

アノテーションのためのアノテーション

徳永 健伸 飯田 龍

東京工業大学 大学院情報理工学研究科

{take,ryu-i}@cl.cs.titech.ac.jp

1 はじめに

自然言語処理の分野では、この20年間コーパスの構築と機械学習を利用する手法によって多くの成果があがっている。この手法では、特定の課題を解くために、その課題の正解をアノテーションしたコーパスを構築し、それを訓練データとして機械学習の技術を使ってその課題を解く。従来、専門家ができるだけ広範な言語現象を網羅することを考えながら人手で規則を記述していたのに対し、この手法では、必ずしも専門家ではない複数のアノテータがコーパスに情報をアノテーションし、アノテーションされた事例数を増やすことによって言語現象の網羅性を高めることを目指している。専門家の役割は、複数のアノテータが一貫性を維持しながら大規模なコーパスにアノテーションができるようなアノテーションのガイドラインを整備することになる。十分な量のコーパスから規則性を抽出するのは専門家ではなく、機械学習の手法によっておこなわれる。

コーパスと機械学習による手法は形態素解析や統語解析など自然言語処理の基礎技術のみならず、情報抽出や自動要約などの応用処理においても、その有効性が示されている。特に TREC, MUC, CoNLL, SemEval といった評価型会議では、アノテーションされたコーパスが参加者共通のデータとして利用され、手法間の定量的な比較がおこなわれている。しかしながら、すべての問題がこの手法で解決できるわけではなく、この手法が支配的になることに対する懸念もある (Reiter, 2007; Steedman, 2008; Krahmer, 2010; Church, 2011)。

コーパスと機械学習による手法では、コーパスにアノテーションされた課題に対する正解と、テキスト中の局所的、表層的な情報を利用して課題を解こうとする。しかしながら、課題解決のために利用する情報を選択する際、アノテータが実際に利用している情報をふまえて検討することはほとんどなかった。これはコーパスのアノテーションにおいてはアノテーションの結果のみが重視され、アノテータがアノテーションの過程においてどのような情報を利

用してアノテーションをおこなっているかにはほとんど関心が払われていなかったことと関連する。これまでにアノテーション中のアノテータの意志決定過程を収集した例はほとんどない。

工学的な観点からは、現在利用しているような情報で十分な性能が得られるのなら、アノテータがアノテーション過程で実際に利用している情報を使う必要は必ずしもないだろう。実際、形態素解析や統語解析では、テキストの表層的な情報を使うだけでも高い性能を得ている (Neubig et al., 2011)。しかしながら、テキストの表層的な情報だけを使ってすべての課題が十分な性能で解けるわけではない。たとえば、言語資源も豊富で研究の進んでいると考えられる英語の共参照解析の問題でも未だに F 値で 0.7 程度の性能に留まっており (Haghighi and Klein, 2010)、省略も考慮しなければならない日本語の共参照解析では F 値はさらに低く、0.4 程度に留まっている (Iida and Poesio, 2011)。これらの課題を解くためにはこれまで使われてこなかった情報に目を向ける必要がある。

このような背景をふまえ、我々はコーパス中の各アノテーション事例について、アノテータがそのアノテーションをおこなった根拠をアノテーションすることを提案する。このような根拠はそのアノテーションの妥当性を説明する情報であり、オブジェクトレベルのアノテーション情報に対するメタレベルの情報だと考えることができる。アノテーション事例にアノテーションされた根拠を分析することによって、当該課題を解くためにこれまで機械学習の素性として利用されてこなかった、より有用な情報を明らかにできる可能性がある。

以下、2 節では、アノテーションの根拠を分析するために収集するデータについて、3 節では考えられるデータの利用について述べる。4 節ではこのプロジェクトを遂行する上での研究課題についてまとめ、5 節で関連研究について述べる。

2 データの収集

個々のアノテーションの根拠を分析するために、我々は互いに相補的な二種類のデータを収集することを考えている。ひとつはアノテータがアノテーション中におこなう行動などの観察可能なデータである。これにはアノテーションツールの操作や作業中の視線などが含まれる。もうひとつはアノテーション中のアノテータの思考過程のように外部からは自然に観察できない潜在的なデータである。

観察可能なデータを収集するためには、そのための仕組みが必要である。たとえば、アノテータがアノテーションツールを使う行動を記録するためにはキーボード入力やマウス操作などの情報がタイムスタンプと一緒に記録できる機能が必要となる。アノテーション中の視線を計測するには視線計測装置が必要である。最近の視線計測技術の発展はめざましく、適当な設定をすれば画面上のアノテータの視線座標をテキストの文字単位の精度で計測することが可能である。これらの仕組みを使えばアノテーション中のアノテータの行動を詳細に記録することが可能になる。また、自動的に記録されたこれらの低レベルのデータは、アノテーション課題に応じてさらに解釈しやすい形に変換する必要がある。たとえば、述語項構造のアノテーションでは、マウスのクリックやドラッグ操作は、述語や項に対応する語や句の同定や述語と項の関係の同定といった操作に変換できる。同様にある時刻における視線座標はそれに対応する語や句への注視に変換できる。これらを時間同期させて分析することによって、たとえば、述語と項の関係を同定する前にアノテータがどの語や句を注視し、どのようなタイミングで関係を付けたかといった情報が得られる。これらの情報はアノテータが個々のアノテーションをする際の意志決定過程を推測する上で重要な手掛りとなる。また、これらの情報はアノテータのアノテーション作業を阻害することなく収集できるという利点がある。

前述のような観測装置では潜在的なデータを収集すること難しい。人間の認知過程を言語化して収集する手法として思考表出法 (think-aloud protocol: TAP) (Ericsson and Simon, 1984) がある。TAPは、作業中の被験者に自分の考えていることを逐一声に出して報告させ、それを記録する手法である。TAPを使えば、アノテータの思考過程を直接言語化して記録することが可能となるが、アノテータはアノテーション作業に加えて、思考過程の報告を同時におこなう必要があるため、アノテータの認知負荷が高くなってしまい、アノテーション本来の作業に支障をきたす可能性がある。我々はアノテータの認知負荷を軽減するために、アノテータを二名一組

で作業させ、その間で交される対話を記録し、対話分析の手法を使うことによってアノテータの思考過程を言語化することを計画している。対話中の発話は思考過程の間接的な証拠にしかならないが、対話は協調作業においては自然な行為であり、TAPに比べると認知負荷の軽減が期待できる。

3 データの利用

収集したデータの利用法として以下のような利用が考えられる。

【言語処理に有用な情報の発見】 前述したようにコーパスと機械学習による手法では、まず、対象となる課題の正解情報をコーパスに人手でアノテーションする。この時のアノテータの意志決定プロセスを理解することは、従来、機械学習の素性として利用されなかった情報を発見する手掛りを与える。たとえば、共参照解析のためのアノテーションをする際にアノテータが先行詞を探索するのにどのような順でどの範囲まで探索するのかといった情報は収集したデータから得ることできる。これにより、人間がどのように短期記憶中に先行詞の候補を保持し、それをどう処理しているのかを間接的に観測できる可能性がある。

【アノテーションの質の評価】 アノテーションの質は複数のアノテータによって同じ対象に同じ基準でアノテーションをした時の一致率や κ 係数で評価するのが一般的である (Carletta, 1996)。これはコーパスに付与されたアノテーション全体に対するいわばマクロな評価であるが、収集したデータは個々のアノテーション事例に付与されるので、それを基にして各アノテーション事例ごとの評価をおこなうミクロな評価が可能になる。たとえば、情報を付与するまでに長い時間がかかる、二人一組でアノテーションしている場合に二人の判断が異なるなどの情報はそのアノテーション事例が難しい事例であることを示唆する。

【アノテータの質の評価と育成】 従来、アノテーションの質に比べてアノテータの質に関する議論はほとんどなされてこなかった。コーパスと機械学習による手法では、精度の高いアノテーションは解析システムの精度に直接影響するので、それを作り出す質の高いアノテータを育成することは重要な課題である。アノテータの質は他のアノテータとの一致率や絶対的な正解が付与された参照コーパスとの比較によって評価することができる。これはやはりアノテーション結果に基づくマクロな評価である。これに対して、収集したデータを利用するとアノテーション結果ではなく、アノテーション中の振舞いを

熟達したアノテータの振舞いと比較することによってアノテータの質の動的な側面の評価が可能となる。具体的な行動指針を与えることができるという意味で、これはアノテータを育成するという観点からも効果的な情報となる。

4 研究課題

本節では今後プロジェクトを進める上で解決すべき研究課題について述べる。言うまでもなくこれはすべてを網羅したリストではない。

【データ収集環境の整備】 2節で述べたように、特に観察可能なデータを収集するためには、作業記録が残せるアノテーションツールや視線計測装置などが必要である。また、これらのソフトウェアとハードウェアが協調的に動作できるように整備も必要となる。さらに、二人一組でアノテーションする場合は、アノテーションツールを二人が協調して使えるような改修も必要となる。アノテーションツールにも様々な種類のアノテーションがおこなえる汎用性が必要である。

【アノテーション課題の選定】 形態素解析などのようなすでにコーパスと機械学習に基づく既存手法で十分な性能が得られている課題を対象とするよりは、意味解析や談話解析などの十分な成果が得られていない課題を優先して取り組む必要がある。

【アノテータの組の構成基準】 二人一組でアノテーションを協調しておこなう際には二人のアノテーション経験・能力の差を考慮する必要がある。どのような組合せがアノテーションの根拠となる情報を引き出すのに有効かを実験を通して明らかにしなければならない。

【観測可能なデータと潜在的データの統合】 2節では、二種類のデータを収集することについて述べた。これらを統合して分析する手法の確立が必要である。

【アノテーション根拠の表現】 前項と関連して、各アノテーション事例にアノテーションする根拠をどのように表現するかも重要な研究課題である。表現内容についてはアノテーション課題に依存する部分があると予想できるが、表現形式と一般的なガイドラインは規定しておく必要がある。

5 関連研究

Rosengrant (2010) は、被験者が電気回路の問題を解く過程で TAP と視線計測を併用してデータを収集し、これらの異なる種類のデータを合せて分析する gaze scribing という分析手法を提唱している。

被験者はコンピュータ画面のペイントツールに提示された回路図を見ながら、問題として指定された抵抗値などを計算する。この際、頭の中で考えた計算過程などを発話しながら、かつペイントツールを使ってメモすることを求められる。回路理論に精通した度合いに応じて被験者を玄人と素人に分類し、両者の問題解決の戦略の違いを分析している。また、多様な問題解決の課題に gaze scribing を適用することの重要性を強調している。ただし、gaze scribing はその手法が十分形式化されておらず、確立された手法というには程遠い。アノテーションも問題解決のひとつであると考えれば、我々が提案するデータの収集・分析も Rosengrant (2010) と同じ研究の方向性であるといえるが、我々は語や句などのより細かい粒度の対象に対する注視を扱い、協調作業をするアノテータ二人の対話から有用な情報を抽出することを目指しているため、データ収集に関する技術的な課題と同時に分析における課題も異なってくる。

二名で協調的に問題を解決する手法としてソフトウェア工学の分野ではペア・プログラミングの手法が知られている。ペア・プログラミングの本来の目的はソフトウェアを効率的に開発することであり、必ずしも二人の間でやりとりされる対話を分析することではないが、二人のプログラマの熟練度の組合せが結果にどのように影響するかなどの分析はおこなわれている (Lui and Chan, 2006)。また、ソフトウェア開発の過程における視線情報を分析する研究 (Bednarik and Tukiainen, 2008) やペア・プログラミングと視線情報を組み合わせた分析もある (Pietinen et al., 2008)。これらの研究は 4 節で述べたアノテータの組を構成する基準を策定する上で参考になる。

アノテーションについても最近、作業中のアノテータの振舞いを分析する研究がいくつかおこなわれるようになってきている。Tomanek et al. (2010) は能動学習でどの事例を追加するかを選択する基準として、個々のアノテーション事例の難しさの指標をアノテータの振舞いから推定することを試みている。あらかじめ定義されたテキスト領域の注視時間や注視の移動パターンからアノテータの認知負荷を推定し、それを基にアノテーション事例の難しさを見積っている。課題として固有名抽出を対象とし、分類する固有名を画面の中心に置き、周辺文脈を制御して複数の条件で実験をおこなっているため、自然なアノテーション作業の分析というよりは計画された心理実験となっている。また、Demirşahin et al. (2012) はペア・プログラミングの考え方をアノテーションに応用したペア・アノテーションを提案し、

トルコ語の談話関係のアノテーションをおこなっている。ペア・アノテーションの効用として、不注意なミスが減る、判断の揺れが起きたときの判断が早くなる、アノテーションの動機付けが維持され、集中できる、などを挙げている。ただし、アノテータ間でやりとりされる対話内容の細かい分析はされていない。

視線が簡単に計測できるようになったおかげで、最近では心理言語学や認知科学などのさまざまな研究分野で視線情報の利用が盛んになってきている (Duchowski, 2002)。特にテキストの読解や算出と視線との関係については心理言語学の分野で盛んに研究がおこなわれているが (Griffin and Bock, 2000; Richardson et al., 2007)、前述のような様々な問題解決において視線情報を分析に利用する研究は比較的新しい。アノテーションはテキストの読解を基に指示どおりのアノテーションをおこなうため、テキスト読解と問題解決の両方の側面を持っていると考えられる。我々のプロジェクトは視線情報を用いたテキスト読解の研究で得られた知見を生かしながら、分析対象を問題解決に広げるものであるといえる。

6 おわりに

本稿ではコーパスアノテーションにおいて、アノテーションされた個々の事例にアノテータがそのアノテーションをおこなった根拠となる情報をアノテーションすることを提案した。これらのメタな情報を分析することにより、従来は利用されていなかった情報を発見し、それを使うことによって言語処理の性能を改善することが期待できる。そのための具体的なデータ収集の可能性について議論し、実現にあたっての研究課題について述べた。プロジェクトの最初の試みとして、述語項構造のアノテーションにおけるアノテータの視線情報とアノテーションツールの操作履歴を収集し、その分析をおこなっている (光田 他, 2013)。今後は4節で述べた研究課題について研究を進める予定である。

参考文献

Roman Bednarik and Markku Tukiainen. 2008. Temporal eye-tracking data: evolution of debugging strategies with multiple representations. In *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08)*, pages 99–102.

Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

Kenneth Church. 2011. A pendulum swung too far. *Linguistic Issues in Language Technology*, 6(5):1–27.

Işın Demirşahin, İhsan Yalçınkaya, and Deniz Zeyrek. 2012. Pair annotation: Adaptation of pair programming to corpus annotation. In *Proceedings of the 6th Linguistic Annotation Workshop*, pages 31–39.

Andrew T. Duchowski. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, and Computers*, 34(4):455–470.

K. Anders Ericsson and Herbert A. Simon. 1984. *Protocol Analysis – Verbal Reports as Data*. The MIT Press.

Zenzi M. Griffin and Kathryn Bock. 2000. What the eyes say about speaking. *Psychological Science*, 11(4):274–279.

Aria Haghighi and Dan Klein. 2010. Coreference resolution in a modular, entity-centered model. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 385–393.

Ryu Iida and Massimo Poesio. 2011. A cross-lingual ILP solution to zero anaphora resolution. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011)*, pages 804–813.

Emiel Krahmer. 2010. What computational linguists can learn from psychologists (and vice versa). *Computational Linguistics*, 36(2):285–294.

Kim Man Lui and Keith C.C. Chan. 2006. Pair programming productivity: Novice–novice vs. expert–expert. *International Journal of Human-Computer Studies*, 64(9):915–925.

光田 航, 飯田 龍, 徳永 健伸. 2013. テキストアノテーションにおける視線と操作履歴の収集と分析. 言語処理学会第 19 回年次大会発表論文集.

Graham Neubig, Yosuke Nakata, and Shinsuke Mori. 2011. Pointwise prediction for robust, adaptable Japanese morphological analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 529–533.

Sami Pietinen, Roman Bednarik, Tatiana Glotova, Vesa Tenhunen, and Markku Tukiainen. 2008. A method to study visual attention aspects of collaboration: eye-tracking pair programmers simultaneously. In *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08)*, pages 39–42.

Ehud Reiter. 2007. The shrinking horizons of computational linguistics. *Computational Linguistics*, 33(2):283–287.

Daniel C. Richardson, Rick Dale, and Michael J. Spivey. 2007. Eye movements in language and cognition: A brief introduction. In Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson, and Michael J. Spivey, editors, *Methods in Cognitive Linguistics*, pages 323–344. John Benjamins.

David Rosengrant. 2010. Gaze scribing in physics problem solving. In *Proceedings of the 2010 symposium on Eye tracking research & applications (ETRA '10)*, pages 45–48.

Mark Steedman. 2008. On becoming a discipline. *Computational Linguistics*, 34(1):137–144.

Katrin Tomanek, Udo Hahn, Steffen Lohmann, and Jürgen Ziegler. 2010. A cognitive cost model of annotations based on eye-tracking data. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, pages 1158–1167.