

熟練者による解説文内イベントの出現と その根拠のアノテーション

亀甲 博貴 森 信介

京都大学 学術情報メディアセンター

{kameko, forest}@i.kyoto-u.ac.jp

1 はじめに

近年、シンボルグラウンディング課題への注目は大きくなっている。非テキスト情報に紐付けられたテキストを容易かつ大規模に収集することが可能となり、シンボルグラウンディング研究の多くは画像とテキストのペアを筆頭にこれらの非テキスト情報と紐付けられたテキストを活用して行われている [8, 1]。これらのコーパスを扱う際に、これらの対応が正しいことを仮定してその対応関係の獲得を目指している。しかし、自動で収集されたこれらのテキスト・非テキスト情報のペアは、必ずしも正しく対応しているとは限らない。シンボルグラウンディング研究において、このテキスト情報と非テキスト情報の間のギャップを解析することは、より高度なシステムの構築のために不可欠であると考えられる。

我々は将棋の解説文に着目している。将棋の解説文はプロ棋士同士の対局に対して第三者のエキスパートが解説を自然言語で付与したものである。これらの解説文はゲームの履歴中の各局面に紐付いており、概ね対応している局面に関する言及であるが、過去の状態や未来の予測への言及も含み、必ずしも対応している局面への言及とは限らない。我々は収集した解説文付き棋譜に対し、将棋ドメインの固有表現 (*s-NE*) を定義し、解説文のアノテーションを行った [6]。本稿では、この *s-NE* がゲーム中に発生するイベントであるとみなし、このイベントが実際の状態におけるどの時点で発生しているかに着目する。

また、我々はこれらのアノテーションを行う際に熟練者のアノテーターが状態のどこに着目しているかの根拠情報のアノテーションを同時に行った。人間のエキスパートは盤面の一部を指し示しながら解説を行うことがあるが、このような着目部分を示しながらの言及はゲームの解説に限らず多くの場面で見られる。人間が与えられた状態のうちどの部分を根拠に判断を行っているかを解析することは、コンピュータによる教師システムを構築する上で重要であると考えられる。このうち、画像などに比べて状態がシンボル化されているゲームを対象に根拠情報のアノテーションを行うことで、アノテーションを行う際の物理的な誤差を気にせず扱うことが期待できるため、このような根拠アノテ

ションの第一歩として適している。

我々は、この対応関係の解析を目的として、対象ゲームの熟練者によるイベント情報のアノテーションを行った。本稿では、このアノテーションによって得られた情報について記す。

2 イベントの根拠アノテーション

本稿では、イベントの出現情報に関する 3 層からなるアノテーションを定義する。提案するアノテーションは時系列ラベル、出現確率ラベル、根拠ラベルからなる。なお、本稿では対象の *s-NE* として戦型 (St), 囲い (Ca), 指し手名 (Mn) を選択した。これらは状態と強く結びつく固有表現であり、状態との対応が取りやすいことが期待できる。

図 1 にアノテーションに用いた Web インターフェースを示す。

2.1 時系列ラベル

時系列ラベルは対象のイベントが紐付いている状態に対してどの位置に現れるかを示す。

Present イベントが現状態に現れていることを示す。

例) 対する先手は美濃囲い_{Ca} に組んだ。

Past イベントが現状態には現れておらず、過去の状態に現れていたことを示す。

例) 85 手目の突き捨て_{Mn} は、この変化でも生きてくる。

Future イベントが現状態には現れておらず、未来の状態に現れることが予測されることを示す。

例) このあと居飛車_{St} に組む可能性が高そうだ。

Not イベントが現状態にも過去の状態にも現れておらず、未来の状態も現れないことが予測されることを示す。

例) (プレイヤー名) は居飛車_{St} 党。



後手番になった藤井は、平日手局記録に四間飛車/Stを採用。今度はどういった戦型になるのだろうか。

①1. 現れている ②2. 将来現れる ③3. 過去現れた ④4. 現れていない ⑤5. ラベル付けできない

将来現れる (「2. 将来現れる」を選んだ場合)。その確信度。
 ①1. 高い確率で現れる ②2. それなりの確率で現れる ③3. 現れるかもしれない (現れる可能性もあるが確率は低い) ④4. 解説のための悪手

将来現れる場合 (「2. 将来現れる」を選んだ場合)。現局面からの手順。
 指し手符号の列での記入をお願いします。(例: ▲7六歩△3四歩)

Submit

ラベル付与済みデータ

ラベルを変更したいコメントをクリックして下さい。
 更新の新しいものが上に並んでいます。

1. 後手番になった藤井は、**千日手/No**四間飛車を採用。今度はどういった戦型になるのだろうか。
 1. 本局は**千日手/No**指し手である。残り時間は三浦1時間7分。藤井1時間0分。対局開始時刻は1時13分。(相簿: コメント入力欄)

Instruction

不明な点がございましたら、亀甲 (kameki[at].ji.kyoto-u.ac.jp [at.] を半角アットマークに置き換える) までご連絡ください。

開始時

こちらで事前に指定した user id を入力して start ボタンを押してください。
 中断する場合はアラフツを閉じてください。
 再開時に同じ user id を入力することで対局を再開することができます。

ラベル付け時

図 1: アノテーションのための Web インターフェース。

Undecidable アノテーターが上記の 4 種類からタグを決定できなかったことを示す。このタグは原則として選ばれないことを期待している。

図 2 に 4 種類のタグが示す状態の例を示す。

2.2 出現確率ラベル

時系列ラベルにおいて、Present や Past は既に確定している情報であるが、Future は未確定の未来の状態の予測に基づく言及である。そこで、時系列ラベルにおいて Future を選んだイベントについて、そのイベントが発生する確率のラベル付けを行った。

- High** そのイベントが高い確率で発生する。
- One-of** 2 通り以上の起こりうる展開があり、対象のイベントはそのうちの 1 通りの展開において発生する。
- Low** 他により有力な展開があり、対象のイベントが発生する確率は低い。
- For-Comment** 解説文においては、説明のために本来は発生しないであろう展開について言及することもある。For-Comment タグが付与されたイベントは、このような解説のための言及における展開中に現れる。

	将棋倶楽部 24	将棋ウォーズ
Antr. 1	7 段 (1.12%)	5 段 (1.1%)
Antr. 2	6 段 (2.76%)	5 段 (1.1%)
Antr. 3	3 段 (14.73%)	3 段 (6.9%)
Antr. 4	3 段 (14.73%)	3 段 (6.9%)
第一著者		2 級 (69.9%)

表 1: アノテーター (Antr.) の棋力。「将棋倶楽部 24」と「将棋ウォーズ」はそれぞれオンライン将棋対局サーバで、全てのプレイヤは対局結果に応じてレートが付与される。括弧内の値はその段位が上位 X パーセントのプレイヤであることに相当することを示す。参考: http://shogi-naka.com/cgi-bin/24_ranking_checker.cgi, https://nandemoplus.com/shogiwars_occupancy/.

また、そのイベントが起こる展開について、手順の一例を付与した。

2.3 根拠ラベル

四間飛車という戦法は、飛車という駒を左から 4 列目に移動させて戦う戦法である。このように $s-NE$ の一部は状態の要約であり、解説文中にこのような $s-NE$ が現れた場合は多くの場合状態の一部と対応している。この対応を解析するため、状態のうち $s-NE$ に対応する部分を根拠ラベルとして付与した。

人間のエキスパートが解説をする際には、盤面の一部を指し示しながら言語による解説をする光景が多く見られる。根拠ラベルはこの指し示すべき盤面の一部とも対応していると考えられ、非言語情報を介しての解説生成を目指す上で根拠の解析は重要である。

アノテーターは根拠を図 1 に示すように盤面の一部をクリックして示す。インターフェースは盤面のマスなどをクリックすることで黄色くなり、これが根拠として選択されていることを示す。盤面は初期局面から現局面までのうちから 1 局面を選択することができ、アノテーターは 1 局面を選択した上でその局面に根拠ラベルを付与する。根拠として選択できるのは以下の要素である。

- 最終手
- 持ち駒
- 盤面上の各マス

3 統計情報

4 人の将棋の熟練者によってアノテーションを行った。表 1 にアノテーターの情報を示す。アノテーターの棋力はオンライン対局サーバにおける対局結果から付与されており、参加する将棋プレイヤーの中で上位に位置する棋力を有することから、高い棋

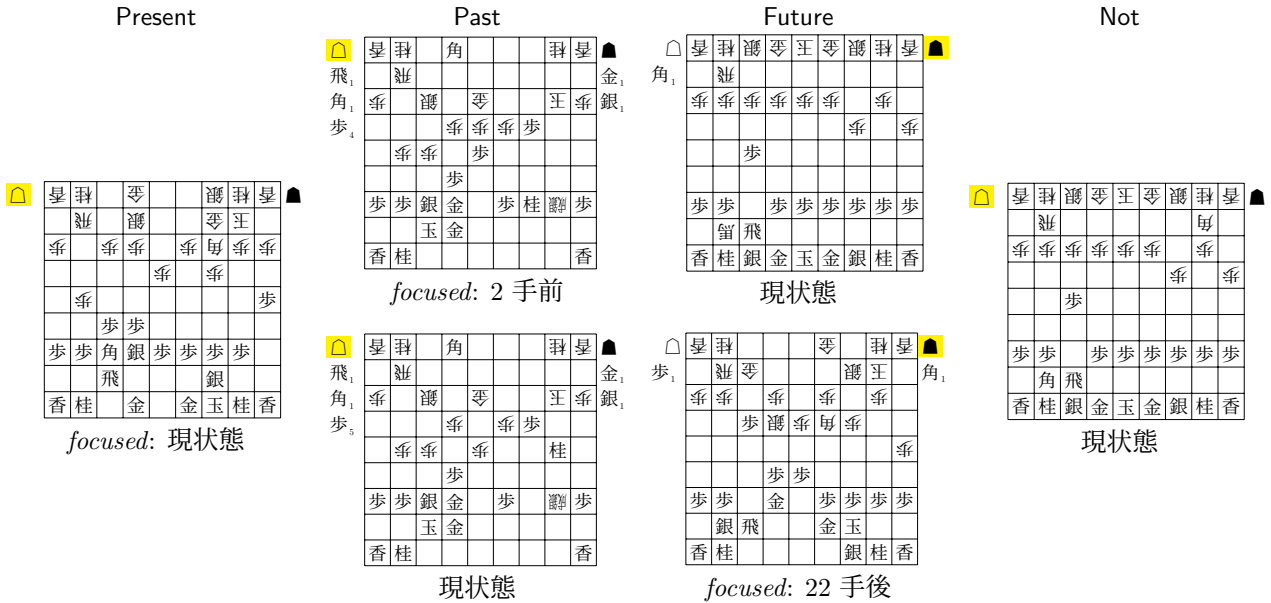


図 2: 時系列関係ラベルの例. Past と Future では, *focused* とした状態で対象のイベントが発生している.

St	Ca	Mn	Sum
56	39	124	219

表 2: コーパス内の *s-NE* の数.

	Present	Past	Future	Not	Undec.
Antr. 1	100	68	10	41	0
Antr. 2	91	82	13	33	0
Antr. 3	93	91	14	21	0
Antr. 4	100	78	15	26	0
第一著者	83	63	16	56	1

表 3: 各アノテーターが付与した時系列ラベルの数.

力を有していることが分かる. なお, 比較のため第一著者も同様のアノテーションを行った.

表 2 にアノテーション対象となった *s-NE* の数を示す. また表 3 に時系列ラベルの付与数を示す. アノテーターは Undecidable を付与せず, 全ての *s-NE* に対して 4 種類のラベルのうちいずれかを付与している.

表 4 にアノテーターペア毎の Cohen's Kappa 係数 [2] を示す. Cohen's Kappa 係数はアノテーションの結果がどの程度一致しているかを示す指標である. 0 から 1 の値を取り, 1 に近いほど高い同意が得られていることを示す. 表からどのアノテーターペアについても 0.7 程度の値を示しており, 高い同意が得られていることが分かる. また, 第一著者を除いた 4 人のアノテーターについて, 複数人のアノテーション結果について拡張した Fleiss' Kappa 係数 [3] は 0.74 となり, こちらも高い同意が取れていることを示している.

表 5 にアノテーターペア毎の根拠ラベルの一致

	1	2	3	4	A
Antr. 1		0.70	0.71	0.68	0.70
Antr. 2			0.78	0.78	0.72
Antr. 3				0.79	0.66
Antr. 4					0.68
第一著者					

表 4: 各アノテーターペア毎の Cohen Kappa 係数. "A" は第一著者を表す.

	1	2	3	4	A
Antr. 1		0.48	0.50	0.39	0.38
Antr. 2			0.50	0.44	0.44
Antr. 3				0.53	0.37
Antr. 4					0.66
第一著者					

表 5: 各アノテーターペア毎の根拠ラベルの F 値. "A" は第一著者を表す.

率を示す. 根拠ラベルの一致率は, 根拠として提示した箇所を真値とした F 値によって計測した. 表から, 時系列ラベルの一致率の高さと比べると根拠ラベルの一致率が低い.

根拠ラベルの一致率の低さの原因として 2 つの理由が考えられる. 1 つはアノテーターがそれぞれ独自の思考でアノテーションを行っているので, そもそも判断根拠が異なることが考えられる. もう 1 つは根拠を提示すること自体の難しさによるアノテーションの揺らぎによるものである. 人間の思考の決定過程を考察する上で, これらの解析を行うことは今後の課題である.

4 本コーパスを利用した展望

4.1 シンボルグラウンディング性能評価

本コーパスはシンボルグラウンディングモジュールの性能評価に用いることができる。機械学習によってテキストと非テキスト情報の対応付けを学習する際には、多くの場合大量のテキスト・非テキストペアを教師データとしてその対応関係を獲得する。しかし自動で獲得した大量のテキスト・非テキストペアが真に対応しているとは限らず、この対応関係の獲得の精度を評価する際にはこのギャップを考慮する必要がある。本コーパスは対象ドメインの熟練者によって対応関係をアノテーションしているため、より正確な対応関係の評価が可能である。

また、本コーパスは時系列情報の対応付けの評価に用いることもできる。我々は以前の研究で将棋の解説文に対して認識的モダリティ情報のアノテーションを行った [5] が、これは本コーパスにおける時系列ラベルと密接な関係にある。当該コーパスにおける時間関係のラベルや事実性関係のラベルは、本コーパスにおける時系列ラベルや出現確率ラベルと対応しており、イベントとそれを形容するモダリティ表現の対応関係になっている。

4.2 説明のための根拠提示

本コーパスにおける根拠ラベルは、人間が解説をする際に指し示す盤面情報であるとみなすことができる。非テキスト情報をテキストで説明する文生成を行うことを考えた際に、実際に人間が行っているように非テキスト情報の着目点を指し示しながら説明文生成を行うことを考える。既存手法ではアテンション機構の挙動を可視化 [7] するなど考えられる。またモデルの根拠提示手法としてモデルの勾配を可視化するなどの手法 [4] も取られており、これらを文生成に活用することも考えられる。本コーパスの根拠ラベルは、これらの根拠提示モデルの評価に用いるなど、より賢い説明生成モデルの構築のための活用が期待できる。

5 おわりに

本稿では将棋の解説文を対象にイベントの出現とその根拠のアノテーションを行った。本コーパスで得られた情報は説明生成などのシンボルグラウンディング課題への活用が期待できる。より賢いシステム構築のためには対象ドメインの熟練者の意思決定過程を解析する必要がある。本コーパスや本稿で提案するアノテーションはその解析に適している。

謝辞

本研究は JSPS 科研費 18K11427, 19K20341 の助成を受けたものです。

References

- [1] Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO captions: Data collection and evaluation server. *CoRR*, abs/1504.00325, 2015.
- [2] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.
- [3] Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.
- [4] Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky. Visualizing and understanding neural models in NLP. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 681–691, San Diego, California, June 2016. Association for Computational Linguistics.
- [5] Suguru Matsuyoshi, Hirotaka Kameko, Yugo Murawaki, and Shinsuke Mori. Annotating modality expressions and event factuality for a Japanese chess commentary corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 2018.
- [6] Shinsuke Mori, John Richardson, Atsushi Ushiku, Tetsuro Sasada, Hirotaka Kameko, and Yoshimasa Tsuruoka. A Japanese chess commentary corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2016.
- [7] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *arXiv*, 2015.
- [8] Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *TACL*, 2:67–78, 2014.