

将棋解説の自動生成のための局面からの特徴語生成

亀甲 博貴^{1,a)} 浦 晃¹ 三輪 誠² 鶴岡 慶雅¹ 森 信介³ 近山 隆¹

概要: 将棋の対局を観戦する上でコンピュータ将棋プログラムの形勢判断は有益な情報である。自然言語による局面の解説をプログラムが行えるようになれば、より有益な情報を提供できると考えられる。本稿では将棋の解説文を生成するモデルを提案し、その根幹である局面からの特徴語生成について調査した。学習の対象とする文を、意味でクラス分けする分類器を学習しその予測を用いることで限定した。特徴語生成はF値0.38と期待より低い値となったものの、解説文の生成においては、特定の局面については正しい文を生成できることが分かった。

Generating Characteristic Words from Positions for Shogi Commentary

HIROTAKA KAMEKO^{1,a)} AKIRA URA¹ MAKOTO MIWA² YOSHIMASA TSURUOKA¹ SHINSUKE MORI³
TAKASHI CHIKAYAMA¹

Abstract: Evaluation of positions and moves provided by computer Shogi programs is helpful for people watching a game of Shogi. Such evaluation information will be even more useful if it can also be given in natural language. In this paper, we propose a model for generating Shogi commentary and investigate how we could generate characteristic words from given positions. We limited the set of sentences for experiments by semantic classification. Although the F-score of generating characteristic words was 0.38, lower than we expected, we could generate the right commentary for some specific positions.

1. はじめに

近年のコンピュータ将棋プログラムの棋力向上は目覚ましく、プロ棋士と公式に対局が行われるに至るまでその強さが認知されるようになった。また有力な将棋プログラムの中には、将棋ファンが入手し気軽に扱えるようになってきているものも多い。これらの将棋ソフトの多くは局面の形勢判断を評価値として返す機能を持つことから、プロ棋士同士の対局においてその形勢を将棋ソフトを用いて判断する光景もよくみられるようになった。棋力の高いプレイヤーの対局を観戦する上で、ソフトが判断する形勢や最善手は大いに役立っているといえる。

しかしこれらのプログラムが返すのは最善手や形勢のみであり、なぜその指し手が選択されるのかを理解するには観戦者にある程度の棋力が求められる。そのため観戦者がプロ棋士同士の対局を楽しむために、タイトル戦など注目されている公開対局ではプロ棋士が解説を行っている場合が多い。棋力の上でプロ棋士に近づきつつあるプログラムによる局面の解説が可能となれば、解説者がつかないような対局において将棋ファンの観戦に有益であると期待できる。

本稿ではコンピュータ将棋プログラムによる将棋の解説文生成のための第一歩として、局面を表現する特徴的な単語を予測する手法を提案する。それらの単語は将棋ファンが局面を理解するのに役立つとともに、解説文の生成への入力として利用できることが期待できる。

本稿では2章で関連研究について述べ、3章で提案手法を紹介する。4章で提案手法の評価について述べ、5章でまとめを述べる。

¹ 東京大学大学院工学系研究科

Graduate School of Engineering, The University of Tokyo

² マンチェスター大学コンピュータ科学科

School of Computer Science, The University of Manchester

³ 京都大学学術情報メディアセンター

Academic Center for Computing and Media Studies, Kyoto University

a) kameko@logos.t.u-tokyo.ac.jp

2. 関連研究

2.1 ゲームにおける解説文生成

チェスの解説を生成する研究として、評価関数の各特徴量を用いたものがある [1]. これはチェスの局面を評価する際に用いる特徴量に応じて、あらかじめ与えたルールによって解説文を出力するというものである。チェスは人の手で設計された評価関数でも十分に強いコンピュータプレイヤーが作成できており、この研究ではそのような人の手で設計された特徴量から解説文の生成を行っている。対して近年の有力な将棋プログラムは機械学習を用いての評価関数の自動獲得 [2] を行っているものが主であり、これを人間が理解することは困難である。故にチェスと同様に評価関数が用いている特徴量からルールベースで解説文の生成を行うのは困難であることが予想される。

将棋の解説文を生成するシステムとして、コンピュータ将棋の読み筋を出力するものがある [3]. 棋力の高いコンピュータ将棋プログラムの1つである GPS 将棋の読み筋と評価値を、あらかじめ与えたルールによって人が読みやすい形に加工し、ミニブログサービスの1つである Twitter *1 に投稿することで広く将棋ファンに向けて公開するというものである。探索による最善手順とその評価値を出力している。

このシステムは単に最善手と評価値を返すのみではなく、観戦において有益であろう情報も提供する。例えば読み筋があったり必至や詰めろがかかっていたりすることが分かれば、最善手がそれらに該当することを示す。また現状では読み筋がないが持ち駒が増えれば詰みが生じるという局面も存在する。そのような場合、相手はその駒を渡さないように指さなければならぬ。そういった局面を発見するため、手駒を増やしての詰み探索を行う。加えて指し手の狙いが何かを示すため、手番を変更し連続で指せるとしたらどう指すかを探索する。これらの探索は、特に手番を変更しての探索は通常の指し手決定の際には行われぬものであるが、探索の簡単な工夫により容易に探索結果が得られかつ観戦者に有益な情報となる。

観戦時に将棋プログラムが持つ大きな役割の1つに、形勢判断がある。[3] では、プログラムが返す評価値についての統計をとり考察している。なお GPS 将棋の評価関数は evaluation curve [4] という勝率と関連させる手法を用いて調整されており、その評価値は勝率と密接な関わりを持つ。[3] ではインターネット対局サーバ『将棋倶楽部 24』 *2 から得られるアマチュアの対局棋譜とプロ棋士の対局棋譜について、GPS 将棋が返す評価値がある値となった局面数の頻度の統計を取っている。図示するとどちらも 0 を中心に

山型になっているが、アマチュアの棋譜に比べてプロ棋士の棋譜は 0 点付近の頻度が突出している。原因として、プロ棋士の対局ではミスが少なく五角の局面が多く生じることと、実際には形勢に差がついているにも関わらず GPS 将棋がその差を理解できず五角と判断してしまうことの 2 つの可能性があるとされている。

またリアルタイムでの解説を行った際に予想した指し手が何手先まで一致したかを報告している。定跡がほぼ終わっていると期待される 35 手以降と、終盤に入っていると予想される 90 手以降について、それぞれ一致深さの頻度を調査し、35 手以降については 804 局面中 58.26% の 471 局面、90 手以降については 146 局面中 62.27% の 96 局面が 1 手以上一致した。つまり半数以上の局面について指し手の予想が 1 手以上あたっている。

コンピュータ将棋プログラムが出力する読み筋やその評価値は多くの将棋ファンの将棋観戦に有益な情報であるが、一方でその読み筋の意味するところを解説するには至っていない。多様な局面において充実した解説を将棋ファンが理解できる形で提供するためには、自然言語を用いる必要があると考えられる。

2.2 その他の分野における自然言語や単語の生成

自然言語生成は天気予報の生成 [5] や仮想人間との対話 [6]、ジョークの生成 [7] など幅広い分野を対象とする。自然言語生成の一例として、天気の予測データから天気予報を生成する研究がある [5]。例えば 3 時間毎の風向きと風速を予測した時系列データが与えられたとき、天気予報ではただ予測データを羅列するのみではなく、データを簡潔に説明する文章が求められる。具体的には、事前に用意されたデータの変動モデルに合致するように入力された時系列データをいくつかに分割し、発話する内容を決定する。その後、用意された発話ルールに実際の観測データを埋める形で発話する。発話内容を事前に用意したモデルに当てはめることで、自然かつ意味のある文章の生成に成功している。将棋の解説文生成との大きな相違点として、出力すべき文章がある程度自明であると考えられることが挙げられる。例えばある時刻での予想風速が急激に速くなっていれば、そこについて言及すればいいのだろうと容易に想像がつく。一方で将棋の解説は、解説すべき個所が局面に明らかになっているわけではない。局面上の様々な特徴を勘案して解説文を生成する必要があるため、何を出力すべきかという問題は将棋の解説文生成の方がより難しいものであるといえる。

将棋の解説文生成と似た性質を持つと考えられる研究に、画像と単語の関係をモデル化するタスクがある [8]。データから単語を生成する点や、対象とするデータ中に生成すべき単語を明示的に意味するものがないという点に本研究との共通点が見られる。この問題に取り組む研究の1つで

*1 GPS 将棋のアカウント: <https://twitter.com/gpsshogi>

*2 <http://www.shogidojo.com/>

ある [9] では、画像とそれを説明するタグが与えられたとき、これらを Restricted Boltzmann Machine (RBM) で学習することで与えられた画像からタグを生成する、あるいは与えられたタグがつけられそうな画像を選び出すシステムである。RBM は教師なしのニューラルネットワークの 1 つで、訓練データを復元するように学習を行う。入力の一部が欠けている場合でもそれを復元するようなモデルであるが、入力として画像を表現するベクトルとタグを表現するベクトルを結合したものをを用いる。例えば画像からタグを生成する場合には、タグを表現するベクトルが欠けているものとみなして画像を表現するベクトルのみを入力すると、タグを表現するベクトルが復元される。この研究はまた、画像とタグそれぞれを独立に RBM で学習を行った上で隠れ層を入力として用いて RBM を学習することで、より高い精度となったことを報告している。画像や自然言語を入力する前にそれぞれに特有の前処理を行い特徴ベクトルに変換するが、その後のネットワーク内では画像特有、あるいは自然言語特有の処理を行うわけではなくベクトル間の関係を扱っている。そのため将棋の局面についても同様に特有の処理を行い局面を表現するのに十分な特徴ベクトルを用いることで、同様のネットワークを用いることができると考えられる。

自然言語生成の問題は大きく何を話すかとどのように話すかの 2 つに分けられるが、画像から単語を生成するような問題はこのうち何を話すかを決定していることになる。話す内容が決定され、そこから自然言語文を生成する研究も行われている。前述の天気予報の例のようにテンプレートベースの手法もあるが、別の手法の 1 つとして言語モデルを用いるものがある [10]。言語モデルとはある単語列 w_0, w_1, \dots, w_{i-1} があつたとき、次に w_i が続く確率を定義したものである。具体的には単語列とその他の状態 $attr_i$ から $P(w_i | w_0, w_1, \dots, w_{i-1}, attr_i)$ が定義される。このときある文の生成確率は

$$P(w_0, w_1, \dots, w_n) = \prod_{i=0}^n P(w_i | w_0, w_1, \dots, w_{i-1}, attr_i) \quad (1)$$

で定義され、これが大きい文はより生成されやすい文である。

他にも文法ベースの手法があげられる。自然言語処理の分野でよく用いられる文法の 1 つに文脈自由文法 [11] と呼ばれるものがある。これは文の各要素を展開する文法セットを持ち、展開していくことで文法規則に反しない自然言語文が生成されるというものである。この問題を重み付きグラフ問題と捉え解く研究 [12], [13] がなされ、一定の成果をあげている。また何を話すかとどのように話すかを決定する手法を組み合わせる自然言語生成がロボカップの実況 [14] や天気予報の分野 [15] に適用され、人間による解説には及ばないものの発話内容の選択と発話の自然さが

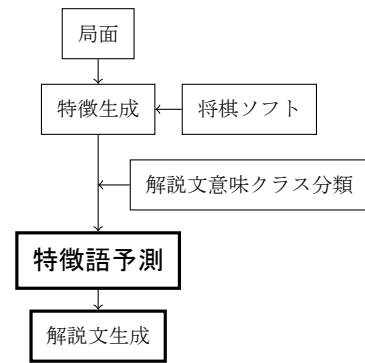


図 1 解説文生成システム

ともに人間のそれに近いものが生成できたことが報告されている。

3. 提案手法

3.1 解説文生成システムの全体像

本研究が想定している解説文生成システムの全体像を図 1 に示す。事前にシステムが生成すべき意味クラスを定め、各クラスにおいて特徴語の予測器を学習しておく。与えられた局面に対してどの意味クラスの解説文を生成すべきかを決定し、対応する予測器に入力する。予測器は特徴語数次元のベクトルを返す。このベクトルが生成すべき解説文が持つ意味を表現するものとし、対応する解説文を生成する。

本研究ではこのうち予測器の学習と解説文の生成を試みる。特に予測器の学習は将棋の局面と自然言語である解説文の関係を取る、解説文生成の根幹をなすものであるといえる。

3.2 解説文の意味による分類

特徴語の予測に先立ち、解説文の意味による分類を行う。

将棋の解説文の内容は、次の一手の予想や戦型の解説、対局者や会場の様子など、多岐にわたる。このうち対局者や会場の様子など、将棋のゲームとしての面に関係しない描写は将棋プログラムに出力を期待される解説文ではないといえる。またゲームに直接関係する解説文であっても、例えば次の一手の予想と戦型の解説を同時に行う必要はなく、それぞれ独立の予測システムを構築できると考えられる。

そこで本稿では、解説文をその意味でクラス分けする分類器を提案する。例えば将棋の解説文においては、次の指し手に言及するクラスや形勢について言及するクラス、囲いや戦法について言及するクラスなどが考えられる。以下ではこれを意味クラスとよぶことにする。解説文の意味クラスは、例えば「矢倉」や「穴熊」といった単語は戦型に言及する文章に多く現れるなど、現れる単語と強く関係するだろうことが期待される。そこで入力に用いる特徴は、

それぞれの文に各単語が何度出てきたかをベクトルとして表現する Bag of Words を用いる。学習器としてはパーセプトロンを用いる。各解説文には人手による意味クラスのアノテーションが必要となるが、アノテーション数を少なくするために能動学習 [16] を行う。

3.3 特徴語の予測

本稿ではある局面が与えられたときにその局面を解説する際に現れそうな単語を予測するシステムを提案する。

入力として、局面を表現する特徴ベクトルを用いる。用いた特徴は盤上の各駒の位置・持ち駒と、コンピュータ将棋プログラム「激指」が評価関数で用いている特徴からなる。

出力として、教師値は対象となる各単語が解説文中に現れるまたは現れないことを示すベクトルを用いる。予測する際は各単語について 0 から 1 までの範囲の値を持つベクトルを返す。

学習器には 3 層ニューラルネットワークを用いる。ニューラルネットワークは各単語の現れやすさを表現する対象単語数次元のベクトルを出力するよう学習する。

3.4 解説文の生成

前項で得られた単語の予測結果は、その局面において解説すべき内容を表現すると考えられる。そこで、単語の予測結果を用いた言語モデルを作成し解説文を生成する手法を提案する。

入力として局面の特徴・直前の 2 語に加え、上記の予測結果を用いる。出力の教師値は学習データ中に実際に現れた単語あるいは終端記号のみが 1 でその他が 0 のベクトルを用いる。これによりある局面・単語列に続く次の 1 語としてそれぞれの単語がどの程度現れやすいかが得られることが期待される。

4. 評価

本稿では解説の対象を「戦型に言及する文」に限定し、提案手法の評価を行った。なおここでいう「戦型に言及する文」とは、「矢倉」や「穴熊」といった自玉の囲いの話題と、「四間飛車」や「腰掛け銀」といった戦法の話者を指すものとする。

4.1 対象とする棋譜

学習には、有料で配信されているプロ棋士の解説付き棋譜（順位戦と名人戦）*3 のうち 2011 年以前のものを用いた。対象となるのは 104,903 局面・349,750 文である。ただし注釈や解説者の雑談など、将棋の解説ではない文も多く含む。

4.1.1 棋譜についているコメントの考察

本稿で対象棋譜としたのは順位戦と名人戦のものである。順位戦は A 級・B1 級・B2 級・C1 級・C2 級の 5 段階のクラスに分かれており、この中でリーグ戦により順位を決める棋戦である。A 級の優勝者が名人戦挑戦者となり、名人のタイトルをかけた 7 番勝負を行う。タイトル称号がかかっている名人戦は将棋ファンの間でも特に注目の対局であり、その挑戦者を決める A 級リーグも注目度は高い。それに比べると下位クラスの対局は注目度が低い。そのためか公開されている棋譜についている解説コメントの量も、クラスにより差が現れている。また最近のものほどコメント量が多いという特徴もみられる。表 1 にその数を示す。

クラスはまたコメントの性質とも大きく相関がある。棋譜につけられているコメントは、将棋の解説文の他にその局面までの互いの消費時間や棋士の食事の内容など、局面の解説と関係のないコメントも多い。表 2 に、いかにもゲームと関係のなさそうな単語を含むコメントの数を示す。下位のクラスではそのようなコメントに対してゲームを解説するコメントが少なく、上位のクラスや名人戦の棋譜に比べて相対的に有効なコメントが少ない傾向にある。

これから、学習データ中の有効なコメントの割合を高める 1 つの方法として、下位クラスの棋譜を扱わないことが考えられる。有効コメントの割合が高くなることは期待できるが、有効コメントの絶対数は減少する。学習に用いるコメント数が不十分であるとすると、下位クラスの棋譜を一括して対象外とするか否かの検討が必要である。

4.2 解説文の分類

解説文を「戦型に言及する文」か否かで 2 クラス分類を行う分類器を学習した。形態素解析には KyTea [17] を用いた。一般に形態素解析に用いられる MeCab [18] と比べ、KyTea は分野適応しやすいという特徴がある。本実験ではアノテーションを行った将棋の解説文を用いてチューニン

表 1 順位戦・名人戦のクラスとコメント量の関係（コメント行数/局数）*4

期	名人戦	A	B1	B2	C1	C2
70 期	1,979/7	8,259/45	6,816/78	8,235/120	11,363/164	13,323/217
69 期	1,185/4	8,124/45	6,392/78	6,484/120	8,407/157	10,218/213
68 期	1,971/7	8,213/45	6,927/78	6,801/120	7,622/155	8,615/217
67 期	1,234/6	7,126/45	4,898/78	5,534/117	7,156/155	8,013/215
66 期	1,382/7	5,359/45	4,513/78	4,422/109	5,606/144	7,548/225
65 期	728/6	4,388/46	2,875/78	3,373/115	4,461/140	5,848/227
64 期	0/7	720/45	372/78	425/115	693/148	1,007/228

表 2 順位戦・名人戦のクラスとコメントの性質の関係

単語	名人戦	A	B1	B2	C1	C2
ここまでの消費時間は	132	912	1,607	2,555	3,115	4,627
残り時間は	123	686	762	1,142	1,427	1,745
昼食の注文は	29	195	205	294	355	467
夕食の注文は	6	165	195	257	264	363

*4 棋譜中のコメント行数。1 行に複数文が含まれているものや空の行もあるため、文の数と厳密には一致しない。

*3 <http://www.meijinsen.jp/>

グを行ったものを用いている。また「▲7六歩」や「△8八角」といったような指し手表現は局面に強く依存するものが多いと予想されるため、全て名詞「MOVE」に置換し同一の単語とみなした上で形態素解析を行った。

能動学習の際には、まずランダムな50文にアノテーションを行った。その後平均化パーセプトロンにより2クラス分類器を学習し、2クラス間のスコア差が最も小さい文を選び出しアノテーションを行った。以上の操作により、最初の50文を含めて1,000文のアノテーションを行い学習時の教師データとした。

分類器が「戦型に言及する文」とした解説文200文を取り出し実際に戦型に言及しているか確認したところ、162文は実際に戦型に言及していた。解説文をランダムに500文抽出して同様のクラス分けを行うと戦型に言及している文は20文しか存在しなかったことから、分類器が戦型に言及するとした文はある程度信頼できることが分かる。

4.3 単語の予測

与えられた局面に対して現れうる単語を予測する予測器を学習した。評価実験の対象とする局面は、4.2により学習された分類器が「戦型に言及する文」であるとした解説文が付与されている局面とした。このうち8割を学習用局面、2割を評価用局面とした。予測の対象となる単語は、学習に用いる棋譜中に多く出現しておりかつ「戦型に言及する文」に特有であると考えられる単語とした。本稿で対象とした単語を表3に示す。

各単語についてネットワークの出力を元に2値分類し、評価用局面に対して対象単語が解説文に現れるか否かと実際にはどうであったかを見た。分類には3層パーセプト

表3 対象単語と出現回数

単語	出現回数
穴熊	2,524
矢倉	1,357
囲い	1,174
美濃	1,087
換わり	679
振り	598
ゴキゲン ^{*5}	523
銀冠	471
四間	455
模様	396
中飛車	376
石田	319
三間	290
腰掛け	274
棒銀	243
横歩	177
居飛車	156

ロンを用いた。ニューラルネットワークの実装には Fast Artificial Neural Network Library (FANN) [19] を用いた。入力とする特徴は学習用局面に現れたもののみを用い、評価用局面のみに現れた特徴は全てないものとみなした。入力は4,547次元、出力は17次元となった。同様の学習を、対象単語を学習用局面に付与された対象解説文中の全ての単語に拡張して行った。その際の出力は3,852次元となった。ニューラルネットワークは隠れ層を50次元とした。また隠れ層はハイポボリックタジェント、出力層はシグモイド関数を通し、出力の範囲を[0, 1]とした。出力は[0, 1]の連続値であるが、一定の閾値以上であるか否かで現れるか否かの2値分類問題と見て、F値を比較した。閾値はそれぞれのネットワークごとにF値の高いところとした。結果を表4, 5に示す。

False Positive が大きく、本来発話すべきでない局面での発話が非常に多い。

全ての出力を合わせると、17語に限定したものはF値0.30、全ての単語を対象としたものはF値0.38となった。全ての単語を対象にしたネットワークの方が平均して高いF値を示し、単語が出てくるか否かをより高い精度で当てていることが分かる。しかし2値分類問題として見たときこのF値はあまり高いとはいえず、単語が出現するか否かの2値分類問題としてとらえることは難しいことが分かる。

出力がどの程度元の文と近いかを見るのに、normalized Discounted Cumulative Gain ($nDCG$) [20] を用いる。 $nDCG$ はWeb検索エンジンの評価のために考案された指標で、関連度の高いものが高い順位であることを評価する。具体的には p 個の報酬値 rel_i とそれらの正しい順位が与えられたときその報酬値がどの程度正しいかを評価するもので、

表4 出現単語予測の精度 (17語に限定) 閾値0.34

単語	TP	FN	FP	TN	Pre	Rec	F
四間	22	51	163	1,801	0.12	0.30	0.17
銀冠	12	88	131	1,806	0.08	0.12	0.10
石田	3	62	91	1,881	0.03	0.05	0.04
振り	11	178	121	1,727	0.08	0.06	0.07
腰掛け	1	67	92	1,877	0.01	0.01	0.01
棒銀	2	54	137	1,844	0.01	0.04	0.02
三間	1	57	102	1,877	0.01	0.02	0.01
美濃	39	173	155	1,670	0.20	0.18	0.19
矢倉	139	88	413	1,397	0.25	0.61	0.36
囲い	33	190	138	1,676	0.19	0.15	0.17
換わり	81	73	120	1,763	0.40	0.53	0.46
穴熊	382	88	422	1,145	0.48	0.81	0.60
横歩	1	40	100	1,896	0.01	0.03	0.01
模様	6	61	85	1,885	0.07	0.09	0.08
ゴキゲン	44	55	81	1,857	0.35	0.44	0.39
中飛車	81	66	155	1,735	0.34	0.55	0.42
居飛車	2	84	114	1,837	0.02	0.02	0.02

*5 KyTea の実際の出力は「ゴ/名詞/ゴキゲン/名詞/きげん」

$$DCG_p = rel_1 + \sum_{i=2}^p \frac{rel_i}{\log_2(i)} \quad (2)$$

$$nDCG_p = \frac{DCG_p}{IDCG_p} \quad (3)$$

で定義される。IDCG_p は与えられた報酬値が正しい順位の通りに並んでいると仮定したときの理想的な DCG_p である。この値が 1 に近ければ近いほど、報酬値の順位が正当なものであるといえる。

ここでは、前項で挙げた 17 個の単語の順位についての nDCG を比較する。出力する単語の数を 17 個に制限した場合 nDCG = 0.923 であるのに対し、学習データ中の全ての語彙を対象にした場合 nDCG = 0.950 と、制限するよりよい結果になった。これは学習器に 3 層パーセプトロンを用いていることによると考えられる。ネットワークが中間層を持つことで複数の出力が共起するだろうことが予想される。

単語の数を制限しないものについてスコアの一番高い単語がどれほど正解であるかを判定すると、その正解率は 0.997 となり、特にスコアの一番高い単語については信頼できるものであるということがわかった。

4.4 解説文生成モデルの作成

解説文を生成するにあたり、単純な言語モデルの作成を試みた。単語予測に用いた特徴は前項で用いた現局面の特徴と直前の 2 語、前項で得た単語予測ベクトルである。学習にはパーセプトロンを用いた。出力する単語は貪欲法で、最も出現確率の高いものを選択する。直前の 1 単語まで正確であったとしたときに次の 1 単語を出力する実験を行った。

出力を見ると、ほとんどが文章にならない単語の羅列

表 5 出現単語予測の精度 (全単語対象) 閾値 0.25

単語	TP	FP	FN	TN	Pre	Rec	F
四間	24	49	48	1,916	0.33	0.33	0.33
銀冠	15	85	75	1,862	0.17	0.15	0.16
石田	0	65	4	1,968	0	0	0
振り	45	144	158	1,690	0.22	0.24	0.23
腰掛け	24	44	38	1,931	0.39	0.35	0.37
棒銀	2	54	7	1,974	0.22	0.04	0.06
三間	23	35	43	1,936	0.35	0.40	0.37
美濃	59	153	283	1,542	0.17	.28	0.21
矢倉	111	116	272	1,538	0.29	0.49	0.36
囲い	58	165	210	1,604	0.22	0.26	0.24
換わり	104	50	65	1,818	0.62	0.68	0.64
穴熊	362	108	443	1,124	0.45	0.77	0.57
横歩	1	40	17	1,979	0.06	0.02	0.03
模様	0	67	9	1,961	0	0	0
ゴキゲン	62	37	41	1,897	0.60	0.63	0.61
中飛車	85	62	48	1,842	0.64	0.58	0.61
居飛車	1	85	10	1,941	0.09	0.01	0.02



図 2 相振り飛車の局面



図 3 早繰り銀 vs. 腰掛け銀の局面

となってしまう。その中で「一/手/損/角/換わり/と/な/つ/た/。」という解説のついている局面についてはほぼ正確に同じ文を生成していた。「一手損角換わりとなった」というコメントのつく局面は多く、決まり文句となっているものについては出力できていると考えられる。他にも、図 2 の局面に対して、前半部分は正しく出力できていないが「飛/と/振/つ/て」に続いて「相/振/り/飛車/が/確定/し/た/。」という解説文がついている局面に「飛車が」が「飛車で」になっただけの出力が得られた。文として正しい形にはなっていないが、早繰り銀・腰掛け銀・棒銀の 3 すくみを解説している局面において棒銀や腰掛け銀という単語を出力する局面も見られた。図 3 の局面がそれであるが、この局面は棒銀ではない。そのような局面に対して棒銀という単語を生成していることから、局面を直接表現するのではなく解説に現れやすい単語を生成していることが伺える。

4.5 考察

本研究ではある局面が入力されたときに特徴語を予測す



図 4 矢倉 3七銀 (△6四角まで)

る問題に取り組んだが、問題を大幅に簡単にしている。具体的には対象局面に戦型に言及する文がついていることが分かっており、かつ頻出単語のみに絞っている。それにも関わらず単語の予測が期待した程度の精度ではできていない。1つにモデルや特徴の設定の調整不足によるものであると考えられる。また問題自体の難しさとして、盤外の環境に解説文が大きく左右される点もあると考えられる。

一例として、矢倉 3七銀型の定跡^{*6}で始まる棋譜を見る。対象棋譜は 41 局存在し、囲いの定跡中であることから戦型に関する解説文も多く付与されている。このうち 25 手目▲3七銀がこの定跡の名前になっている指し手だが、この局面に対するコメントに、「後手が指せると見られていたが近年では先手の勝率があがってきた」という旨のコメントが見られた。これは読みによるものではなく棋界の流行や実戦の結果をもとにしており、将棋プログラムが考慮するところではない。棋譜から将棋の解説を生成する問題であると捉える上では、こういった解説を出力することはモデルに含まれず、そもそも解けない問題である。こういった解説文をどのように扱うかの検討が必要であると考えられる。

5. おわりに

本稿では将棋の解説文生成モデルの全体像を提案した。またその一部である単語の予測について具体的なモデルを含め提案した。学習には人がつけた解説文を用い、簡単な分類器と能動学習を行うことで、低いコストで学習に用いる解説文を選択できることを示した。単語の分類問題として考えるとその精度はあまり高くないが、単語の予測が発話の意味を表現すると仮定し、提案手法を用いることで実

^{*6} 手順が前後して同一局面になることもあるが、ここでは▲7六歩△8四歩▲6八銀△3四歩▲6六歩△6二銀▲5六歩△5四歩▲4八銀△4二銀▲5八金右△3二金▲7八金△4一玉▲6九玉△5二金▲7七銀△3三銀▲7九角△3一角▲3六歩△4四歩▲6七金右△7四歩▲3七銀△6四角(図4)に手順を含め完全に一致している棋譜のみを見た。

際の発話内容に近いベクトルを獲得した。解説文生成のためにバイグラム言語モデルを作成したが、自然な文章の生成には至らなかった。

発話の意味の予測は実際につけられた解説文にある程度近いベクトルが得られたとはいえ、学習の対象を戦型に言及する文という狭い範囲に限ったことを考慮すると期待したほどの精度は得られなかった。意味クラス分類を精査する、あるいは予測のモデルを検討する必要があると考えられる。

解説文の生成について、第一に日本語の言語生成モデルを考える必要がある。本稿の実験では全て形態素を単位としてのトライグラムを用いたが、特に日本語では助詞の連続など、単純なトライグラムでは十分にモデル化できないだろうと考えられる。本研究では将棋の解説文という非常に狭い範囲を対象としているため、語彙数は少ない一方で最小単位を大きくすると空間が極端にスパースになってしまう恐れがある。将棋の解説文を表現しうる言語モデルを検討したのちに局面を表現する特徴と組み合わせることが今後の課題である。

参考文献

- [1] Sadikov, A., Možina, M., Guid, M., Krivec, J. and Bratko, I.: Automated chess tutor, *Proceedings of the 5th international conference on Computers and games, CG'06* (2007).
- [2] 保木邦仁: 局面評価の学習を目指した探索結果の最適制御, *ゲームプログラミングワークショップ 2006 論文集*, pp. 78-83 (2006).
- [3] 金子知適: コンピュータ将棋を用いた棋譜の自動解説と評価, *情報処理学会論文誌*, Vol. 53, No. 11, pp. 2525-2532 (2012).
- [4] Takeuchi, S., Kaneko, T., Yamaguchi, K. and Kawai, S.: Visualization and Adjustment of Evaluation Functions Based on Evaluation Values and Win Probability, *22nd national conference on Artificial intelligence (AAAI '07)* (2007).
- [5] Sripada, S., Reiter, E. and Davy, I.: SumTime-Mousam: Configurable marine weather forecast generator, *Expert Update*, Vol. 6, No. 3, pp. 4-10 (2003).
- [6] DeVault, D., Traum, D. and Artstein, R.: Making grammar-based generation easier to deploy in dialogue systems, *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pp. 198-207 (2008).
- [7] Binsted, K., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., Pain, H., Waller, A. and O'Mara, D.: Computational humor, *Intelligent Systems, IEEE*, Vol. 21, No. 2, pp. 59-69 (2006).
- [8] Huiskes, M. J. and Lew, M. S.: The MIR Flickr Retrieval Evaluation, *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval* (2008).
- [9] Srivastava, N. and Salakhutdinov, R.: Multimodal Learning with Deep Boltzmann Machines, *Advances in Neural Information Processing Systems 25 (NIPS2012)*.
- [10] Ratnaparkhi, A.: Trainable approaches to surface natural language generation and their application to conver-

- sational dialog systems, *Computer Speech & Language*, Vol. 16, No. 3, pp. 435–455 (2002).
- [11] Chomsky, N.: Three models for the description of language, *Information Theory, IRE Transactions on*, Vol. 2, No. 3, pp. 113–124 (1956).
 - [12] Konstas, I. and Lapata, M.: Concept-to-text generation via discriminative reranking, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12 (2012).
 - [13] Konstas, I. and Lapata, M.: Unsupervised concept-to-text generation with hypergraphs, *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT '12* (2012).
 - [14] Chen, D. L. and Mooney, R. J.: Learning to sportscast: a test of grounded language acquisition, *Proceedings of the 25th international conference on Machine learning, ICML '08* (2008).
 - [15] Angeli, G., Liang, P. and Klein, D.: A simple domain-independent probabilistic approach to generation, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10*, pp. 502–512 (2010).
 - [16] Settles, B.: Active learning literature survey, *University of Wisconsin, Madison* (2010).
 - [17] Neubig, G., Nakata, Y. and Mori, S.: Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis., *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*, pp. 529–533 (2011).
 - [18] Kudo, T.: MeCab : Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.net/>, (online), available from (<http://ci.nii.ac.jp/naid/10019716933/>) (2005).
 - [19] Nissen, S. and Nemerson, E.: Fast artificial neural network library, *Available at leenissen.dk/fann/html/files/fann-h.html* (2000).
 - [20] Järvelin, K. and Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques, *ACM Trans. Inf. Syst.*, Vol. 20, No. 4 (2002).