

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成25年 5月20日現在

機関番号：11301

研究種目：若手研究（B）

研究期間：2011～2012

課題番号：23700157

研究課題名（和文）自然論理と確率論理の融合による文間関係認識

研究課題名（英文）Unifying Natural Logic and Probabilistic Logic for Recognizing Textual Relations

研究代表者

渡邊 陽太郎 (WATANABE YOTARO)

東北大学・大学院情報科学研究科・助教

研究者番号：70583326

研究成果の概要（和文）：自然言語上での推論を実現する自然論理（Natural Logic）と、データから有用な特徴を自動学習する枠組である確率的手法である対数線形識別モデルを融合した文間関係認識手法を確立した。Natural Logicによる文間関係の推論の過程を、統計的手法によるアライメントおよび確率モデル上の推論のプロセスと捉え、文間の対応付けと文間関係の推論を単一のモデルで表現することによって、提案手法は、言語に内在する言語に関する規則を精緻に捉え、データからの自動学習による頑健さの双方を備えることを可能とした。複数のデータセットによる評価を実施し、提案手法の有効性を実証した。

研究成果の概要（英文）：The principal investigator established a novel probabilistic model for recognizing textual relations that unifies Natural Logic, a logic over natural language, and a statistical probabilistic model. By modeling an inference process of a textual relation as alignment between units in sentences and inference over a factor graph, it makes possible to precisely handle several linguistic phenomena in natural languages, and provide robustness by automatically learning linguistic relations from a set of datasets. Evaluation over multiple datasets had been conducted, and the effectiveness of the model had been proved.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
交付決定額	3,200,000	960,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：含意関係認識 自然論理 アライメント

1. 研究開始当初の背景

二文間の意味的關係を認識する文間関係認識は、文間の表現の違いを吸収しつつ、それらの同義性、矛盾性を判定する問題であり、情報検索、質問応答、情報抽出、文書要約、機械翻訳など、様々な自然言語処理の応用分野において極めて重要な技術である。文間関係認識に関する既存研究の多くは、文間での対応箇所（単語、係り受けなど）の類似性によって文間の意味的關係を認識するという単

純な手法を用いてきた。しかし、文間関係認識の問題は、語彙レベルでの意味関係、量化詞、文脈に依存する含意関係の向きの変化、事態の事実性、事態間関係、数値表現など、様々な言語現象を捉える必要のある複合的な問題であるため、単純な類似度だけではこれらの言語現象の影響を精緻に捉え、文間の意味關係を適切に導くことは困難である。多様な言語現象を精緻に捉え、精緻かつ頑健に文間関係認識を実現する枠組が、文間関係認識技術の向上、および前

述の多様なアプリケーションの発展のためには不可欠である。

2. 研究の目的

本研究では、認知言語学を背景とした理論である Natural Logic (Lakoff, 1970) に基づく、精緻かつ頑健な文間関係認識技術を実現することを目的とする。まず、二文間での対応箇所(単語、係り受けなど)の意味関係(同義、含意、排他など)を結合する規則を網羅的に把握することにより、精緻な文間関係認識のための基盤を確立する。そして、意味関係の結合規則の逸脱に対する頑健性を高め、さらには不確実性を考慮した言語・世界知識の活用ができるよう、統計的機械学習の枠組を導入することによって、頑健な文間関係認識を実現する。

3. 研究の方法

自然論理を導入した精緻な文間関係認識を実現するために、まずは先行研究である (MacCartney et al., 2008) の Natural Logic に基づく文間関係認識手法において用いられている、二文間の対応箇所(単語や依存構造等)の意味的關係を結合する規則を拡充し、文間関係の説明が可能な事例のカバレッジを上げることによって、広範な言語現象に対応可能な文間関係認識基盤を確立する。次に、規則に基づく推論の頑健性を確保するため、統計的機械学習手法を導入した文間関係認識モデルを設計、最適な結合順序で意味的關係を推論するアルゴリズムと、モデルのパラメータ推定手法の開発をおこなう。文間関係認識のためのデータセットを用いた評価実験、および大規模文書集合を用いたオープンメイン文間関係認識の実証実験を行い、分析とモデルの改善のプロセスを繰り返すことによりシステムを実用レベルまで引き上げる。

4. 研究成果

(1) 文間関係認識のためのデータセットの構築

様々な言語現象を含む文間関係認識のためのデータセットを構築した。日本語の文間関係認識データは、NTCIR-9 RITE (Shima et al. 2011) など既に構築されているものの、含まれている言語現象が Natural Logic で研究されている言語現象の範囲を超えているものがかなり含まれていることや、個々の言語現象が該当する事例数が限定されているなどの問題があり、Natural Logic による文間関係認識手法の評価用データとしては適さなかった。そこで、言語現象としていくつか

のクラスを設定し、それぞれのクラスに該当する事例を作成していくことでデータセットを構築した。

データセットに含める言語現象は、英語の含意関係認識の先行研究 (Cooper et al., 1996, Bentivogli et al, 2010, MacCartney 2009) において議論されているカテゴリに基づき選択した。具体的には、語彙の意味関係(同義、反義、上位-下位)、量化子、修飾、否定、並列、関係節、同格、時間表現、数量表現、能動態/受動態、叙実動詞、機能的関係などである。構築したデータセットに含まれる各言語現象の分布は下記の通りである。

言語現象	数
量化子	182
修飾	55
数量/時間	53
語彙の意味関係	44
含意/叙実動詞	36
エンティティやイベント間関係	27
並列	27
項のミスマッチ	26
否定+語彙の意味関係	23
言い換え	21
統語的關係	19
述語のミスマッチ	17
部分-全体	13
条件	8
同格	7
世界知識	3
その他	37
計	598

上記のデータは、作業員1名により開発した。まず、Wikipedia等の記事を利用して一方の文を作成し、もう一方の文は、対象のカテゴリに該当する文を手で作文するという方法で文対を作成している。

(2) Natural Logic と統計的確率モデルを融合した文間関係認識手法の確立

(MacCartney et al, 2008, 2009) において提案されている、Natural Logic に基づく意味関係および意味関係の演算規則を、統計的機械学習の枠組の一種である対数線形識別モデルに導入することによって、文の要素間の対応関係と文間意味関係を同時に認識する手法を確立した。

研究代表者が開発した文間関係認識モデルの全体像を図1に示す。提案手法は、MacCartney らの手法と同様、一方のテキスト(T)から他方のテキスト(H)への編集操作に対して意味関係を与える。編集操作は、置換(Substitution)、削除(Deletion)、挿入

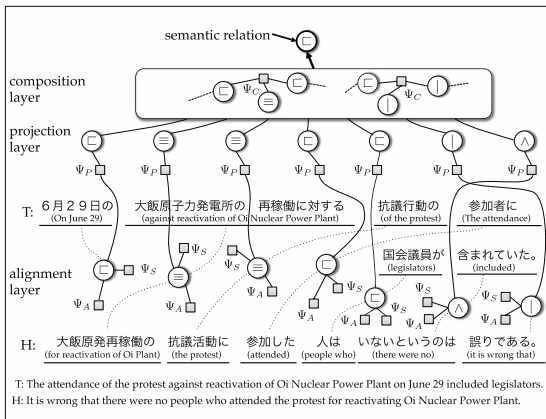


図 1: Natural Logic と統計的確率モデルを融合した文間関係認識モデル

(Insertion) の 3 種類からなり、文間の編集操作の集合によって文間のアライメントが表現される。それぞれの編集操作には、Natural Logic にて定義されている 7 種類の意味関係のいずれかが与えられる。文間の意味関係は、アライメントの局所的な意味関係から、文脈を考慮した意味関係の射影規則、および意味関係の結合規則を用いて、構成的に導かれる。

アライメントとその意味関係、意味関係の導出過程で導かれる意味関係は提案モデルでは隠れ変数として扱われる。開発したモデルの学習には、隠れ変数を持つ対数線形識別モデルの最適化手法の一つである周辺尤度最大化を適用することで、Natural Logic の意味関係と整合するアライメントの特徴が自動学習される。対数線形識別モデルの学習時には、文の要素間のあらゆるアライメントを考慮する必要があるが、この計算は実行可能ではないため、提案手法では、(MacCartney et al, 2008) によって提案された MANLI アルゴリズムを、意味関係を含めた編集を扱えるよう拡張し、N-best 解のみを用いて学習するような工夫を施すことによって、モデルの学習を実現した。

評価実験には、(A) 前述の研究代表者が開発したデータセットと、(B) 含意関係認識のタスクである NTCIR-10 RITE-2 のデータセットを用いた。(A) のデータセットを用いた評価実験では、正解のアライメントを教師あり学習し、文間関係の推論をおこなうベースライン手法と比較し、提案手法で実現した、文間関係からのアライメントの推定の有効性を確認した。また、(B) のデータセットを用いた実験では、RITE-2 タスクに参加した日本語文間関係認識システムの中で、認識が最も難しい「矛盾」の関係について最も高い性能を達成した。

以上の成果は、言語処理分野の主要な国際会議である COLING 2012 に採録されている。ま

た、関連する文間関係認識技術に関して、情報処理学会論文誌および国際ジャーナルである ACM Transactions on Asian Language Information Processing (TALIP) に採録され、その有用性が認められている。

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 2 件)

① Yotaro Watanabe, Junta Mizuno, Eric Nichols, Katsuma Narisawa, Keita Nabeshima, Naoaki Okazaki and Kentaro Inui. Leveraging Diverse Lexical Resources for Textual Entailment Recognition. ACM Transactions on Asian Language Information Processing (TALIP) - Special Issue on RITE, Volume 11 Issue 4, Article No. 18, December 2012. (査読あり)
<http://dl.acm.org/citation.cfm?id=2382600>
 DOI: 10.1145/2382593.2382600

② 水野 淳太, 渡邊 陽太郎, Eric Nichols, 村上 浩司, 乾 健太郎, 松本 裕治. 文間関係認識に基づく賛成・反対意見の俯瞰. 情報処理学会論文誌, Vol.52, No.12, pp.3408-3422, December 2011. (査読あり)
<http://id.nii.ac.jp/1001/00079546/>

[学会発表] (計 13 件)

① Katsuma Narisawa, Yotaro Watanabe, Junta Mizuno, Naoaki Okazaki and Kentaro Inui. Is a 204 cm Man Tall or Small? Acquisition of Numerical Common Sense from the Web. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). 2013 年 8 月 (発表日未定), ソフィア, ブルガリア. (査読あり)

② Yotaro Watanabe, Junta Mizuno and Kentaro Inui. THK's Natural Logic-based Compositional Textual Entailment Model at NTCIR-10 RITE-2. In Proceedings of the 10th NTCIR Conference. 2013 年 6 月 19 日. 国立情報学研究所, 東京. (査読なし)

③ Kazeto Yamamoto, Naoya Inoue, Yotaro Watanabe, Naoaki Okazaki and Kentaro Inui. Discriminative Learning of First-order Weighted Abduction from Partial Discourse Explanations. In Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013). 2013 年 3 月

28日, サモス, ギリシャ. (査読あり)

④ 大西真輝, 水野淳太, 福原裕一, 渡邊陽太郎, 乾健太郎. 文節間限定関係に基づく文間弱対立関係認識. 言語処理学会第19回年次大会発表論文集. 2013年3月15日, 名古屋大学. (査読なし)

⑤ 成澤克麻, 渡邊陽太郎, 水野淳太, 岡崎直観, 乾健太郎. 数量の大小の自動判定: 「彼は身長が2mある」は高いか低いかな. 言語処理学会第19回年次大会発表論文集. 2013年3月14日, 名古屋大学. (査読なし)

⑥ Junta Mizuno, Eric Nichols, Yotaro Watanabe and Kentaro Inui. Organizing Information on the Web through Agreement-Conflict Relation Classification. In Proceedings of the Eighth Asia Information Retrieval Societies Conference (AIRS 2012). 2012年12月17日, 天津, 中国. (査読あり)

⑦ Yotaro Watanabe, Junta Mizuno, Eric Nichols, Naoaki Okazaki and Kentaro Inui. A Latent Discriminative Model for Compositional Entailment Relation Recognition Using Natural Logic. In Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012), pages 2805-2820. 2012年12月11日, ムンバイ, インド. (査読あり)

⑧ Naoya Inoue, Kazeto Yamamoto, Yotaro Watanabe, Naoaki Okazaki and Kentaro Inui. Online Large-margin Weight Learning for First-order Logic-based Abduction. In Proceedings of the 15th Information-Based Induction Sciences Workshop. 2012年11月7日, 筑波大学東京キャンパス文京校舎. (査読なし)

⑨ 渡邊陽太郎, 水野淳太, 岡崎直観, 乾健太郎. Natural Logic と条件付確率場の融合による構成的文間関係認識. 情報処理学会 自然言語処理研究会 研究報告 (2012-NL-208). 2012年9月2日, 東北大学. (査読なし)

⑩ 山本風人, 井之上直也, 渡邊陽太郎, 岡崎直観, 乾健太郎. 誤差逆伝播を利用した重み付き仮説推論の教師あり学習. 情報処理学会 自然言語処理研究会 研究報告 (2012-NL-206). 2012年5月11日, 東京工業大学. (査読なし)

⑪ 成澤克麻, 渡邊陽太郎, 水野淳太, 岡崎直観, 乾健太郎. 数量表現を伴う文における含意関係認識の課題分析. 言語処理学会第18回年次大会発表論文集. 2012年3月16日, 広島市立大学. (査読なし)

⑫ 渡邊陽太郎, 水野淳太, 岡崎直観, 乾健太郎. 隠れ変数を持つ識別モデルによる文間意味関係の学習. 言語処理学会第18回年次大会発表論文集. 2012年3月16日, 広島市立大学. (査読なし)

⑬ Yotaro Watanabe, Junta Mizuno, Eric Nichols, Katsuma Narisawa, Keita Nabeshima and Kentaro Inui. TU Group at NTCIR9-RITE: Leveraging Diverse Lexical Resources for Recognizing Textual Entailment. In Proceedings of the 9th NTCIR Workshop Meeting on Evaluation of Information Access Technologies (NTCIR-9). 2011年12月8日, 国立情報学研究所. (査読なし)

6. 研究組織

(1) 研究代表者

渡邊 陽太郎 (WATANABE YOTARO)
東北大学・大学院情報科学研究科・助教
研究者番号: 70583326

(2) 研究分担者

なし

(3) 連携研究者

なし