

情報科学論文における問題解決手法と評価表現の付与仕様の検討

Towards Annotating Problem Solutions and Evaluation Expressions
in Information Science Papers白井 穂乃^{*1} 井之上 直也^{*1} 乾 健太郎^{*1*2}

Hono Shirai

Naoya Inoue

Kentaro Inui

^{*1}東北大学

Tohoku University

^{*2}理化学研究所 AIP センター

RIKEN Center for Advanced Intelligence Project (AIP)

The ultimate goal of our research is to extract problem solutions and its evaluation expressions from Information Science papers. This paper presents the preliminary annotation scheme for this task and conducts an annotation study in Natural Language Processing papers. By conducting the annotation study, we discuss the remaining issues of the present annotation scheme and its future direction.

1. はじめに

近年、学術論文の出版数が急増している。米 American Journal Experts の 2016 年次報告書^{*1}によると、薬学の分野では 560,000 本、工学では 280,000 本、生物分野では 260,000 本を超える論文が 1 年のうちに出版されたと言われている。我々研究者は、こうした膨大な学術文献の中から関連分野の情報を適切に取捨選択し、把握することが年々難しくなっている。

このような状況を打開すべく、学術論文から有用な情報を自動抽出するための様々な研究が盛んに行われている。例えば、Teufel ら [9] は、学術論文の各文を「背景に関する記述」「先行研究に関する記述」などに分類する Argumentative Zoning というタスクに取り組んでいる。また、文献からの情報抽出を関係抽出の問題として定式化し、評価型ワークショップを開催する動きもある。例えば、BioNLP [3] では、生物医学分野の文献を対象として、タンパク質等の専門用語の認識タスク、たんぱく質間の関係や、物質とその副作用などを抽出する関係抽出タスクの研究が行われている。また、ScienceIE [1] では、物理学、材料科学、計算機科学を対象として、フレーズの抽出や同義語・下位語の関係抽出などの研究が行われている。また、引用評価極性解析 (Citation Sentiment Analysis) の分野では、文献内で引用している文献に対する著者の感情極性を解析する研究が盛んに行われている [6]。最後に、文献間の引用関係に基づいて、文献の重要度や、技術のトレンドなどを自動的に解析する取り組み (Citation Network Analysis) も盛んに行われている [7]。

このような背景のもと、我々は、計算機科学分野の学術論文の自動解析の研究に取り組んでいる。計算機科学の論文の中で主に展開される議論のうちの一つは、ある問題 (タスク) に対する解決手法を提案し、提案手法と既存手法のメリット・デメリットを述べるものである。そこで我々は、まずこうした情報の自動抽出を実現するために、文献内のあるタスクを解くための手法 (問題解決手法と呼ぶ)、とその手法に対する特定の観点に基づく評価を抽出することを目指している。例えば、次のような一文が論文内にあったとしよう。

- (1) *The results indicate that the whole-sentence-based classifier performs the best.*

連絡先: 白井穂乃, 東北大学大学院情報科学研究科, 乾研究室,
hshirai@ecei.tohoku.ac.jp

^{*1} <https://www.aje.com/jp/arc/dist/docs/>

International-scholarly-publishing-report-2016.pdf

例 (1) では、*the whole-sentence-based classifier* という手法に対して、*performs the best*, すなわち「性能という観点」において、「良い」という評価がなされていることを自動認識するのが、我々の目的である。このように個々の論文における問題解決手法とそれに対する評価を自動認識し、その結果を集約することにより、分野全体の技術の特徴を手早く俯瞰できるようになると期待できる。

自然言語処理の分野では、このような観点付き評価の抽出問題は、評価分析 (Sentiment Analysis)、または意見マイニング (Opinion Mining) として、古くから取り組まれてきた [5]。しかしながら、こうした解析を学術論文のドメインに適用する試みはこれまでになく、タスクの具体的な設計方針 (何を評価対象とみなすか、何を観点とみなすかなど)、コーパス等の言語資源も整備されていない。

そこで本稿では、学術論文ドメインにおける観点付き評価の抽出タスクの設計方針を明らかにするための予備調査について報告する。より具体的には、ある一つのタスクについての自然言語処理の学術論文を対象とし (2. 節)、計算機科学ドメインに特化した観点付き評価分析のタスクを設計する (3. 節)。また、この仕様に基いて、実際にコーパスアノテーションを行った結果を報告する (4. 節)。

2. データ

まず、本稿が研究対象とするデータについて説明する。本稿が対象とする論文は、ACL Anthology Corpus [2] に収録されている自然言語処理の論文である。自然言語処理では、様々な問題 (またはタスク) が定義されており (例えば、形態素解析、構文解析)、それらに関してさまざまな問題解決手段が提案されている。本稿では、観点付き評価抽出タスクの設計に関する知見を得るために、ある一つのタスク、感情分析 (Sentiment Analysis) についての論文を収集した。具体的には、まず Google Scholar^{*2} を用いて、“Sentiment Analysis” をクエリとして検索し、引用関係にある合計 36 本の論文を収集した。このうち、ACL Anthology Corpus に含まれる 28 本の論文からテキストデータを抽出した。本稿では、問題解決手法の特徴や、既存研究との差異について端的に書かれている Introduction の節を用いて後述のアノテーションを行った。今後、これを Related Work の節に拡張することも予定している。

^{*2} <http://scholar.google.com>

3. アノテーションスキーム

自然言語処理のような技術論文は、問題を解決するための技術の提案について書かれている。

本研究では解決する手法 (TERM) とその評価極性、TERM の評価の理由 (REASON) という 2 つの要素についてアノテーションすることとした。

3.1 アノテーションの仕様

アノテーションするラベルを 2 つ定義する。問題を解決する手法・しくみである TERM, その TERM に対する評価を判断する理由・根拠 REASON である。以降の節では、それぞれの定義について説明する。

3.1.1 TERM

問題解決の手法に関する記述をアノテーションするために、TERM というラベルを導入する。具体的には、モデル・アルゴリズムといった仕組み・仕組みの持つ機能・仕組みが動作する方法を表す名詞句を TERM として付与する。例えば、下記の文を見てみよう。

- (2) *We employ a novel adaptive multi-compositionality layer in recursive neural network, which is named as AdaRNN (Dong et al., 2014).*

例 (2) では、*recursive neural network* と *AdaRNN* はそれぞれニューラルネットワークというモデルとその一種であるため、TERM として付与する。

なお、既存研究では問題設定 (e.g. Sentiment Classification, Classification Task) もアノテーション対象としているが、本研究は分野全体の技術の把握を目的としているため、問題設定についてはアノテーション対象としない。

3.1.2 REASON

次に、前述の TERM ラベルが付与された問題解決の手法に対する著者の観点付き評価を捉えるために、REASON というラベルを導入する。REASON は、論文の著者が TERM というターゲットに対する評価を述べているフレーズ (評価表現) である。例えば、下記の例文を見てみよう。

- (3) *The results indicate that the whole-sentence-based classifier performs the best.*

例 (3) では、まず *the whole-sentence-based classifier* が TERM として付与される。著者は、*the whole-sentence-based classifier* に対して、*performs the best* (その性能が最高である) と評価しているため、これを REASON と付与する。なお本稿では、REASON は TERM と同一文内のフレーズに限って付与し、文をまたいだ TERM と REASON の関係は、後処理において共参照関係の解消を行った後に集約することを想定する。また、REASON は必ず単一の TERM に対応しているため、TERM と REASON の間に Link という関係を定義する。

3.1.3 TERM-POSITIVE, TERM-NEGATIVE

さらに本稿では、評価分析 (Sentiment Analysis) の先行研究に倣い、評価をポジティブ・ネガティブの二値で表現し、この評価を TERM に対する属性として付与する (以後、これを TERM-POSITIVE, TERM-NEGATIVE と示す)。例 (3) では、*performs the best* は TERM である *the whole-sentence-based classifier* に対するポジティブな評価であると考えられるため、TERM に「ポジティブ」の属性を付与する。

なお、REASON を付与しない事例として、単に問題解決手法の特徴、性質を述べているだけのフレーズがある。例えば、

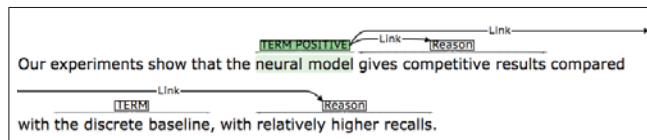


図 1: brat によるアノテーション

表 1: アノテーション結果

ラベル	数
TERM	520
TERM-POSITIVE	107
TERM-NEGATIVE	31
REASON	147

例 (2) では、著者は TERM である *recursive neural network* に対する評価を示していないため、REASON ラベルは付与しない。

3.2 アノテーションの手順

前節の仕様に基いて、下記のような手順でアノテーションを行う。

1. Introduction 内の問題解決手法を表す表現に対して TERM ラベルを付与する。
2. TERM と同一文内において、TERM に対する評価表現が存在するか判断する。あると判断した場合、下記 (a), (b) を実施する。
 - (a) 評価表現に基いて、TERM ラベルを TERM-POSITIVE, TERM-NEGATIVE ラベルに変更する。
 - (b) 評価表現に REASON ラベルを付与する。
 - (c) TERM-POSITIVE, TERM-NEGATIVE から REASON に Link ラベルを付与する。

4. 結果と考察

3. 節で定義した仕様の妥当性を検証するために、2. 節のデータに対して実際にアノテーションを行った。本稿では、アノテーション結果の考察を通して、アノテーションをさらにスケールさせる際の課題について議論する。

なお、アノテーションの際には、brat[8] を用いて図 1 のようなユーザインターフェースを作成し、第一著者 1 名がアノテーションを行った。複数アノテータによるアノテータ間一致率の分析については、今後の課題とする。

アノテーション結果は表 1 のようになった。

以下、アノテーション結果から分かったことをさらに詳しく報告する。

4.1 TERM の種類

TERM のラベルが付与される単語にどのような傾向があるか考察する。まず、model, approach, method など TERM の定義そのものを表す単語を含む名詞句には、TERM のラベルが付与されていることが分かった。

- (4) ... *our model outperforms state-of-the-art methods in both the supervised and semi-supervised settings*

例 (4) では、*model* という単語が含まれる *our model* と *method* という単語が含まれる *state-of-the-art methods* に TERM のラベルが付与された。

一方で、TERM の定義を表す単語が含まれていない名詞句に TERM が付与された事例もある。これらの事例には、文脈によって TERM であるか否かが変わるような名詞句が含まれていた。例えば、以下の 2 つの事例を見てみよう。

- (5) ... *we do not require gold standard sentence-level labels for training.*
- (6) ... *we propose a joint two-level model to address the aforementioned concerns.*

例 (5) において *we* は “*our model*” と同じ意味を持つため TERM であるが、例 (6) では *we* はモデルを提案する著者であるので TERM ではない。このほか、共参照関係を解消する必要がある *it* などの代名詞についても、同様の傾向がみられた。

4.2 TERM-POSITIVE, TERM-NEGATIVE

TERM-POSITIVE, TERM-NEGATIVE が付与された事例には、大きく (1) ドメイン非依存なポジティブ・ネガティブな評価表現を含んでいる事例、(2) ドメイン依存な評価表現を含んでいる事例があった。

ドメイン非依存な評価表現の事例には、例えば次のような事例がある。

- (7) ... *our system can generate high-quality labeled data.*

例 (7) では、*our system* に対して TERM-POSITIVE のラベルが付与され、このときの評価表現として *can generate high-quality labeled data* が付与されていた。この評価表現には、*can* や *high-quality* といった、ドメイン非依存と考えられるポジティブな単語が含まれている。高い質のラベル付きデータを生成できるというフレーズは *our system* をポジティブに評価する表現であると判断できる。このようなドメイン非依存な単語が含まれている場合は感情極性辞書を利用することで自動的に解析することが可能だと考えられる。

一方で、ドメイン依存な評価表現の事例には、例えば次のような事例がある。

- (8) *This approach requires no manually-specified information about the meaning of the connectors, just the connectors themselves*

例 (8) では *This approach* に TERM-POSITIVE のラベルが付与されており、*requires no manually-specified information* に REASON のラベルが付与されていた。ここでは、REASON の部分にはポジティブな単語が含まれていないが、*manually-specified information* (手作業で指定された情報) を必要としないという表現は「タスクを解決する上で必要とする情報が少ない」という観点において TERM にとってポジティブであることを示すため、*requires no manually-specified information* は *This approach* の評価表現である、というドメインの知識に基づくある種の推論が起きている。このようなドメイン依存な知識に基づく評価表現は、アノテータのドメインに関する知識に依存するため、将来アノテータ間の揺れが起りやすいと予想される。

4.3 REASON における観点の種類

最後に、REASON のラベルを付与された事例について、評価表現にどのような観点が含まれているか考察する。REASON のラベルを付与された事例には、(1) TERM の性能、(2) TERM の前提条件、(3) TERM の持つ機能、という観点があった。

まずはじめに、性能の観点における評価表現として、例えば次のような事例がある。

- (9) ... *it is a strong and robust performer*

例 (9) は *it* に TERM-POSITIVE のラベルが付与されており、*a strong and robust performer* という評価表現に REASON が付与されていた。ここで、*performer* は性能という観点を想起させる単語であり、この観点に対して、*strong, robust* という修飾句で、ポジティブな評価を表明している。同様に、例 (4) は、*outperform* が性能という観点を表す単語であり、かつ「上回る」というポジティブな意味を持つ単語であるためポジティブな評価を表明している。

次に、TERM の前提条件における観点の事例を挙げる。例 (8) では、*requires no manually-specified information* に REASON のラベルが付与されていた。ここでは、*This approach* が動作するためには、*manually-specified information* が必要ないこと、すなわち前提条件に関する観点が示されている。*manually-specified information* は、ドメイン知識によりネガティブなもの（準備にコストがかかるもの）であると読み手には分かるため、これが不要ないと表明することによって、この前提条件が暗に「良い」ことを表明している。

最後に、TERM の機能という観点における評価表現として、次のような事例がある。

- (10) ... *neural models use real-valued hidden layers to automatically learn feature combinations, which can capture complex semantic information that are difficult to express using traditional discrete manual features.*

例 (10) では *neural models* に TERM-POSITIVE のラベルが付与されており、それに対する評価表現として *can capture complex semantic information* に REASON のラベルが付与されている。ここでは、*can capture* は TERM の持つ機能という観点を想起させ、*complex semantic information* という難しい情報を捉える、と表明することで機能がポジティブであると評価している。

- (11) *The model can also avoid overfitting to features derived from neutral or objective sentences.*

例 (11) では *The model* に TERM-POSITIVE のラベルが付与されており、その評価表現である *can also avoid overfitting* に REASON のラベルが付与されている。*can also avoid* は TERM の機能という観点を想起させる単語であり、*overfitting* (過学習) というネガティブなものを防ぐことができる、と表明することで機能がポジティブであると評価している。

5. 関連研究

建石ら [10] の研究は、情報科学論文に出現する用語間の関係を構造化するためのタグ付けスキーマを提案している。これは論文中に存在する用語すべてに対して意味クラスを付与し、用語間にも関係のクラスを付与することで論文内容の構造化を行うことを目的としている。また、自然言語処理分野の評

価型ワークショップ SemEval では、論文ドメインでの情報抽出タスクが提案されている。SemEval-2017 の評価タスクである ScienceIE[1] は物理学、材料科学、計算機科学の論文からフレーズと関係の抽出を行うタスクである。フレーズは Task, Process, Material の 3 つのクラスに対応している。フレーズのクラスを分類するタスク、フレーズ同士が類義語・下位語の抽出を試みるタスクが提案されている。SemEval-2018 Task 7 [4] は ACL Anthology Corpus[2] の Abstract について、エンティティ同士の関係を分類するタスクを提案している。エンティティは概念を表現する名詞句と規定されており、比較・結果などの 5 つの関係を抽出・分類するタスクである。本研究では、これら一般的な論文解析タスクとは異なり、文章中の用語を全てを対象にはせず、問題解決手法である TERM のみを抽出している。また、TERM だけでなく TERM に対する評価(極性)とその評価表現 REASON を付与している。

冒頭でも触れたが、自然言語処理の分野では、レビュー文章のドメインにおける観点付き感情極性分析が行われている。例えば、SemEval-2015 Task 12[5] は、ホテルやレストランの料理の価格やサービスの質などの定義した観点に基づいて評価分析を行うタスクである。しかし、我々の知る限り、論文ドメインにおいては観点に基づいた評価分析は行われていない。

6. おわりに

本稿では、情報科学論文から問題解決手法とその評価表現を自動抽出することを見据え、そのアノテーションスキームの設計・検討を行った。アノテーションの結果より、アノテーションをさらにスケールさせる際の注意点、すなわち、問題解決手法ラベル TERM の文脈依存性や、評価表現ラベル REASON 付与時におけるドメイン特化知識の必要性が明らかになった。今後の課題として、付与対象の大規模化、複数アノテータによる付与、アノテータ間一致率の分析、相対的な評価表現(既存手法に比べて「相対的に」ポジティブ等)への対応、自動解析モデルの構築が挙げられる。

謝辞

本研究は、JSPS 科研費 16H06614、JST CREST (課題番号: JPMJCR1513) の支援を受けて行った。

参考文献

- [1] Isabelle Augenstein, Mrinal Das, Sebastian Riedel, Lakshmi Vikraman, and Andrew McCallum. Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 546–555, 2017.
- [2] Steven Bird, Robert Dale, Bonnie J Dorr, Bryan Gibson, Mark T Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. *Proc. Sixth Int. Conf. Lang. Resour. Eval. (LREC 2008)*, pp. 1755–1759, 2008.
- [3] Louise Delèger, Robert Bossy, Estelle Chaix, Mouhamadou Ba, Arnaud Ferré, Philippe Bessieres, and Claire Nédellec. Overview of the bacteria biotope

task at bionlp shared task 2016. In *Proceedings of the 4th BioNLP Shared Task Workshop*, pp. 12–22, 2016.

- [4] Kata Gábor, Davide Buscaldi, Anne-Kathrin Schumann, Behrang QasemiZadeh, Haïfa Zargayouna, and Thierry Charnois. Semeval-2018 Task 7: Semantic relation extraction and classification in scientific papers. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, 2018.
- [5] Tomáš Hercig, Tomáš Brychcín, Lukáš Svoboda, and Michal Konkol. UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proc. 10th Int. Work. Semant. Eval.*, pp. 342–349, 2016.
- [6] Myriam Hernández-Alvarez and José M. Gomez. Survey about citation context analysis: Tasks, techniques, and resources. *Nat. Lang. Eng.*, Vol. 22, No. 3, pp. 327–349, 2016.
- [7] Yuya Kajikawa, Junko Ohno, Yoshiyuki Takeda, Katsumori Matsushima, and Hiroshi Komiyama. Creating an academic landscape of sustainability science: an analysis of the citation network. *Sustainability Science*, Vol. 2, No. 2, p. 221, Jul 2007.
- [8] Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 102–107, Avignon, France, April 2012. Association for Computational Linguistics.
- [9] Simone Teufel, et al. *Argumentative zoning: Information extraction from scientific text*. PhD thesis, University of Edinburgh, 2000.
- [10] 建石由佳, 仕田原容, 宮尾祐介, 相澤彰子. 情報科学論文からの意味関係抽出に向けたタグ付けスキーム. 言語処理学会第 19 回年次大会, 2013.