

現実世界の時間・空間制約を用いた共参照解析の精度向上

村上優樹 鶴岡 慶雅

東京大学大学院 工学系研究科

{murakami, tsuruoka}@logos.t.u-tokyo.ac.jp

1 はじめに

共参照解析は自然言語の意味を理解する上で重要なタスクである。共参照解析の代表的な手法として Soonら [12] による各名詞句の形態的な特徴を用いて分類器を学習するものが挙げられる。これは名詞句が「男性名詞か女性名詞か」「単数か複数か」といった情報や、対応する名詞句が文中でどれほど離れているかなど文の表層情報を用いている。また固有名詞など世界知識が必要な問題に対応するために Rahmanら [10] は固有名詞間の関係性を特徴量に組み込んだ学習による手法を提案している。しかしこれらの手法は文の深い意味を捉えた上で共参照解析を行うわけではない。そこで本研究では文の表層的な情報だけで表現されない現実世界の制約を利用した人称代名詞に関する共参照解析の改善を試みる。

本研究は自然言語の意味を理解するにあたり、背後にある世界の常識、法則などの一部を適切にモデル化することを目指すものであるが、今回は人物移動と人称代名詞に関する共参照解析に着目する。具体的に固有名詞で表現される人物に関する空間・時間的な制約を用いて、既存のよく知られている Stanford Core NLP [6] の共参照解析結果の誤り検出を行い、その有用性を示す。

2 関連研究

2.1 制約を用いた自然言語理解に関する研究

自然言語の意味を考える上で、文の表層的な情報だけでなく、現実の要素をモデル化し文章理解を目指す研究を幾つか紹介する。Changら [5] は、テキストから3D空間を自動的に生成する研究を行っている。その際に3D空間における自然言語の曖昧性を解消するために大量の3Dモデルデータと対応するテキストデータを合わせて学習を行うことで、テキストで表現される3D空間の定量的な意味を捉えようとしている。Seoらは、テキストと図表で示される幾何問題を解く手法 [11] を提案した。この問題では図表で示される辺の長さや面積を選択肢から選ぶタスクである。彼らは

テキストの構文解析における曖昧性を、対応する図表から読み取れる情報を合わせることで解消している。Berzakらは、2通り以上に解釈できる文章と対になった動画データを用いて、どの解釈が正しいかを当てる手法 [1] を提案している。例えば“Sam approached the chair with a bag.”という文は、「Samがカバンを持って椅子に近づく」「Samはカバンのそばにある椅子に近づく」という2通りの解釈ができる。このタスクでは動画の各フレームのオブジェクトとテキストとの対応をとることで、解釈がどちらであるかを判断している。

このように、自然言語の意味を捉えるためには文の表層情報だけでは不十分であるため、これらを適切に処理するための現実と対応付けられた枠組みが必要となる。

2.2 共参照解析

共参照解析とは文章中の複数の名詞句のうち同一のものを表す組を見つけるタスクである。基本手法として各名詞句の形態的な特徴を用いて分類器を学習する手法 [12] が知られている。この形態的な特徴とは、二つの名詞句が「単数か、複数か」「男性名詞か、女性名詞か」などのことである。他にも二つの名詞句が文章中でどの程度離れているか、文字列の表記は同じか、など文の表層的な情報が用いられる。

最近では大規模な世界知識を用いるもの [9][10] も提案されている。これは例えば「オバマは大統領である」など二つの名詞句の関係を大量に用意しておいて、それら全てを特徴量として加え分類器を学習するものであるが、論文中のエラー分析から意味を捉えて判断するという点では不十分であると考えられる。

我々は既存手法では困難だと考えられる共参照解析に対して、人物の移動による制約を利用する手法を提案した [8]。この研究は自然言語の意味を捉えるにあたり、その背後にある世界の常識、法則などの一部をモデル化した世界モデルを用いて自然言語理解を目指すものである。しかしこの手法は人物移動をモデル化する際に、地図情報などを人手で生成しており、幅広

いドメインを自動的に扱うことが困難であった。そこで今回は Wikipedia の情報を用いて、より一般化を目指す。

3 提案手法

3.1 概要

自然言語の意味を捉える際に、背後にある世界の常識・法則などの一部として人物移動における空間・時間的な制約を考える。これは例えば「ある人物は、短い期間中に離れた場所に存在できない」などの基本的なルールを利用する。これらの情報は文の表層情報には現れないため、これを利用することで共参照解析を適切に処理することを目指す。

本研究では Wikipedia における人称代名詞 “He” に関する共参照解析の精度改善を行う。既存の解析器の共参照解析結果を、ある人物 (エージェント) に関する時間と空間の制約を利用し “He” の先行詞が不適切だと判断できる候補を除去することで精度向上をめざす。

今回は古代の世界史 (主にアレキサンダー大王の時代) をテーマにする。具体的な手法を以下に示す。

- 対象とするエージェント (人物)、地名を Wikipedia を用いて選択
- 対象とするエージェントと関連があると考えられる Wikipedia 本文の代名詞 “He” について共参照解析 (照応解析) を既存ツールで行う
- 対象とするエージェント名が明記された文章から制約条件 $c_i \in C$ (いつ誰がどこにいたか) を生成する
- “He” の解析結果からエージェント名を置換し制約条件 $c'_i \in C'$ を生成し、結果が適切であるかどうかを制約条件の集合 C から判断する

例えば、“(1) Alexander found Alexandria in 332 BC.” “(2) He invaded India in 333 BC.” というテキストがあり、既存の共参照解析器により “He” が “Alexander” を示す結果が得られた場合は、(1) と (2) からそれぞれ制約条件 c_1 ($agent = Alexander, location = Alexandria, time = -0331$) と c'_1 ($agent = Alexander, location = India, time = -0332$) が生成され、Alexandria と India の距離と時間条件から共参照解析結果が妥当か判断する。

3.2 対象とするエージェント、地名の選択

英語 Wikipedia の xml ファイルを利用する。この xml ファイルには InfoBox や Persondata など、地名の位置情報や人物のデータが含まれている。テーマの中心とな

るエージェントをアレキサンダー大王とし、“Alexander the Great” の記事からいくつかハイパーリンクをたどることで得られるページのうち、Persondata を持つものをエージェント、経度緯度情報をもつものを地名として採用する。

3.3 制約条件生成

制約条件は「いつ、誰が、どこにいたか」という情報をもつ。“Alexander the Great” のページからいくつかのハイパーリンクで繋がるエージェントのページすべての本文について Semantic Role Labeling (SRL) を行い、その結果からルールにより制約条件を生成する。

SRL とはある動詞に着目し、その動詞に必要な要素 (動く人、動かされる人、場所、時間など) をラベルとして付与するタスクである。“He invaded India in 333 BC.” の解析例は、predicate “invade.01” (A0: He, A1: India, AM-TMP: 333 BC) となる。“01” というのは “invade” と表記される別の意味を持つ文字列を区別するための番号である。各 argument に付与されたラベルは predicate ごとに決められており、例えば “invade.01” の場合は

- A0: invader
- A1: place invaded
- AM-TMP: time

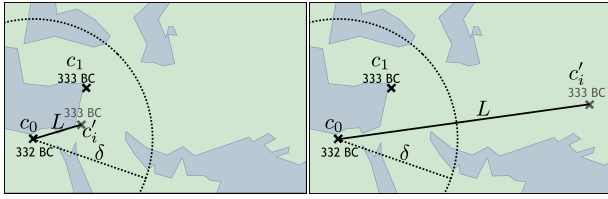
となっている。これらの要素と、SRL 結果とを合わせるとエージェントが 333 BC にどこにいたか、という情報から制約条件が生成できる。その際、本文中の表現がどのエージェント、どの地名に対応するかについては文字列の一致度合いをスコア化することで決定する。

3.4 共参照解析の誤り検出

本文中にエージェント名 f が明記された文から生成された制約条件の集合 C と、エージェント名が “He” であり、既存の共参照解析器で “He” が f と判断され生成された制約条件 $c'_i \in C'$ から、 c'_i に関する “He” が表すエージェントが適切であるかを判断する。

まず Wikipedia から得られた地名の経度緯度情報をもとに地図上に $c_j \in C$ ($j = 1..|C|$), c'_i をそれぞれ点 $P_{c_j}, P_{c'_i}$ としてマッピングする。次にその直線距離 $L = |P_{c_j} P_{c'_i}|$ と時間条件の差 Δt を求める (例えば 332 BC と 331 BC なら $\Delta t = 1$)。それらを用いて制約条件 c'_i に関するスコアを以下のように定義する。

$$Score(c') = \frac{\sum_i Val(c_i)}{|C|}$$



(a) 適切な例 (b) 不適切な例

図 1: 制約条件判断例

$$Val(c_i) = \begin{cases} 0 & (\text{if } L \leq (1 + \Delta t) \times \delta) \\ -1 & (\text{otherwise}) \end{cases}$$

ここで δ はエージェントが1年間に移動できる限界値を表す閾値とする。最終的にこのスコアが全体的な閾値 θ を上回れば制約条件として適切、つまり“He”の共参照解析は適切であると判断する(図1-(a))。逆に下回れば“He”の共参照解析結果が不適切であると判断する(図1-(b))。なお、 $|C| = 0$ となる場合については判断材料がないため適切と判断する。

4 実験

4.1 概要

対象とするテキストデータ中の“He”が誰を表しているのかをStanfordの共参照解析器で求め、その精度(F1値)を時間・空間制約を利用し改善する実験を行った。

4.2 対象とするデータ

対象の中心的なエージェントとして“Alexander the Great”を選択する。対象となるエージェントは、Wikipediaのxmlデータのハイパーリンクを“Alexander the Great”から2つたどり到達する全ページのうちPersondataの項目が含まれるページの人物(1,469人)とする。対象となる地名は、同じくリンクを2つたどり到達する全ページのうち、経度緯度情報を持つページの地名(6,579箇所)とする。制約条件生成のもととなるテキストデータについては、対象となるエージェントの全ページの本文(119,882文)とする。

4.3 制約条件生成

制約条件はエージェント、地名、時間の要素を持ち、例えば“Alexander found Alexandria in 332 BC.”というテキストから“agent=Alexander, location=Alexandria, time=-0331”という制約条件が生成される。“-0331”という表記はChangらのSUTime[4]のフォーマットによるものである。制約条件生成の具体的な実装とルールを以下にまとめる。

- Stanford Core NLP [7] によりテキストに対して tokenize を行う
- Bohnet [3] らと Björkelund [2] らのパイプラインを用いて SRL を行う
- Chang らの SUTime [4] により時間表現を抽出し、ルールにより制約条件を生成する
- Stanford Core NLP [6] によりテキストの共参照解析を行う

SRL 結果による生成ルール

- ある動詞 (predicate) の SRL 結果ごとに制約条件を生成する
- “A0”ラベルが付いた文字列をエージェントとする
- “AM-TMP”ラベルが付いた文字列を時間情報とする
- “AM-LOC” > “A4” > “A3” > “A2” > “A1” のラベル優先度で位置情報とする
- SUTime による時間表現が“-NNNN”で表現されないものは除去する

“-NNNN”の形式で表現される時間表現は紀元前何年という情報を持ち、今回はこれに限定した。またエージェントと位置情報についてはあらかじめ決定されたエージェントリストと地名リストの中から文字列間の一致度合いをスコア化しその最大のものを選択する。具体的には(全体一致) $\times 1000 +$ (単語一致) $\times 50 -$ (文字列長の差)の重みで算出する。今回はエージェント、地名ともにスコアが0以上となるものを制約条件生成の条件とした。その結果生成された制約条件は436個で、そのうち“He”を含むものは71個生成された。この71個について人手によるアノテーションを行い、Stanfordの共参照解析器による結果の正誤についてラベルを付与した。この71個の問題に対してシステムがそれぞれスコアを3.4節で述べた手法で算出し、その値が0以上ならばPositive、0未満ならばNegativeとする。これにより、空間・時間的な制約を満たせないような共参照解析結果をNegativeとして除去することでStanfordの共参照解析結果を改善することができる。ただし、Stanfordの共参照解析器が出力した正しい候補を除去する場合もあるため、本実験では精度とともに相対的な再現率(Relative Recall)を評価する。

4.4 実験結果

エージェントが1年間に移動できる限界値を表す閾値 δ を変化させた時の“He”に関する共有参照解析のF1値を表1と図2に示す。

δ (km)	0	200	400	600	...	2,800	3,000	3,200	3,400	...	10,000
True positive	0	29	34	37	...	47	48	48	48	...	50
False positive	0	11	11	11	...	18	18	18	19	...	21
True negative	50	21	16	13	...	3	2	2	2	...	0
False negative	21	10	10	10	...	3	3	3	2	...	0
Precision	-	0.725	0.756	0.771	...	0.723	0.727	0.727	0.716	...	0.704
(Relative) Recall	-	0.580	0.680	0.740	...	0.940	0.960	0.960	0.960	...	1.000
F1	-	0.644	0.716	0.755	...	0.817	0.828	0.828	0.821	...	0.826

表 1: δ を変えた時の F1 値

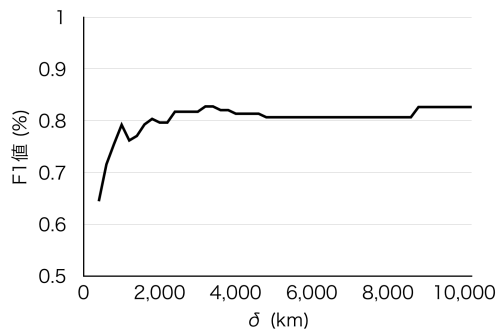


図 2: δ を変えた時の F1 値

まず $\delta = 0$ km より本実験の Stanford Core NLP による共参照解析の適合率 (Precision) は 0.704 であることがわかる。形態素解析や構文解析などのタスクの精度が 90% を超えるものがあるのに対して、共参照解析の精度は一般的に 7 割程度である。F1 値が最も高かったのは $\delta = 3,000$ km 付近で、Stanford の共参照解析器による F1 値をわずかに改善する結果となった。この 3,000km という値は人間の 1 年間に歩く平均距離が 1,800km であることから現実世界とそれほど矛盾しない結果となった。このことは現実の空間・時間的な制約をモデル化して用いていることから説明できる。 $\delta = 3,000$ km における False negative について、具体的な例を以下に示す。

- “After Alexander’s death (323 BC) he made himself master of the territories of the Indian king Porus, ...”

Stanford の共参照解析器では “he” を Alexander だと解釈しているが、Alexander の死後のことについて書かれているため、明らかに解析結果が間違っていることがわかる。このような誤りを時間・空間的な制約を利用することで除去することができた。

5 おわりに

関連研究で見られるように、文の意味を深く理解しようとする研究は最近徐々に増えている。しかし広く共

有されたデータセットが存在しないため、今回の実験でも問題の規模が小さく、適切な結果が得られていない可能性がある。以前の実験 [8] では設定が人手によるところが大きかったため、スケーラビリティを考えた時に問題があったが、本実験では人手による作業を Wikipedia を用いることで自動化したためデータを増やすことは可能である。今後はデータを増やし、より精度の高い実験を行うことが課題である。また、今回は共参照解析の誤り訂正に着目したが、既存のパサラーの出力結果を空間・時間的な制約によりスコアを求めることで正しい解析結果を得る、つまり誤り訂正についても今後取り組みたい。

謝辞

本研究は、JST、CREST の支援を受けたものである。

参考文献

- [1] Yevgeni Berzak, Andrei Barbu, Daniel Harari, Boris Katz, and Shimon Ullman. Do you see what i mean? visual resolution of linguistic ambiguities. In *EMNLP*, 2015.
- [2] Anders Björkelund, Love Hafdel, and Pierre Nugues. Multilingual semantic role labeling. In *CoNLL*, 2009.
- [3] Bernd Bohnet. Efficient parsing of syntactic and semantic dependency structures. In *CoNLL*, 2009.
- [4] Angel X Chang and Christopher D Manning. SUTIME: A library for recognizing and normalizing time expressions. In *LREC*, 2012.
- [5] Angel X Chang, Manolis Savva, and Christopher D Manning. Learning spatial knowledge for text to 3D scene generation. In *EMNLP*, 2014.
- [6] Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. StanfordTM’s multi-pass sieve coreference resolution system at the conll-2011 shared task. In *CoNLL*, 2011.
- [7] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *ACL: System Demonstrations*, 2014.
- [8] Yuki Murakami and Yoshimasa Tsuruoka. Where was alexander the great in 325 bc? toward understanding history text with a world model. In *Proceedings of the First Workshop on LSDSem*, 2015.
- [9] Haoruo Peng, Daniel Khashabi, and Dan Roth. Solving hard coreference problems. In *NAACL*, 2015.
- [10] Altaf Rahman and Vincent Ng. Coreference resolution with world knowledge. In *ACL*, 2011.
- [11] Minjoon Seo, Hannaneh Hajishirzi, Ali Farhadi, Oren Etzioni, and Clint Malcolm. Solving geometry problems: Combining text and diagram interpretation. In *EMNLP*, 2015.
- [12] Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. A machine learning approach to coreference resolution of noun phrases. In *Computational linguistics*, 2001.