

視線情報を利用した日本語述語項構造解析

牧 諒亮 西川 仁 徳永 健伸

東京工業大学 大学院情報理工学研究科

maki.r.aa@m.titech.ac.jp

{hitoshi, take}@cs.titech.ac.jp

1 はじめに

本稿では、統計的学習手法を利用した日本語述語項構造解析に、視線情報を取り入れる手法を提案する。

日本語述語項構造解析では、語彙情報・統語情報ははじめとした種々の素性が取り入れられ、精度の向上が試みられてきた。しかしながらその精度は、タスクの難しさゆえ未だ十分ではなく研究の余地がある。

既存の述語項構造解析では述語の項のみが正例として扱われ、それ以外の候補は一律に負例として扱われてきた。もし項候補の「項らしさ」を判断する指標を与えることができれば、項になりやすい候補とそうでない候補との間で重みに差をつけることが可能となる。正解以外の候補の中にも、潜在的に項になりうるものと項になりそうにないものがあるため、それらを一律に扱わないことが精度の向上に寄与するのではないかと、というのが本研究の意図である。

「項らしさ」の指標を実現するため我々が提案するのが、視線情報を用いた述語項構造解析である。述語項構造アノテーション時に視線を利用した研究 [16] は存在するが、視線情報を利用した述語項構造解析は既存研究が存在せず、その有用性を検証する必要がある。

本稿では、アノテーション作業者のアノテーション作業中の視線を収集し、作業時に作業者がよく見ている候補を「項らしい」候補であると仮定し、停留時間の長い順、停留回数の多い順に順位づけをしたうえで学習をおこなった。その評価のためベースラインとなる従来手法の分類器と、述語・項候補間の距離が短い順に順位づけされたデータを学習した解析器を用意した。結果として、従来手法のベースラインに比べて提案手法が有用であることと、項らしさの判断基準については改善が必要であることがわかった。

2 関連研究

ここでは述語項構造に関わる研究として、日本語でのアノテーションの仕様に関わるもの、素性設計に関

わるもの、解析の手法に関わるもの、視線情報の収集に関わるものを取りあげる。

近年では正しい項が付与されたコーパスを訓練用データとして学習をおこなう手法が主流となっており、日本語の述語項構造解析でも大規模な訓練データが必要とされている。その仕様については飯田ら [8]、松林ら [13] による議論があり、適切な仕様は利用するタスクによっても異なる。実際に述語項構造をアノテーションをしたコーパスとしては、NAIST テキストコーパス [8] や BCCWJDepParaPAS [3,6,7,9,10] などがある。

統計的な日本語述語項構造解析に利用する素性としては、英語文解析における基本的な素性に加え、日本語文のゼロ照応解析にも利用できるような素性が提案されてきた [1]。種々の素性が提案されるなか、有効な素性設計を検討した研究としては松林ら [12] がある。

述語項構造の解析モデルとしては、トーナメントモデル [11]、最大エントロピー法を用いたモデル [1]、決定リストを用いた手法 [5] などが提案されている。特に林部ら [11] は相対的な項らしさを判断するモデルを提案しており、この点で学習時に項らしさの概念を導入する本研究に類似している。

視線情報を述語項構造解析に利用した研究は存在しないが、光田らによって、述語項構造アノテーション時の視線データを収集しアノテーション作業者に対する信頼性を推定する研究 [15,16] がなされている。本研究では光田らによる実験設定を利用して視線データを収集し、より正確にパラメータ推定をおこなうことが目標である。

3 タスク設定

本研究では、NAIST テキストコーパス (NTC) のアノテーション基準 [8] を簡略化しておこなわれた光田ら [16] の設定を踏襲している。

項が述語の前方に出現しているものの中から述語¹を1つ選択し、アノテーション対象の述語とする。次

¹ここでの「述語」にはサ変名詞等、事態性名詞も含まれている。

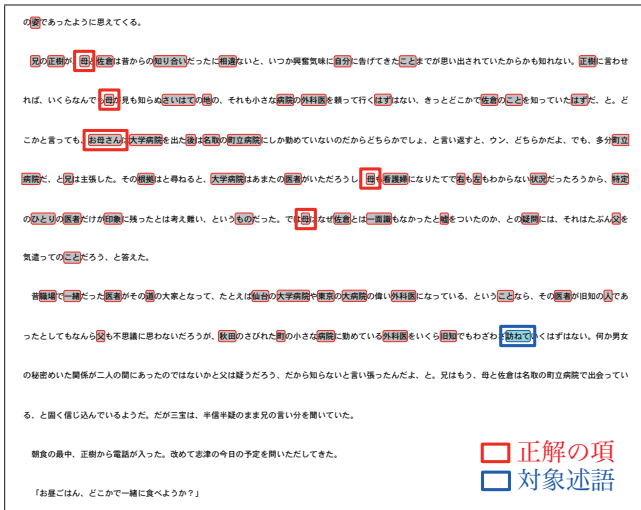


図 1: アノテーション作業画面 (1 課題)

に述語に先行する項候補を列挙し、述語と項候補のペアの集合を入力とする。ペアの集合から 1 つを出力し、その項候補が実際に項であれば正解とする。

なお、今回収集した視線データはガ格のみであるので、以下における「項」はガ格を指していると考えて差し支えない。

4 提案手法

述語項構造アノテーション時に作業者の視線データを収集し、停留回数と停留の合計時間（停留時間）を各項候補ごとに計上する。これらをキーにすることにより項候補に順位づけをおこなうことが可能になる。本稿においては、まず停留回数の多い順で並べ、同順位の場合は停留時間の長い順で並べ、それでも同順位の場合は、述語・項候補間の距離が近い順に並べることによって各候補を順位づけした。ここでおこなった順位づけを「停留順」と呼ぶことにする。

正解となる項候補を最上位に据え、残りを停留順に並べる。このときの順位に応じて各事例に重みを与え、ランキング SVM による機械学習を通じてパラメータ推定をおこなう。この手順を利用すると、評価時には視線データを利用せずに項らしきスコアを点推定²することが可能となる。

5 視線データの収集

5.1 アノテーション実験

図 1 に示すのはアノテーションの作業画面であり、対象の述語は青枠で、ガ格候補は灰色の背景で示されている。視線データを収集する機器の精度を鑑み、作

²他の候補と独立にスコアを計算しているため。

業者が見ている行が判別できる程度の行間を設定し、さらにスクロールが起らないような文章長を設定している。画面に 1 つだけ提示された述語に対して、作業者がガ格の項を候補集合から択一することで 1 課題をこなしたことになる。

図 1 からわかるように、正解となる候補は必ずしも 1 つとは限らず、アノテーション作業者は正解となる項の集合のうち、いずれかを選択すれば正解となる。作業にはマウスを用い、正解となる項をクリックすることで作業が完了する。

作業対象の文章には現代日本語書き言葉均衡コーパス (BCCWJ) コアデータ内の出版サブコーパス・書籍レジスタ (PB) から抽出した 50 課題と、NTC から抽出した 50 課題の計 100 課題を用いた。作業時にはマウスの動作に加え、視線計測装置 Tobii T60 を利用して視線情報を記録した。データ収集のために 10 名のアノテーション作業者を雇用し、各作業者が同一の文章集合に対して作業をおこなった。

これらの設定は先行研究 [15] でおこなわれたデータ収集を踏襲したものであるため大部分では共通の設定をしているが、アノテーション課題の作成と、作業時の指示に関して異なる点がある。前者については格交替をとまなう述語を作業対象から除いている点で異なっている。後者については候補を選ぶ際に、迷った候補があれば各候補をすべてクリックしたうえで択一するよう、指示を変更した点で異なっている。

5.2 視線データの処理

Tobii T60 は一定の間隔で、作業者が見ている画面上の位置を記録している。この記録を利用し、距離閾値 D を 16 ピクセル、時間閾値 T を 100 ミリ秒として停留の抽出をおこなった。1 つの停留は、停留に含まれる視線データの重心からみて、いずれも距離閾値 D 以内に含まれる視線データの集合となっている³。

停留を検出した後に、どの行を読んでいるときの停留であるかを判定するため境界線を引き、さらに文字・項候補への対応づけをおこなっている。

以上の手続きにより各項候補について、停留回数と停留時間を得た。

6 評価実験

アノテーション対象の文書には BCCWJ コアデータ内の出版サブコーパス・書籍レジスタ (PB) を採用した。これらの文書から BCCWJ の長単位 (LUW) で項候補を抽出した。さらに BCCWJDepParaPAS [3, 6, 7, 9, 10]

³抽出アルゴリズムの詳細は、光田 [14] を参照のこと。

から述語項構造を取得し、末尾に直接タグが付与された語をもつ長単位を正解として扱った。アノテーションの対象とする述語は前方文脈にガ格項が存在するものに限定した。

以上の手続きにより1つの述語に対して、正解と項候補を列挙することができたことになる。述語と項候補の組を1事例として、各事例に素性を付与していく。

6.1 素性設計

評価のための素性として、松林ら [12] を参考に基本的な素性を選択した。本稿の目的は新たな素性を提案することではないため、既存研究で使われてきたものを利用している。

表1は今回利用した素性を一覧にしたものである。語彙情報の素性ラベルはBCCWJでの名称を記載した。

距離の素性はBCCWJの「連番」を利用し、述語と項候補の間にある語が増えるほど値が大きくなるよう設計している。

これらの素性を各事例に付与し、訓練用のデータセットと評価用のデータセットを作成する。

6.2 訓練用データセット

1つの述語に対して得られた正解と項候補の集合をまとめて1課題とカウントすることにする。訓練用のデータセットとして、47課題用意した。この47課題に関して、正解を含めた項候補の数は最大99、最低12、中央値が45である。このデータセットをもとに、3種類の訓練事例を作成した。

- **2値分類**：2値分類のための訓練事例。予測には2値分類器を用いる。正解のみに正例ラベルを付与し、残りは負例とした。
- **順位づけ (距離順)**：視線情報を利用せず順位づけした訓練事例。予測にはランキング予測器を用いる。正解を最上位に据え、述語から見て近い順に残りの事例を並べた。ランキングを利用することそのものの効果を検証するため、停留順とは別に、文書から取得可能な順序の例として距離を取りあげ、これをもう1つのベースラインとして用意した。
- **順位づけ (停留順)**：視線情報を利用して順位づけした訓練事例。予測にはランキング予測器を用いる。アノテーションごとに学習事例の順位づけをおこなった⁴。正解を最上位に据え、残りを停留回数・停留時間の大きい順に並べた。停留回数・停

⁴ 9名のデータを利用したため、順位づけ(距離順)とくらべて事例数が9倍になっている。

留時間が等しい候補は、距離順により順位づけをおこなった。これが提案する手法である。

視線データを利用するにあたって、視線情報がうまく取得できていなかった1名のデータを破棄し、9名分のデータを利用した。

6.3 評価用データセット

オープンテストのための評価用データセットとして、2,676課題を用意した。訓練用のデータと候補の数が著しく異なるように、候補数の範囲を訓練用データセットにあわせている。この2,676課題に関して、正解を含めた項候補の数は最大99、最低12、中央値が55である。なおこの評価用データセットには、視線データ収集時に対象から除いた、格交替をとまなう述語も課題として含まれている。この2,676課題をもとに、各候補に素性を付与して評価用の事例を作成した。

6.4 学習・評価

2値分類器の学習にはClassias [4] を利用した⁵。順位づけ(距離順)、順位づけ(停留順)の学習にはSVM^{rank} [2] を利用した。いずれの手法においても評価の際には、各候補についてガ格スコアを求め、候補集合の中で最も値の大きい候補が正解項と一致したときに、ガ格を正解したと判定している。

6.5 結果

学習後にクロズドテストとオープンテストをおこなった結果を表2に示す。オープンテストの評価に利用したデータセットは文内に正解がある課題も含まれているので、正解率を個別に算出している。

クロズドテストでは2値分類が最も正解率が高く、距離順が最も正解率が低くなる結果となった。一方オープンテストでは、全体の正解率で見ると距離順が最も高く、2値分類が最も低い結果となった。さらに内訳をみると、順位づけの手法では文内の正解率が高く、2値分類の手法では文間の正解率が高いことがわかる。

6.6 考察

まずクロズドテストの結果をみると、距離順による順位づけがガ格を全く予測できていないことが読み取れる。距離順による順位づけは、述語に近い項をきわめて強く選好するモデルであるため、述語から遠い文間ガ格を予測できなかったのだと思われる。

⁵ L2正則化項を加えた。学習時のパラメータの最適化はおこなっていない。

表 1: 利用した素性一覧

素性の区分	利用した素性
述語に関わる素性	語彙情報 (語彙素, 語種, 品詞, 活用形, 活用型, 原文文字列)
項候補に関わる素性	語彙情報 (語彙素, 語種, 品詞, 原文文字列), 後続格助詞
述語と項候補に関わる素性	述語と項候補の語彙素ペア, 距離, 文内/文間, 文内/文間と後続格助詞のペア

表 2: 評価の結果

	クローズドテスト	オープンテスト		
	文間	文内	文間	全体
2 値分類	63.8% (30/47)	1.3% (25/1,996)	23.8% (162/680)	7.0% (187/2,676)
順位づけ (距離順)	0.0% (0/47)	42.3% (847/1,996)	4.9% (33/680)	32.9% (880/2,676)
順位づけ (停留順)	34.0% (16/47)	28.0% (559/1,996)	2.9% (20/680)	21.6% (579/2,676)

次にオープンテストにおける文内ガ格について、停留順が2値分類を上回ったことに着目する。2値分類での精度が低かった理由については、文内ガ格に負例しか存在しなかったことが挙げられる。一方停留順については、最初に述語の近くを見るというアノテーション作業者の視線の傾向から「文内」素性が大きな重みを持ったため、文内ガ格をよりよく予測できなかったと考えられる。

同じくオープンテストの文内ガ格において、距離順が停留順を上回ったことにも着目したい。距離順での各素性の重みを調べたところ、格助詞「が」が後続する文内の候補より、格助詞「を」「に」が後続する文内の候補の方が、高いガ格スコアを得られることがわかった。一般に、ガ格よりもヲ格やニ格の方が述語の近くに出現しやすいため、距離順ではその傾向を学習したと考えられる。距離順で文内ガ格の精度が高かったのは、「を」「に」をともなわず文内に正解が現れる課題が多かったからではないかと推察される。

今回の順位づけでは、停留から数値化した作業者の内省を用いるのにとどまり、アノテーション作業者の最終的な判断や、迷った際に選択した候補などは順位づけに用いていない。「項らしさ」をいかに順位づけるかが今後の課題といえる。

7 おわりに

本稿では、日本語述語項構造解析について視線情報を用いた手法を提案し、その評価をおこなった。

今回利用した学習データは数・条件ともに限定的なものであったため、利用するデータ充実させたいうえで再度評価をおこなうことが望ましい。この点については今後の課題とする。

参考文献

- [1] Kenji Imamura, Kuniko Saito, and Tomoko Izumi. Discriminative Approach to Predicate-argument Structure Analysis with Zero-anaphora Resolution. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pp. 85–88, 2009.
- [2] Thorsten Joachims. Training Linear SVMs in Linear Time. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2006.
- [3] Kikuo MAEKAWA, Makoto YAMAZAKI, Toshinobu OGISO, Takehiko MARUYAMA, Hideki OGURA, Wakako KASHINO, Hanae KOISO, Masaya YAMAGUCHI, Makiro TANAKA, and Yasuharu DEN. Balanced corpus of contemporary written Japanese. *Language resources and evaluation*, Vol. 48, No. 2, pp. 345–371, 2014.
- [4] Naoaki Okazaki. Classias: a collection of machine-learning algorithms for classification, 2009. <http://www.chokkan.org/software/classias/>.
- [5] Hiroto Taira, Sanae Fujita, and Masaaki Nagata. A Japanese Predicate Argument Structure Analysis Using Decision Lists. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, pp. 523–532, 2008.
- [6] 浅原正幸, 松本裕治. 『現代日本語書き言葉均衡コーパス』に対する係り受け・並列構造アノテーション. 言語処理学会第19回年次大会発表論文集, pp. 66–69, 2013.
- [7] 浅原正幸, 大村舞, 松本裕治. BCCWJ-DepParaPAS: 『現代日本語書き言葉均衡コーパス』依存構造アノテーションの重ね合わせ (To Appear). 言語処理学会第22回年次大会発表論文集, 2016.
- [8] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕治. 述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から. 自然言語処理, Vol. 17, No. 2, pp. 25–50, 2010.
- [9] 植田禎子, 飯田龍, 浅原正幸, 松本裕治, 徳永健伸. 『現代日本語書き言葉均衡コーパス』に対する述語項構造アノテーション. 第8回コーパス日本語学ワークショップ予稿集, pp. 205–214. 国立国語研究所, 2015.
- [10] 小西光, 小山田由紀, 浅原正幸, 柏野和佳子, 前川喜久雄. BCCWJ 係り受け関係アノテーション付与のための文境界再認定. 第3回コーパス日本語学ワークショップ予稿集, pp. 135–142. 国立国語研究所, 2013.
- [11] 林部祐太, 小町守, 松本裕治. 述語と項の位置関係ごとの候補比較による日本語述語項構造解析. 自然言語処理, Vol. 21, No. 1, pp. 3–25, 2014.
- [12] 松林優一郎, 乾健太郎. 統計的日本語述語項構造解析のための素性設計再考. 言語処理学会第20回年次大会発表論文集, pp. 360–363, 2014.
- [13] 松林優一郎, 飯田龍, 笹野遼平, 横野光, 松吉俊, 藤田篤, 宮尾祐介, 乾健太郎. 日本語文章に対する述語項構造アノテーション仕様の考察. 自然言語処理, Vol. 21, No. 2, pp. 333–377, 2014.
- [14] 光田航. 作業者の振舞いを利用したテキストアノテーションの信頼性推定. 東京工業大学修士論文, 2015.
- [15] 光田航, 飯田龍, 徳永健伸. 単一述語項関係アノテーション課題における視線情報の収集と分析. 情報処理学会第217回自然言語処理研究会, pp. 1–8, 2014.
- [16] 光田航, 飯田龍, 徳永健伸. アノテーションとアノテーション作業者の信頼性推定. 言語処理学会第21回年次大会発表論文集, pp. 553–556, 2015.