クエリと共起する固有名詞に着目した手法が効果を発揮することが明らかになっている。例えば、Bill Gates という検索クエリを考えたとき、検索結果の複数のWeb文書にPaul Allenという人名が共起していれば、それらの文書は同一人物である可能性が高い。固有名詞の利用によって、高い精度で同一人物を判定できる半面、こうした固有名詞は、必ずしも文書に出現するとは限らず、網羅性という面で限界がある。

　網羅性を上げるための方策として最初に上げられるのが、固有名詞以外の一般の単語（名詞）を用いることであるが、一般の単語は、人物との関連性が高い単語ばかりではないため、これらを手掛かりとすることによって、関連のない人物どうしを結びつけてしまうミスが発生しやすくなる。これに対し本研究で

は、一般の単語のうち、有用な単語に高い重みを与え、手掛かりとして用いる手法を提案する。そのために、固有名詞などの精度の高い特徴量を用いた、第一段階クラスタ（クラスタ＝文書のまとまり）を生成し、これら第一段階クラスタの結果を利用することで一般の単語に重みを付与する。この重みづけされた単語を用い第二段階クラスタを生成する。このため、提案手法を、二段階クラスタリング法と呼ぶ。これにより、例えば、野球選手とコンピュータ科学者が同姓同名だった場合に、「ホームラン」や「ボール」、あるいは「メモリ」や「アルゴリズム」といった単語を重みの高い単語として人物の判定に利用することが可能になる。国際ワークショップ（コンテスト）第2回WePSへの参加を行い、17チーム中2 位、F 値=0.81の成績を収めた。この成果をまとめた論文が情報検索分野の世界最難関国際会議ACM SIGIR2010に採録された[学会発表12]。

(2)文字列処理を基礎とするロングテール情報の検索

近年,Web上や組織内に蓄積される電子的文書の量は増大の一途を辿り、それらの文書を人間が把握することが困難となっている。Web全体における文書量の増大のみならず、その中でのトピックの限定された部分集合（特定サイト内のWWW文書集合、Wikipedia、さらには企業内文書集合等）のサイズも増大し、WWW全体と同様に、把握が困難となりつつある。そのような「一貫性のある大規模な文書集合」の例として、Wikipedaに着目し、Wikipedaにに対するクエリ入力支援システムとして、「テキストマイニングによる検索支援システム Wikiwi」を提案した。テキストマイニングとは、与えられたテキスト集合の中での、「言葉の使われ方」（主に、言葉に関する統計的情報）について分析するタスクである。

テキストマイニングとは、与えられたテキスト集合の中での、「言葉の使われ方」（主に、言葉に関する統計的情報）について分析するタスクである。Wikiwiでは、入力されたクエリに対し、「用例抽出」「同義語抽出」という二種類のテキストマイニングをリアルタイムに行い、マイニング結果を提示する。我々のシステムは、テキストや索引をオンメモリに配置したSuffix Array検索がベースとなっているが、Wikipediaが対象であることで、オンメモリにテキストを配置しながら、網羅性の非常に高いクエリ補完システムを実現できる。

図１に、Wikiwiの画面を示す。図中、A.に示されているのが検索窓であり、ユーザがここにクエリを入力すると、システムは入力内容に応じて用例や同義語を提示する。ここで用例とは、クエリの前方、後方にどのような文字列が頻繁に連接しているかを示したものであり（図中「B.後方連接」「C.前方連接」）、この場合、「マンション」の前方に「高層」や「分譲」、後方に「建設」や「管理」といった言葉が続き易いことがわかる。同様に、マンションの同義語・類義語として、「住宅」や「アパート」という文字列が用いられやすいということも提示する（図中、「D.類義語・同義語」）。ユーザは、提示された用例等をクリックすることで、クエリの補完を行える。補完された結果は、そのままGoogleへのクエリとして、検索に用いることができる。データとしては、Wikipediaの2007年5月29日版のスナップショットから、<text>タグで囲まれた部分を取得した。取得したテキストはそのままコーパスとして用い、それ以上の後処理（タグの除去等）は行っていない。

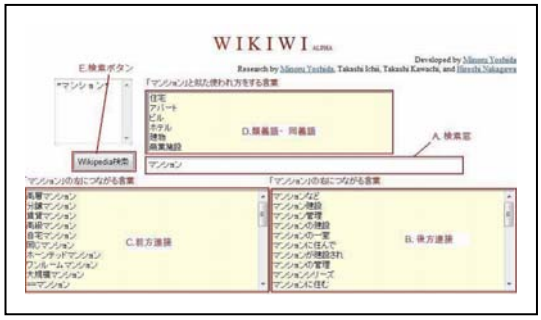

図１ Wikiwiの実行画面

(3)オンライン学習

PassiveAggressive アルゴリズムと呼ばれるオンライン教師あり学習アルゴリズムにおいて、多クラス分類アルゴリズムの研究で成果をあげた。その結果、多クラス分類精度が大幅に改善した。この成果はデータマイニングにおけるトップコンファレンスであるSIAM SDM [学会発表 13]で採択され 同会議 top 12 論文として表彰され、さらに [雑誌論文 10] に採録された。次に、この手法を半教師あり学習に応用する方法を提案、評価した[雑誌論文 8]。

統計的機械学習においては、教師データ（入力と正しい分類の対）を用いて,データの各素性の重みを成分とする重みベクトル $\mathbf{w}$ によって入力ベクトル $\mathbf{x}$ があるカテゴリに分類されれば $\mathbf{x}^T\mathbf{w}>0$ 分類されなければ $\mathbf{x}^T\mathbf{w}<0$ となるような重みベクトル $\mathbf{w}^*$を以下の最適化問題で求める。

$$\mathbf{w}^* = \mathrm{argmin}_{\mathbf{w}} \left(\text{損失}(\mathbf{x}, \mathbf{w}, \text{正しい分類}) + \text{正則化項}(\mathbf{w}) \right)$$

損失は 2 乗誤差あるいはヒンジ誤差などである。正則化項は $\mathbf{w}$ が不必要に複雑になることを避ける目的で導入された。従来、正則化項には解析の容易さから 2 乗ノルム(L2 ノルム)が用いられたが、最近はより強力に複雑さを低減させる各素性値の絶対値の和で表され

るノルム(L1 ノルム)も用いられ始めた。ただし、L1 ノルムを正則化項に用いると、素性の重みをゼロ化する力が強すぎて、w の重要な成分までゼロにしてしまう傾向がある。我々は、この L1 ノルム正則化の問題点を緩和する数理モデルを考案し、高い分類性能と w の簡素さを両立させることに成功した。重要なポイントは、学習において不安定な変化が多い成分のみを素性の重みをゼロにすることである。より具体的には L1 ノルムの各素性対応する重みに、その重みの変化回数を乗ずる。この結果、少ない素性で高い予測性能を持つ学習器 w を構成できた。この結果は機械学習分野の難関国際会議 ICDM2012 などで論文が採択[学会発表 2]された。

(4) プライバシー保護データマイニング
我々は準同型性公開鍵暗号を利用した PPDM をネットワークデータに応用するアルゴリズムを考案した。複数のネットワークプロバイダが存在し、各プロバイダは自分の顧客間のアクセス頻度は知っている。他のプロバイダに関しては、自分の顧客と他のプロバイダの顧客間のアクセス頻度は知っているが、他のプロバイダの顧客間のアクセス頻度は知らないとする。

ここで、各プロバイダは自分のアクセス頻度の知識を他のプロバイダに漏らすことなく、ネットワークの顧客全体に対する接続行列の主固有ベクトルを計算するスペクトラルランキングのアルゴリズムを設計し、プライバシー保護型のリンク解析ができるようになった。このアルゴリズムを応用すると各顧客のページランクを求めることもできる。さらに、顧客同士のアクセス状態（あるいは接続状況）に応じた EM アルゴリズムを実行する PPDM アルゴリズムも提案した[雑誌論文 2]。

(5)トピックモデルと教師なし学習
オンライン教師なし学習アルゴリズムとして、オンライン LDA(潜在ディリクレ配置)法を提案した。LDA のオンライン化は、処理対象を 1 データずつメモリに読み込み、トピックなどに対応する潜在変数を更新するものである。従来は、潜在変数のかなりの履歴データを記憶しておく必要があったが、この提案では直前の潜在変数だけを記憶しておき、新たなデータからの学習結果と重み付けして加算する手法を採り、メモリ量の削減を計った。この重み付けのための重み係数であるが、数理モデルを工夫して最適な重み係数を求めることに成功した。この成果は[学会発表 9] に採択された。

これ以外には Pitman-Yor 過程を利用してロングテール対応を図ったトピックモデル抽出の機械学習アルゴリズムトピックモデルの提案[学会発表 10]、Wikipedia からの気づきにくい有用な情報の抽出[学会発表 14]などの研究成果をあげた。

５．主な発表論文等

〔雑誌論文〕（計 13 件）

① Yo Ehara, Nobuyuki Shimizu, Takashi Ninomoya, Hiroshi Nakagawa: Personalized Reading Support for Second-Language Web Documents. 査読有, ACM Transactions on Intelligent Systems and Technology, 4(2). 2013. (accepted)
② Yang Bin, Hiroshi Nakagawa : Privacy-Preserving EM Algorithm for Clustering on Social Network. 査読有, P.-N. Tan et al. (Eds.): PAKDD 2012, Part I, LNAI 7301, pp. 542-553, 2012. Springer-Verlag Berlin Heidelberg 2012
③ Shingo Takamatsu, Issei Sato, Hiroshi Nakagawa: Probabilistic Matrix Factorization Leveraging Contexts for Unsupervised Relation Extraction. 査読有, PAKDD2011, Springer Lecture Notes Artificial Intelligence (LNAI)6634, Part I. pp.87-99, Shenzhen, China on May 24-27, 2011 (accepted as long paper, acceptance ratio 32/331=9.7%)
④ 大岩秀和, 松島慎, 中川裕志：特徴の出現回数に応じた L1 正則化を実現する教師ありオンライン学習手法. 査読有, 情報処理学会論文誌, Vol.50 TOM 4(3). pp.84-93. 2011.
⑤ 森井正覚, 佐久間淳, 佐藤一誠, 中川裕志：統合したグラフのプライバシ保護リンク解析. 査読有, 情報処理学会論文誌, Vol.50 TOD 4(2). pp.52-60. 2011.
⑥ Takashi Ninomiya, Takuya Matsuzaki, Nobuyuki Shimizu and Hiroshi Nakagawa. Deterministic shift-reduce parsing for unification-based grammars, 査読有, Natural Language Engineering, vol. 17, no. 3, pp. 331-365.(2011)
⑦ 佐藤一誠、中川裕志. Succinct Semi-structured Data Mining Based on FREQT. 査読有, 日本データベース学会論文誌. Vol.9, No.1. pp. 76-81. 2010
⑧ 松島慎、佐藤一誠、二宮崇、中川裕志. PA アルゴリズムにおけるラベルなしデータからの学習. 査読有, 日本データベース学会論文誌. Vol.9, No.1. pp. 82-87. 2010
⑨ Minoru Yoshida. Hiroshi Nakagawa. Mining Numbers in Text Using Suffix Arrays and Clustering Based on Dirichlet Process Mixture Models. 査読有, (PAKDD 2010) Part II. Springer LNAI 6119. pp.230-237. 2010

⑩ 松島慎、清水伸幸、吉田和弘、二宮崇、中川裕志．多クラス識別問題における Passive- Aggressive アルゴリズムの効率的厳密解法．査読有，電子情報通信学会論文誌：情報爆発特集号、Vol.J93-D. No.6. pp724-732．2010

⑪ Nobuyuki Shimizu, Masashi Sugiyama, Hiroshi Nakagawa: Spectral Methods and Text Mining: Automatic Expansion of User，査読有，*IEICE Transactions*, E93-D, 6, special issues on Info-Plosion, pp. 1378-1385, June 2010.

⑫ 吉田稔，中川裕志，寺田昭．コーパス検索支援のための動的同義語候補抽出，査読有，人工知能学会論文誌 25(1), pp.122-132．2010.

⑬ 柴山直樹、中川裕志．確率的潜在意味解析における特異値行列の非対角化の解釈とその評価．査読有，人工知能学会論文誌 Vol.26, No.1, pp.262-272, 2010 年 10 月．

〔学会発表〕（計 24 件）

① Yo Ehara, Issei Sato, Hidekazu Oiwa,and Hiroshi Nakagawa: Mining words in the minds of second language learners: learner-specific word difficulty. 査読有，25th International Conference on Computational Linguistics (COLING 2012), pp.799-814. Mumbai, India. Dec. 8-15, 2012 (long paper)

② Hidekazu Oiwa, Shin Matsushima, and Hiroshi Nakagawa: Healing Truncation Bias : Self-weighted Truncation framework for Dual Averaging. 査読有，12th IEEE International Conference on Data Mining(ICDM), pp.575-584.Brussels. Dec. 10-13, 2012 (long paper, acceptance ratio 10.7%)

③ Issei Sato, Ken-ich Kurihara, Hiroshi Nakagawa. Practical Collapsed Variational Bayes Inference for Hierarchical Dirichlet Process. 査読有，18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDD 2012),pp.105-113, Beijing, China , Augst 12- August 16 , 2012, (Research Track 133 / 734 = 18.1%).

④ Issei Sato, Hiroshi Nakagawa. Rethinking Collapsed Variational Bayes Inference for LDA. 査読有，29th International Conference on Machine Learning (ICML 2012) pp. 999-1006, Edinburgh, Scotland, June 26-July 1 , 2012, (long paper acceptance ratio 27.3%).

⑤ Shingo Takamatsu, Issei Sato, Hiroshi Nakagawa. Reducing Wrong Labels in Distant Supervision for Relation Extraction. 査読有，ACL 2012. pp.721-729, Jeju, Korea on July 8-14, 2012 (accepted as oral presentation: 19%)

⑥ Bing Yang, Issei Sato, Hiroshi Nakagawa: Privacy-Preserving EM Algorithm for Clustering on Social Network. 査読有，The 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2012). Kuala Lumpur, Malaysia, May29-June 1, 2012 (accepted as short paper. 88/241=36.5%).

⑦ Bing Yang, Issei Sato, Hiroshi Nakagawa: Secure Clustering in Private Networks. 査読有，11th IEEE International Conference on Data Mining(ICDM), pp.894-903. Vancouver, Canada. Dec. 11-14, 2011 (accepted as long paper, acceptance ratio 18%)

⑧ Hidekazu Ooiwa, Shin Matsushima, Hiroshi Nakagawa. Probabilistic Frequency-aware Truncated methods for Sparse Online Learning. 査読有，The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2011) , Springer Lecture Notes Artificial Intelligence (LNAI)6911, ECML PKDD , Part II . pp.533-548, Athens, Greek. Sept. 5-9, 2011 (accepted. acceptance ratio 120/599=20%)

⑨ Issei Sato, Kenich Kurihara, Hiroshi Nakagawa. Deterministic Single-Pass Algorithm for LDA. 査読有，Neural Information Processing Systems Conference (NIPS2010). 2010.

⑩ Issei Sato, Hiroshi Nakagawa. Topic Models with Power-Law Using Pitman-Yor Process. 査読有，16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,(KDD2010) pp.673-682 . 2010

⑪ Bin Yang,Hiroshi Nakagawa,Issei Sato,Jun Sakuma: Collusion - Resistant Privacy - Preserving Data Mining, 査読有，16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,(KDD2010)pp.483-492 .2010

⑫ Minoru Yoshida, Masaki Ikeda, Shingo Ono, Issei Sato, Hiroshi Nakagawa: Person Name Disambiguation by Bootstrapping, 査読有，*The 33rd Annual* ACM SIGIR *Conference*. pp. 10-19, July, 2010.

⑬ Shin Matsushima, Nobuyuki Shimizu, Kazuhiro Yoshida, Takashi Ninomiya, Hiroshi Nakagawa. Exact Passive - Aggressive Algorithm for Multiclass Classification Using Support Class, 査読有, the 2010 SIAM International Conference on Data Mining (SDM'2010) pp.301-314. 2010. This paper is selected as top 12 papers of SDM2010

⑭ Yohei Noda, Yoji Kiyota, Hiroshi Nakagawa: Discovering Serendipitous Information from Wikipedia by Using its Network Structure, 査読有, In Proceedings of *4th Int'l AAAI Conference on Weblogs and Social Media* (*ICWSM 2010*), poster session, pp. 299-302, Washington, D.C., USA, 2010

⑮ Issei Sato, Kenichi Kurihara, Shu Tanaka, Seiji Miyashita and Hiroshi Nakagawa . Quantum Annealing for Variational Bayes Inference 査読有, The 25th Conference on Uncertainty in Artificial Intelligence (UAI2009) http://www.cs.mcgill.ca/~uai2009/proceedings.html. 2009

⑯ 吉田 稔, 中川 裕志 , 渋谷 久恵, 前田 俊二 : テキストマイニングによる機器異常診断支援の試み: 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（第 10 回日本データベース学会年次大会）F5-4, 2012 年 3 月.

⑰ 江原遥, 佐藤一誠, 中川裕志:ブートストラップ法のための能動学習. 言語処理学会第 18 回年次大会. F1-1. 2012 年 3 月.

⑱ 谷田和章, 荒牧英治, 佐藤一誠, 吉田稔, 中川裕志: ソーシャルメディアによる風邪流行の予測. 言語処理学会第 18 回年次大会. A3-5. 2012 年 3 月.

⑲ 谷田和章, 荒牧英治, 佐藤一誠, 吉田稔, 中川裕志: ソーシャルメディアを用いた風邪薬販売量の予測, 言語処理学会第 18 回年次大会, 広島, 2012 年 3 月.

⑳ 谷田和章, 荒牧英治, 佐藤一誠, 吉田稔, 中川裕志: Twitter による風邪流行の推測, 人工知能学会情報編纂研究会第 6 回研究会, 東京, 2011 年 10 月.

㉑Minoru Yoshida, Hiroshi Nakagawa: Web People Search: Person Name Disambiguation and Other Problems, Tutorial of The 2nd Asian Conference on Machine Learning (ACML2010), Tokyo, Japan, Nov.8, 2010

㉒佐藤 一誠, 中川裕志. Latent Dirichlet Allocation における決定論的オンラインベイズ学習. 情報処理学会自然言語処理研究会. 2009-NL-193. 2009

㉓Masaki Ikeda, Shingo Ono, Issei Sato, Minoru Yoshida and Hiroshi Nakagawa. Person Name Disambiguation on the Web by TwoStage Clustering. 2nd Web People Search Evaluation Workshop (WePS 2009), 18th WWW Conference, Madrid, Spain, April 21. 2009

㉔野田 陽平, 清田 陽司, 中川 裕志: Wikipediaからの意外性のある情報の抽出, NLP若手の会 第4回シンポジウム, 京都大学 .2009

〔図書〕（計１件）

① 中川裕志： 情報法,（宇賀克也, 長谷部恭男 編：第 8 章 データベースサービスとコンテンツ）, pp.133-159, 有斐閣, 2012 年 9 月

〔産業財産権〕
○出願状況（計 0 件）
○取得状況（計 0 件）

〔その他〕
ホームページ 公表論文リスト掲載
http://www.r.dl.itc.u-tokyo.ac.jp/node/46/

６．研究組織
(1)研究代表者
中川裕志（Nakagawa Hiroshi）
東京大学・情報基盤センター・教授
研究者番号：20134893

(2)研究分担者
吉田稔（Yoshida Minoru）
東京大学・情報基盤センター・助教
研究者番号：40361688

清田陽司（Kiyota Youji）
東京大学・情報基盤センター・助教
研究者番号：10401316

佐藤一誠（Sato Issei）
東京大学・情報基盤センター・助教
研究者番号：90610155

(3)連携研究者
二宮崇（Ninomiya Takashi）
東京大学・情報基盤センター・講師
研究者番号：20444094