

履歴書

萩原 正人 (HAGIWARA, Masato)

更新: 2010年4月21日

連絡先	〒153-0043 東京都目黒区東山 1-30-16 長谷川コーポ 202号室	携帯: 090-6806-3457 E-mail: hagnosis@gmail.com
学歴	2006年4月 - 2009年3月: 名古屋大学 大学院 情報科学研究科 博士課程後期課程 情報システム学専攻ソフトウェア論講座 博士論文: 「Modeling and Selection of Context for Better Synonym Acquisition」 (日本語訳: 類義語獲得の性能向上のための文脈のモデル化と選択)	
	2004年4月 - 2006年3月: 名古屋大学 大学院 情報科学研究科 博士課程前期課程 情報システム学専攻 ソフトウェア論講座 * 飛び級制度により進学 GPA: 3.8 修士 (情報科学) 修士論文: 「Utilization of Probabilistic Latent Semantics for Automatic Thesaurus Construction」 (日本語訳: シソーラス自動構築のための確率的潜在意味の利用)	
	2001年4月 - 2004年3月: 名古屋大学 工学部 電気電子・情報工学科 専門課程 GPA: 3.9	
	2001年3月: 愛知県立国府高等学校卒業	
研究経験	1. 大規模コーパスからの語彙知識獲得 自然言語処理において、同義語や、上位・下位関係などの語彙知識を大規模コーパスから自動獲得することは重要な問題で、幅広い応用範囲があります。語彙の関連度を測定する指標として、文脈の類似度を利用した「分布類似度」が広く用いられてきましたが、この分布類似度を用いる文脈情報の抽出、選択、拡張に注目し、いくつかの拡張および選択手法を提案しました。WordNet 等、既存のシソーラスを用いた評価により、本手法の有効性を確認しました。さらに、確率的潜在意味モデル(PLSA)等を適用することにより、スパースネス等の問題を解決し、性能の向上を図りました。 また、文脈パターンと機械学習を組み合わせた語彙知識獲得にも取り組んでいます。最近の成果では、ベクトル空間モデルや分布類似度を用いた従来手法と比較し、50%以上もの性能向上を実現しました。さらに、グラフカーネルを用いて、日本語の非分かち書き文から、意味カテゴリ (同じ意味クラスに属する名詞の集合) を自動的に抽出できることを示しました。	
	2. 日本語クエリ書き換え Web 検索エンジンにおいて、クエリの訂正技術はロバストな検索を実現する上で重要です。Microsoft Research において取り組んだインターンのプロジェクトでは、日本語のクエリ書き換えのための統一的アプローチを提案・実装しました。本手法は、基本的には従来の英語スペル訂正手法に基づいていますが、(1) 表記・綴りの上では異なるが意味的に類似した候補への書き換えを扱うことができ、(2) 意味的類似度に基づくカーネル法を用いて、書き換え候補間の類似度を計算する際のスパースネスを軽減できるという点に新規性があります。また、書き換えモデルに対する訓練およびテストセットを検索セッションログから効率的に生成する手法を提案しました。クエリ書き換えタスクにおいて、提案手法は英語のクエリ訂正に関する従来手法と比較して優れた性能を示しました。	
技術スキル	言語: C, C++, C#, Python, Ruby, (D)HTML, JavaScript OS: Windows, Linux LAMP アーキテクチャに基づいた Web アプリケーション経験 (5年以上)	
語学スキル	日本語: ネイティブ 英語: 会話・読み書き上級 TOEIC 960点 (2007年取得) 中国語 (標準語): 会話・読み書き中級 (2009年 中国語検定3級取得)	

受賞・研究貢献

- * 言語処理学会第 16 回年次大会(NLP2009) 最優秀発表賞 受賞 発表:「グラフカーネルに基づく非分かち書き文からの意味的語彙カテゴリの抽出」
- * NLP 若手の会 第 3 回シンポジウム 奨励賞受賞
発表:「意味的類似度を利用した日本語クエリ書き換えのための統一的アプローチ」
- * 21 世紀 COE プログラム「社会情報基盤のための音声・映像の知的統合」 第 22 回 IMI セミナー 優秀賞受賞 発表:「シソーラス自動構築における確率的潜在意味の利用」
- * ACL-IJCNLP 2009 (Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing) Student Research Workshop プログラム委員
- * その他, 国内論文誌の査読経験あり.

その他 特筆事項

- * 名古屋大学大学院情報科学研究科へ成績優秀のため飛び級入学し, 当該研究科において初の飛び級事例となりました. 学部 3 年次に卒業論文相当の研究に取り組みました (題目:「クロスリンガル情報検索のための多言語シソーラス自動構築に関する研究」)
- * 大学院の学生を中心とした英語発表・プレゼンテーションについての勉強会を発足・運営し, 週 1 回, 各専門分野の議論を通じて英語力の向上を図りました (1 年間).
- * 大学における留学生のメンターに選ばれ, 言語・文化・専門分野の面で指導およびサポートを行いました (2 年間).
- * 日本語・英語間の技術翻訳実務 (生物医学・歯学・特許翻訳等の分野) に従事しました (3 年間).

興味・趣味

コンピュータによる作曲・バンドでの楽器演奏
読書・語学 (プログラミング言語・自然言語)

職歴

2009年4月 - (現在): **RD(エンジニア) - バイドゥ株式会社** (東京 / 上海オフィス勤務)

- * バイドゥモバイル検索 (<http://m.baidu.jp/>)のランキング・アルゴリズムの向上, ページ解析技術の開発, および機械学習を用いたスパム・フィルタ・モジュールの開発を担当しました.
- * モバイル検索の絵文字検索および自然言語処理技術を用いた絵文字意味解析を開発しました. この成果については NLP2010 にて発表済みです (業績リスト参照). また, 絵文字の意味解析およびインデックス技術について特許出願中です.
- * BaiduType (日本語仮名漢字変換(IME) <http://type.baidu.jp/>) の技術監督を担当しました.
- * 日本語形態素・構文解析, 同義語マイニング, 固有名詞資源整備など, その他様々な自然言語処理トピックに取り組み, 検索エンジンの質の向上に貢献しました.

2008年4-7月: **リサーチ・インターン - Microsoft Research (米国ワシントン州)** (メンター: 鈴木久美氏)

- * 検索クエリの誤り訂正・異表記や翻字の正規化を行う高性能な日本語のクエリ書き換え手法を提案し, 従来手法に対する有意性を確認しました.
- * 数10GBの検索ログを扱うクエリ書き換えシステムを Visual C#, SQL Server, Ruby を用いて実装しました. 本システムは, Microsoft の Live Search (<http://www.live.com/>) に近々組み込まれる予定です.
- * クエリ書き換えペアをセッションログから効率的かつ自動的に生成する手法を開発しました.
- * 本プロジェクトの結果を第3回 NLP 若手の会にて発表し, 奨励賞を受賞しました. さらに, 国際会議 NAACL HLT 2009 に論文が採択されました.

2006年11月 - 2007年8月: **開発エンジニア - 情報処理推進機構 (IPA) 2006年度下期未踏ソフトウェア創造事業** (PM: David J. Farber 教授)

- * 未踏ソフトウェア創造事業に, GPS 付き情報端末を対象とした位置情報適応型メタ SNS "Serendi: A Location-Aware Social Networking Platform" (<http://serendi.org/>) が採択されました.
- * 自然言語処理およびネットワーク解析技術を用いて, ユーザー間の「Compatibility (相性)」を解析するモジュールを開発しました. PHP, JavaScript, Ruby, MySQL, ActiveRecord を使用しました.
- * 同時参加人数50人以上の大規模なユーザーテストを実施し, 本システムの信頼性を確認しました.

2006年4月 - 2007年3月: **リサーチ・アシスタント - 名古屋大学**

- * 名古屋大学において採択された21世紀COEプログラム「社会情報基盤のための音声・映像の知的統合」に関連するプロジェクトの研究開発に従事しました.
- * シソーラスをはじめとする言語資源構築の性能改善のため, 語彙類似度計算に用いる文脈の拡張・選択手法をいくつか提案・実装しました.
- * ここでの研究成果は, トップレベルの国際会議および学術雑誌に数多く採録されました (主要論文参照).

2005年9月 - 2006年3月, 2006年9月 - 2007年3月: **ティーチング・アシスタント - 名古屋大学**

- * 「線形代数および演習」と「線形代数および演習」の講義を担当しました.

2005年8-9月: **ソフトウェア・エンジニア・インターン - Google Inc. (米国カリフォルニア州)**

(メンター: Dekang Lin 氏・Jun Wu 氏)

- * Google 本社における2ヶ月間のインターンシップに参加しました. 当時インターン制度が始まって1年しか経過しておらず, 数少ない日本人インターンとして選ばれたうちの一人です.
- * 日本語クエリ・サジェスションのプロジェクトに取り組みました. この時の成果は, 現在の Google 検索結果の画面上下部に表示されるクエリ・サジェスションの基礎技術として採用されています.
- * MapReduce など Google 社の保有する並列計算アルゴリズムおよび大規模クラスターなどのインフラを駆使して開発を行いました.

その他
開発経験

frippa (<http://www.frippa.com/>)

* 地域型個人間取引仲介・案内・3行広告 Web サービス frippa のシステム全体の開発を担当しました。現在約 2,000 人のユーザーが参加しており、活発な個人間取引サイトへと成長しました。

* Linux, MySQL, ActiveRecord, Ruby 等をベースにした独自 MVC フレームワーク上で動作しています。

* 自然言語処理に基づいた類似商品推薦機能を実装しました。

* 2007 年度名古屋大学下宿用品リユース市との共同プロジェクトにより登録物品データベースを技術提供するなどの社会貢献実績もあります。

その他に、RINEN.inc (<http://rinen.cc/>), Anchor (<http://anchor.vc/>) など IT 系ベンチャー企業においてアルバイトとして開発経験があり、Ajax や Flash 等を駆使したユーザーインターフェースを担当しました。

- 論文誌 (査読あり)
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Supervised Synonym Acquisition Using Distributional Features and Syntactic Patterns. *Journal of Natural Language Processing*, Vol. 16, Num. 2, pp. 59-83, 2009.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. A Comparative Study on Effective Context Selection for Distributional Similarity. *Journal of Natural Language Processing*, Vol. 5, Num. 5, pp. 119-150, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Effective Use of Indirect Dependency for Distributional Similarity. *Journal of Natural Language Processing*, Vol. 15, Num. 4, pp. 19-42, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Bootstrapping-based Extraction of Dictionary Terms from Unsegmented Legal Text. *New Frontiers in Artificial Intelligence: JSAI 2008 Conference and Workshops, Revised Selected papers, Lecture Notes in Computer Science*, Vol. 5447, pp. 213-227, 2009.
- 国際会議・ワークショップ (査読あり)
- Masato Hagiwara and Hisami Suzuki. Japanese Query Alteration Based on Lexical Semantic Similarity. *Proc. of NAACL HLT 2009*, pp. 191-199, 2009.
- Nobuyuki Shimizu, Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama, Hiroshi Nakagawa. Metric Learning for Synonym Acquisition, *Proc. of COLING 2008*, pp. 793-800, 2008.
- Masato Hagiwara. A Supervised Learning Approach to Automatic Synonym Identification based on Distributional Features. *Proc. of ACL 2008 Student Research Workshop*, pp. 1-6, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Bootstrapping-based Extraction of Dictionary Terms from Unsegmented Legal Text. *Proc. of JURISIN 2008*, pp. 63-72, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Context Feature Selection for Distributional Similarity. *Proc. of IJCNLP 2008*, pp. 553-560, 2008.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Effective Proximity Distance for Word-Based Context. *Proc. of SNLP 2007*, pp. 105 - 110, 2007.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Effectiveness of Indirect Dependency for Automatic Synonym Acquisition. *Proc. of CoSMo 2007*, pp. 1 - 8, 2007.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. Selection of Effective Contextual Information for Automatic Synonym Acquisition. *Proc. of COLING/ACL 2006*, pp. 353 - 360, 2006.
- Masato Hagiwara, Yasuhiro Ogawa, Katsuhiko Toyama. PLSI Utilization for Automatic Thesaurus Construction. *Proc. of IJCNLP 2005*, pp. 334 - 345, 2005.
- 国内会議・ワークショップ (査読あり)
- 萩原正人, 小川泰弘, 外山勝彦. 類義語自動獲得における間接依存関係の有効性. 言語処理学会第13回年次大会ワークショップ「言語的オントロジーの構築・連携・利用」論文集 pp. 43-46, 2007.
- その他国内会議・ワークショップ (査読なし)
- 萩原正人, 水野貴明. モバイル検索システムのための絵文字に対する意味解析. 言語処理学会第16回年次大会予稿集, PA2-24, 2010.
- 萩原正人, 鈴木久美. 意味的類似度を利用した日本語クエリ書き換えのための統一的アプローチ. NLP 若手の会 第3回シンポジウム, 発表 20, 2008.
- 萩原正人, 小川泰弘, 外山勝彦. 分布類似度のための文脈素性選択. NLP 若手の会 第2回シンポジウム, 発表 11, 2007.
- 萩原正人, 小川泰弘, 外山勝彦. perplexity を用いた類義語獲得の自動評価, 言語処理学会第12回年次大会予稿集, B4-4 (pp. 767-770), 2006.
- 萩原正人, 小川泰弘, 外山勝彦. 文脈として有効な近接語の距離. 2006年度電気関係学会東海支部連合大会講演論文集, O-394, 2006.
- 萩原正人, 小川泰弘, 外山勝彦. シソーラス自動構築における PLSI の利用. 情報処理学会 第166回自然言語処理研究会, 2005-NL-166, pp.71-78, 2005.
- 萩原正人, 小川泰弘, 外山勝彦. 索引語ペアを用いた多言語シソーラスの自動構築. 第3回情報技術フォーラム(FIT2004), pp. 209-212, 2004.