着手者フィルタと合法手提示を用いた将棋解説文生成手法の改良

大田 皐介[1,a)]　中沢 実[1]

概要：筆者らは，棋譜コメントの検索結果と将棋 AI の読み筋を大規模言語モデルに入力する将棋解説文生成手法を提案した．この手法には，検索された別局面の指し手を入力局面に適用する誤りや，先手と後手を混同する誤りがみられた．本稿では，直前の着手者に基づいて検索対象を限定する着手者フィルタと，入力局面の全合法手をプロンプトに付加する合法手提示を導入する．延べ 3,713 局面を対象とする自動評価では，両者を併用した改良手法の指し手符号整合率が基本手法を上回った．一方，読み筋被覆率の改善は確認できなかった．主観評価の結果，改良手法の平均評点は 4 項目すべてで高かったが，Wilcoxon の符号順位検定では，多重比較補正後の統計的有意差は認められなかった．生成例の分析では，改良手法で誤った駒取りや指し手への言及が減る例と，読み筋に基づく具体的な説明が欠落する例が確認された．これらの結果から，改良手法では指し手符号の整合性の向上と先後混同率の低下がみられた一方，読み筋に基づく具体的な説明が欠落する場合もあり，解説内容の改善には課題が残ることが示された．

キーワード：将棋，解説文生成，検索拡張生成

Improving Shogi Commentary Generation with Last-Mover Filtering and Legal-Move Prompting

Kosuke Ota[1,a)]　Minoru Nakazawa[1]

Abstract: We improve our shogi commentary generation method, which provides a large language model with retrieved game-record comments and engine variations. To reduce the misuse of retrieved moves and confusion between the two sides, we introduce last-mover filtering and legal-move prompting. These restrict retrieval by the preceding mover and add all legal moves to the prompt, respectively. Automatic evaluation on 3,713 position instances showed improved move-notation consistency and reduced side confusion, but no improvement in variation coverage. Human evaluation yielded higher mean ratings on all four criteria, but Wilcoxon signed-rank tests found no significant differences after multiple-comparison correction. Example analysis revealed fewer erroneous references to captures and moves in some cases, but omissions of concrete explanations based on engine variations in others. Thus, although the combined method improves formal consistency, preserving concrete explanations remains a challenge for improving commentary content.

Keywords: Shogi, Commentary Generation, Retrieval-Augmented Generation

1. はじめに

将棋 AI は局面の評価値や候補手を提示できるが，指し手の意味やその後の展開を理解するには将棋の知識を要する．自然言語による解説を併せて提示できれば，局面理解を支援できると期待される．一方，大規模言語モデル（以下，LLM）が自然な文章を生成しても，その説明が盤面や指し手の意味と一致するとは限らない．

筆者らは，局面と棋譜コメントの対応を学習した Shogi-CLIP によりコメントを検索し，将棋 AI の読み筋とともに LLM に入力する手法を提案した [1]．この先行研究の主観評価では，コメント検索を用いた手法は，読み筋を中心とする比較手法よりも分かりやすさ，正確性，有用性の平均評点が高かった．一方，先手を主語として後手の指し手を

[1] 金沢工業大学
Kanazawa Institute of Technology, Nonoichi, Ishikawa 921–8501, Japan
[a)] c6500562@st.kanazawa-it.ac.jp

© 2026 Information Processing Society of Japan

説明する，入力局面では指せない手に言及する，駒の配置を誤って記述するなどの問題がみられた．これらは，検索コメントと読み筋を用いて文章を生成する際に，入力局面との整合性が保たれないという課題である．

本稿では，直前の着手者が一致する局面に付随するコメントに検索対象を限定する着手者フィルタと，入力局面の全合法手をプロンプトに付加する合法手提示を提案する．本研究の目的は，検索コメントを利用する際の入力局面との不整合を抑制し，その効果が解説文の品質にも及ぶかを検討することである．着手者フィルタと合法手提示の有無を組み合わせた 4 条件について，生成文に含まれる指し手符号の整合性，読み筋への言及，先手・後手の混同を自動評価する．また，将棋経験者による主観評価を行い，基本手法と，着手者フィルタと合法手提示を併用した改良手法を，分かりやすさ，正確性，自然さ，有用性の 4 項目で比較する．さらに，手法間の評点差が大きい局面について生成例を分析し，形式的な整合性の改善と，解説内容の妥当性・具体性に残る課題を区別して検討する．

2. 関連研究と基本手法

2.1 将棋・チェスの解説文生成

亀甲らは，局面から特徴語を予測するモデルと対数線形言語モデルを組み合わせ，将棋解説文を生成した [2]．山内らは，棋譜の手順と将棋 AI の予測手順をテキスト化し，T5 による解説文生成に利用した [3]．同研究では，予測手順を生成文に含められる場合にも，指し手の意味を誤って説明する例が報告されている．本研究では，読み筋に加えて人手の棋譜コメントを検索し，指し手の意味を説明するための参考情報として用いる．

チェスでは，Kim らがチェスの専門モデルから抽出した概念情報を用いて LLM による解説生成を支援し，内容の正確性や情報量，言語表現の品質を評価している [4]．将棋でも，水匠チームは知識データベースから検索した情報を LLM に与える解説システムを提示した [5]．本研究は，検索コメントを利用する既存の構成に対し，着手者と合法手の情報を与えることで，入力局面と生成文との不整合の抑制を試みる．

2.2 検索拡張型の基本構成

基本構成は，以下の 4 つの処理からなる．

(1) 入力局面の特徴を Shogi-CLIP の局面エンコーダで表現する．
(2) 棋譜コメントの埋め込みとの類似度により上位 2 件を検索する．
(3) dlshogi の評価値と読み筋を取得する．
(4) コメント，読み筋，BOD 形式の局面図と直前の指し手を LLM へ与える．

本稿では，先行研究 [1] の学習済み Shogi-CLIP モデルを使用する．

以下では，着手者フィルタと合法手提示を用いない条件を基本手法，両処理を併用する条件を改良手法と呼ぶ．基本手法は先行研究 [1] で提案したコメント検索を含む構成に対応する．

3. 改良手法

3.1 着手者フィルタ

基本手法では，生成文中で先手と後手を混同する誤りがみられた．入力局面とは直前の着手者が異なる局面のコメントを提示すると，コメント中の指し手や狙いを，先手と後手を取り違えて説明に取り込む可能性がある．そこで，直前の着手者が入力局面と一致するコメントに検索対象を限定する着手者フィルタを導入する．

通常の対局では，入力局面の手番と反対側が直前の着手者となる．検索候補の局面も着手後の手番から直前の着手者を求め，入力局面の直前の着手者と一致する候補だけを残す．その候補集合内で類似度上位 2 件を取得する．これにより，解説対象と異なる側の着手に付随するコメントの参照を抑制する．

3.2 合法手提示

基本手法では，検索された別局面の指し手を入力局面に適用し，入力局面では指せない手に言及する誤りがみられた．そこで，入力局面で選択可能な指し手を明示するため，全合法手をプロンプトに付加する合法手提示を導入する．

cshogi[6] を用いて，入力局面で手番側が指せる全合法手を列挙し，指し手符号の一覧をプロンプトに付加する．これにより，検索コメント中の指し手を入力局面で指せるかどうかを判断するための情報を与え，指せない手への言及の抑制を図る．

4. 自動評価

4.1 実験設定と指標

着手者フィルタと合法手提示の有無を組み合わせ，両処理を用いない基本手法，着手者フィルタのみ，合法手提示のみ，両処理を併用する改良手法の 4 条件を比較した．各条件の生成文に含まれる指し手符号と先手・後手の記述を評価した．評価には，Shogi-CLIP の学習に用いた棋譜とは別に取得した，2025 年 6 月の順位戦 66 局の棋譜 [7] を用いた．対象は棋譜コメントが付与された延べ 3,713 局面で，序盤 667，中盤 1,559，終盤 1,487 局面である． 30 手目までを序盤，31 から 70 手目を中盤，71 手目以降を終盤とした．各条件で局面ごとに解説文を 1 件生成し，計 14,852 件すべてで応答を得た．生成には Llama-3-ELYZA-JP-8B を Ollama 0.20.0 で用い，temperature=0 とした．

生成文から，▲7 六歩などの KIF 形式の指し手符号を抽出した．局面 i の生成文から抽出した符号の出現列を P_i，

© 2026 Information Processing Society of Japan

表 1 指し手符号整合率／読み筋被覆率

条件	序盤	中盤	終盤
基本手法	0.716/0.387	0.615/0.417	0.564/0.369
着手者フィルタのみ	0.743/0.308	0.652/0.273	0.629/0.257
合法手提示のみ	0.799/0.329	0.729/0.324	0.720/0.327
改良手法	0.765/0.314	0.716/0.321	0.713/0.334

表 2 先後混同率と対象文数

条件	混同文数	対象文数	先後混同率
基本手法	383	3,388	0.113
着手者フィルタのみ	172	2,005	0.086
合法手提示のみ	150	2,299	0.065
改良手法	150	2,227	0.067

入力局面で手番側が指せる全合法手の符号集合を L_i，生成時に提示した読み筋の符号集合を R_i とする．指し手符号整合率 C と読み筋被覆率 V を次式で定義する．

$$C = \frac{\sum_i \sum_{t \in P_i} \mathbf{1}[t \in L_i \cup R_i]}{\sum_i |P_i|} \tag{1}$$

$$V = \frac{\sum_i |\,\mathrm{set}(P_i) \cap R_i|}{\sum_i |R_i|} \tag{2}$$

ここで，$\mathbf{1}[\cdot]$ は条件を満たす場合に 1，それ以外に 0 となる指示関数，$\mathrm{set}(P_i)$ は重複を除いた符号集合である．整合率では同一符号の反復を出現回数分数え，被覆率では同一符号を 1 件として数える．各指標は条件・進行段階ごとに分子と分母を合算して算出する．

さらに，先手または後手を主語とし，▲・△付きの符号を含む文について，主語に対応する節に逆側の符号が現れる割合を先後混同率とした．

これらは生成文の形式的な特徴に基づく指標である．指し手符号整合率では，指し手の狙い，取る駒，その後の展開に関する説明の妥当性は評価できない．また，符号の一致を調べるため，読み筋中の指し手を正しい局面や順序で説明していることも保証しない．読み筋被覆率は符号への言及の範囲を測る指標であり，指し手の意味を説明せずに符号を列挙しても高くなり得る一方，符号を用いない説明は捉えられない．先後混同率の対象も，所定の主語と符号を含む文に限られる．主語を省略した文や，符号を伴わない記述の誤りは対象に含まれないため，条件ごとの評価対象文数を併記する．

4.2 結果

表 1 に結果を示す．改良手法の指し手符号整合率は，序盤・中盤・終盤のすべてで基本手法を上回った．序盤では 0.716 から 0.765，中盤では 0.615 から 0.716，終盤では 0.564 から 0.713 に上昇した．一方，合法手提示のみの条件が全段階で最も高く，着手者フィルタの併用による整合率の上乗せは確認できなかった．読み筋被覆率は，基本手法の序盤 0.387，中盤 0.417，終盤 0.369 に対し，改良手法ではそれぞれ 0.314，0.321，0.334 であり，全段階で下回った．

表 2 に先後混同率を示す．基本手法と比べて，合法手提示のみの条件で先後混同率の低下が最も大きかった．着手者フィルタを併用した改良手法でも，合法手提示のみの条件に近い低い値を示した．これらの指標では捉えられない解説内容の妥当性を検討するため，次節以降では主観評価と生成例の分析を行う．

表 3 主観評価の項目

項目	質問内容
分かりやすさ	分かりやすい説明がされているか
正確性	局面を正確に説明できているか
自然さ	日本語として自然か
有用性	局面を理解するために役立つと思うか

5. 主観評価

5.1 対象局面と生成設定

アンケートによる主観評価では，前節の自動評価とは別に生成した解説文を対象とした．自動評価に用いた 66 棋譜から，相居飛車，対抗形，相振り飛車を各 1 局選定した．3 棋譜の各着手直後の計 337 局面について両手法の生成結果を保存し，各戦型の序盤・中盤・終盤から 2 局面ずつ，計 18 局面を選定した．

生成モデルは Llama-3-ELYZA-JP-8B であり，Ollama 0.20.0 を用いた．temperature=0 とし，各局面・条件につき 1 件の解説文を生成した．同一局面の両手法では，盤面，読み筋，システムプロンプト，モデル設定を共通とした．改良手法にのみ着手者フィルタと合法手提示を適用した．18 局面中 8 局面では着手者フィルタを適用しても検索コメントが変わらず，入力の差分は合法手一覧のみであった．

5.2 提示方法と分析単位

評価者は将棋経験者 14 名で，有段者 8 名，級位者 6 名であった．各局面の図と生成文を提示し，表 3 の 4 項目について，1 を最低，5 を最高とする 5 段階評価を求めた．自由記述も任意で収集した．

各評価者について，項目ごとに手法別の 18 局面の平均評点を求めた．この 14 組の平均評点に対して，両側の Wilcoxon の符号順位検定 [8] を行った．統計的な分析単位は評価者とし，局面ごとの評点を独立した標本としては扱わない．項目ごとに平均評点差が 0 の評価者を符号順位の計算から除き，差の絶対値が等しい場合は平均順位を用いた．その順位を固定し，符号の全組合せから両側 p 値を算出した．4 項目の p 値には Holm 法による多重比較補正 [9] を適用し，有意水準を 5%とした．

© 2026 Information Processing Society of Japan

表 4 14 名の平均評点と Wilcoxon の符号順位検定（両側）．差は改良 − 基本，補正は 4 項目に対する Holm 法

項目	基本	改良	差	p	補正 p
分かりやすさ	3.05	3.14	0.09	0.435	0.869
正確性	2.84	3.00	0.15	0.145	0.434
自然さ	3.79	3.85	0.05	1.000	1.000
有用性	2.74	2.93	0.19	0.034	0.137

表 5 棋力別の平均評点差（改良 − 基本）

項目	有段者 8 名	級位者 6 名
分かりやすさ	0.13	0.04
正確性	0.22	0.07
自然さ	0.20	−0.15
有用性	0.20	0.19

5.3 全体結果

表 4 に平均評点と Wilcoxon の符号順位検定の結果を示す．平均評点差は丸め前の値から算出した．改良手法の平均評点は全項目で基本手法を上回り，その差は分かりやすさ+0.09，正確性+0.15，自然さ+0.05，有用性+0.19 点であった．平均評点差が 0 でない評価者数は，分かりやすさ 11 名，正確性 12 名，自然さ 12 名，有用性 11 名であった．ただし，Holm 補正後はいずれの項目にも統計的有意差は認められなかった．

有用性は，未補正 p 値（$p = 0.034$）が 4 項目で唯一，有意水準 5%を下回ったため，評価者ごとの平均評点差を確認した．14 名中 8 名で改良手法の平均評点が高く，3 名で同点，3 名で基本手法が高かった．このように，平均評点の上昇は全評価者に共通するものではなかった．また，有用性の補正後 p 値は 0.137 であり，本評価の結果から改良手法の優位性を結論付けることはできない．

5.4 棋力別の比較と評価者間の一致度

表 5 に棋力別の平均評点差を示す．有段者群では改良手法の平均評点が全項目で高く，級位者群では自然さのみ低かった．

評価者間で解説文の評価順位がどの程度一致するかを調べるため，Kendall の一致係数 W[10] を算出した．分析には，18 局面それぞれについて基本手法と改良手法で生成した計 36 解説文の評点を用い，同順位を補正した．全解説文に同じ評点を付けた評価者も含め，全体 14 名，有段者 8 名，級位者 6 名の結果を表 6 に示す．有段者群の W は全項目で級位者群より数値上高い結果となった．これらは解説文の評価順位の一致度であり，基本手法と改良手法のどちらを支持するかの一致度ではない．

6. 生成例の分析と考察

6.1 分析対象

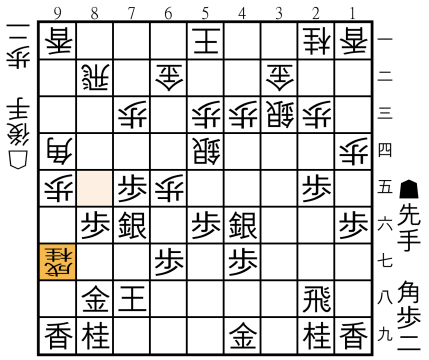
手法間の平均評点差が大きかった局面のうち，改良手法の評点が高い図 1・図 2 の局面と，基本手法の評点が高い図 3 の局面を分析対象とした．図 1〜図 3 に対象局面と両手法の生成文全文を，表 7 に平均評点差を示す．

表 6 36 解説文の評価順位に関する Kendall の一致係数 W

項目	全体 14 名	有段者 8 名	級位者 6 名
分かりやすさ	0.19	0.28	0.20
正確性	0.20	0.25	0.23
自然さ	0.15	0.25	0.19
有用性	0.19	0.26	0.25

表 7 図 1〜図 3 の局面の平均評点差（改良 − 基本）

項目	図 1	図 2	図 3
分かりやすさ	0.79	0.93	−0.71
正確性	1.21	1.36	−0.79
自然さ	0.71	0.36	−1.14
有用性	1.00	1.21	−1.07



(a) 基本手法

後手の△ 9 七桂成は、先手の歩を取り、成桂を生み出しています。先手は桂を失いましたが、桂を取り返す手段を探しています。先手は桂を取り返すことを目指しています。

先手の▲ 7 七金は、金を動かすことで、後手の角を取り、銀を取ります。先手は角と銀を取り、優位に立つことを狙っています。

形勢は互角です。

(b) 改良手法

後手の△ 9 七桂成は、先手の歩を取り、成桂を生み出しています。先手は桂を取ることで、後手の攻めを止める狙いがあります。後手は桂を成ることで、先手の陣形を崩すことを目指しています。

先手の▲ 7 七金は、金を動かすことで、後手の攻めを止める狙いがあります。後手は角を打つことで、先手の陣形を崩すことを目指しています。先手は金を動かすことで、後手の攻めを防ぐことを狙っています。

形勢は互角です。

図 1 相居飛車の中盤と両手法の生成文．局面は第 84 期順位戦 A 級 1 回戦，中村八段対永瀬九段の棋譜（名人戦棋譜速報 [7]）より引用．

6.2 図 1 の局面：駒取りと駒打ちに関する記述

図 1 の局面は△ 9 七桂成と後手が先手の歩を取った局面である．両手法に共通して入力した読み筋は，▲ 7 七金△ 7 六角▲ 9 七桂であった．

基本手法では「先手の▲ 7 七金は、金を動かすことで、後手の角を取り、銀を取ります。」と生成された．しかし，▲ 7 七金は駒を取る手ではないため，このような駒取りに関する誤記述が，基本手法の正確性の評点を下げた一因と考えられる．

改良手法では，▲ 7 七金について「後手の攻めを止める狙いがあります。」と述べ，角と銀を取るという誤記述はみられなかった．▲ 7 七金は 7 六の銀に金の利きを加える手であり，この生成文は金による守りを抽象的に説明したものと解釈できる．誤った駒取りの説明がなくなり，守りの狙いに触れたことが，改良手法の正確性の平均評点（3.21）が基本手法（2.00）を上回った一因と考えられる．一方，有段者の自由記述では，改良手法にも具体的な手順や攻め方の説明を求める指摘があり，説明の具体性には課題が残った．

着手者フィルタの適用により，改良手法では基本手法とは異なる「△ 3 三角と打って先手玉のコビンを狙いにいく手」というコメントが検索された．生成文にも「後手は角を打つことで、先手の陣形を崩すことを目指しています。」という記述がみられた．検索コメントと生成文は，角を打って先手側を攻めるという点で共通しており，検索コメントの変化がこの記述につながった可能性がある．

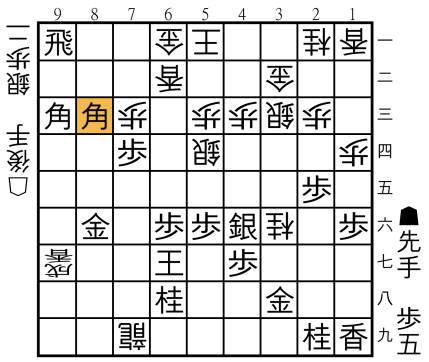
6.3 図 2 の局面：検索コメント中の指し手の適用

図 2 の局面は先手が▲ 8 三角と打った局面である．両手法の検索コメント，読み筋，局面情報は同一であり，入力の差分は合法手一覧のみであった．検索コメントには「防ぐには△ 7 三歩が手軽かつ堅い」という説明が含まれていた．

基本手法では「角が利いているため、後手は△ 7 三歩を指すことが難しくなっています。」と生成された．しかし，入力局面の 7 三には既に後手の歩があり，△ 7 三歩は合法手ではない．検索コメント中の指し手を，入力局面で選択できる手として扱った可能性がある．また，指し手の狙いに関する説明にも不整合がみられた．

改良手法では，非合法手である△ 7 三歩への言及はみられなかった．基本手法と改良手法の平均評点は，正確性がそれぞれ 2.29 と 3.64，有用性が 2.29 と 3.50 であった．非合法手への言及がなくなったことが，改良手法の正確性と有用性が基本手法より高く評価された一因と考えられる．

本例では，検索コメントに含まれる△ 7 三歩が，改良手法に付加した合法手一覧には含まれていなかった．合法手提示によって入力局面で選択できる指し手が明示され，△ 7 三歩への言及が抑制された可能性がある．一方，改良手法に対しても，どの金に角が利いているのかを具体的に示してほしいという自由記述があり，説明の具体性には課題が残った．



(a) 基本手法

先手の▲ 8 三角打は、後手の金に利いていますが、特に狙いはありません。角を打つことで、後手の陣形を乱すことを目指しています。角が利いているため、後手は△ 7 三歩を指すことが難しくなっています。

後手の△ 5 二銀打は、先手の角を狙っています。銀を打つことで、先手の角を取ることを目指していますが、先手は▲ 5 五歩で銀を退ける可能性があります。

形勢は互角です。

(b) 改良手法

先手の▲ 8 三角打は、後手の金に利いています。先手はこの角を生かして、後手の陣形を崩すことを狙っています。形勢は互角です。

図 2 相居飛車の終盤と両手法の生成文．局面は第 84 期順位戦 A 級 1 回戦，中村八段対永瀬九段の棋譜（名人戦棋譜速報 [7]）より引用．

6.4 図 3 の局面：読み筋に基づく具体的な説明の欠落

図 3 の局面は，先手が▲ 3 五歩と仕掛けた局面である．着手者フィルタを適用しても検索結果は変化せず，両手法で同じコメント「先手は早繰り銀からいきなり仕掛ける。」「早繰り銀から仕掛けを決断した。」が得られた．読み筋も共通で，△ 3 五歩▲ 3 五銀△ 5 四銀であった．両手法への入力は，合法手一覧の有無のみが異なっていた．

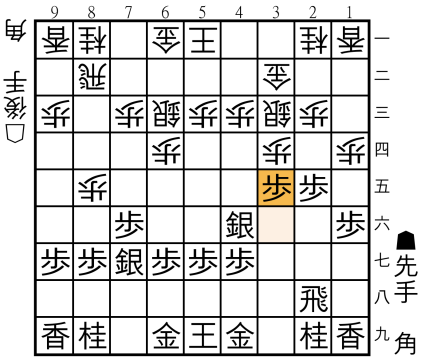
基本手法では「後手が△ 3 五歩と応じることで、先手は銀を動かして歩を取り、攻勢を強めようとしています。」と生成され，歩の交換と銀の進出を関連付けて説明していた．一方，改良手法では，先手・後手のいずれの指し手も相手の陣形を崩す狙いとして説明し，「後手は、先手の歩を取ることで、先手の陣形を崩すことを目指しています。」と記述していた．基本手法にみられた銀で歩を取り返すという具体的な説明が，改良手法では欠落していた．

基本手法と改良手法の平均評点は，自然さがそれぞれ 4.43 と 3.29，有用性が 4.00 と 2.93 であった．有段者の自由記述では，基本手法の説明が簡潔で分かりやすいと評価

© 2026 Information Processing Society of Japan

される一方，改良手法には後手の△3五歩を陣形を崩す手段とする説明の不適切さや，「手段」という語の反復に関する指摘があった．

合法手一覧の追加による入力情報の変化が，説明する指し手の選択に影響した可能性がある．本例では，▲3五銀による取り返しの説明の欠落と表現の反復が，自然さや有用性の評点を下げた一因と考えられる．



(a) 基本手法

先手の▲3五歩は、早繰り銀から仕掛ける手段で、先手が攻めの意欲を示しています。後手が△3五歩と応じることで、先手は銀を動かして歩を取り、攻勢を強めようとしています。形勢は互角です。

(b) 改良手法

先手の▲3五歩は、早繰り銀から仕掛ける手段です。先手は、後手の陣形を崩すことを狙っています。

後手の△3五歩は、先手の歩を取る手段です。後手は、先手の歩を取ることで、先手の陣形を崩すことを目指しています。

形勢は互角です。

図 3 相居飛車の序盤と両手法の生成文．局面は第 84 期順位戦 A 級 1 回戦，中村八段対永瀬九段の棋譜（名人戦棋譜速報 [7]）より引用．

6.5 自動評価・主観評価・生成例に基づく考察

改良手法では，指し手符号整合率が基本手法を上回り，先後混同率も低かった．これらは，入力局面と整合しない指し手や，着手者を取り違えた記述を抑制する可能性を示している．生成例でも，図 1・図 2 では基本手法にみられた駒取りや非合法手に関する誤記述がなくなり，正確性と有用性の評点が高かった．このことから，不適切な言及の抑制は，利用者が解説を正確で役立つと評価することにつながる可能性がある．

一方，読み筋被覆率は改良手法が全段階で基本手法を下回った．被覆率だけでは説明内容の不足を判断できないが，図 3 では後続手による取り返しの説明が欠落し，自然さと有用性の評点も低かった．したがって，改良手法では不適切な言及が減る一方で，局面理解に必要な展開説明まで省略される場合があると考えられる．合法手提示によって指せる手を明示することと，その手の狙いや後続手との関係を説明することは，分けて検討する必要がある．

主観評価全体では，改良手法の平均評点が 4 項目すべてで高かったものの，多重比較補正後の有意差は認められなかった．生成例でみられた誤記述の減少と具体的な説明の欠落は，形式的な整合性の改善が解説品質の一様な向上につながらない可能性を示している．ただし，自動評価と主観評価は異なる生成文を対象としており，指標値と評点の関係を直接検証したものではない．

また，自動評価では合法手提示のみの整合率が全段階で最も高く，先後混同率も最も低かった．図 1 では着手者フィルタによる検索コメントの変化が説明に影響した可能性はあるが，併用による追加の改善は明確ではない．今後は，検索コメントの入力局面への適用を考慮するとともに，誤記述を抑えながら具体的な展開説明を保つ生成方法を検討する必要がある．

7. 限界と今後の課題

主観評価は 14 名と 18 局面を対象とする探索的な評価であり，結果を他の評価者集団や対局に一般化するには追加の検証が必要である．棋力別の人数も少なく，棋力による効果の違いを検証するには追加の評価者が必要である．また，Wilcoxon の符号順位検定は評価者ごとの平均評点差の分布に対称性を仮定するが，本評価では差が 0 でない評価者が各項目 11 から 12 名と少なく，この仮定を十分に確認できていない．

主観評価では，評価者の棋力や局面の難易度に応じた解説の調整を行っていない．対抗形の中盤の局面では，有段者からも「局面が難しいです」という自由記述があり，評点には解説文の品質だけでなく，局面自体の理解の難しさも影響した可能性がある．今後は，棋力に応じて用語や説明の詳しさを調整した解説文を生成するとともに，評価者の棋力に応じた難易度の局面を選定し，解説が利用者の局面理解に役立つかを検証する必要がある．

8. おわりに

本稿では，検索コメントと読み筋を用いる将棋解説文生成手法に，着手者フィルタと合法手提示を導入した．4 条件の自動評価では，両処理を併用した改良手法の指し手符号整合率が基本手法を上回り，評価対象文における先後混同率は低かった．一方，読み筋被覆率の改善は確認できず，整合率は合法手提示のみの条件が全段階で最も高かった．別に生成した解説文に対する 14 名の主観評価では，改良手法の平均評点が全項目で高かったものの，Wilcoxon の符号順位検定では多重比較補正後に統計的有意差は認められなかった．したがって，本実験で確認された形式的な整合性の改善から，分かりやすさ，正確性，自然さ，有用性の改善まで結論付けることはできない．

© 2026 Information Processing Society of Japan

生成例の分析では，改良手法で誤った駒取りや指し手への言及が減る例と，読み筋に基づく具体的な説明が欠落する例が確認された．指し手符号の整合性に加え，局面情報と説明内容との対応を評価する必要がある．今後は，検索コメントの入力局面への適用を考慮し，読み筋に基づく手順や狙いを具体的に説明できる生成方法を検討する．また，棋力に応じて用語や説明の詳しさを調整した解説文の生成を検討する．

参考文献

[1] 大田皐介，中沢実：局面に対応したコメント検索を用いた大規模言語モデルによる将棋解説文生成，ゲームプログラミングワークショップ 2025 論文集，pp. 1–8 (2025).

[2] 亀甲博貴，三輪誠，鶴岡慶雅，森信介，近山隆：対数線形言語モデルを用いた将棋解説文の自動生成，情報処理学会論文誌，Vol. 55，No. 11，pp. 2431–2440 (2014).

[3] 山内悠輔，河原大輔：手順のテキスト化による将棋解説文生成，言語処理学会第 30 回年次大会発表論文集，pp. 3197–3202 (2024).

[4] Kim, J., Goh, J., Hwang, I., Cho, J. and Ok, J.: Bridging the Gap between Expert and Language Models: Concept-guided Chess Commentary Generation and Evaluation, Proc. NAACL-HLT 2025, pp. 9497–9516 (2025).

[5] 水匠チーム：水匠アピール文書，第 35 回世界コンピュータ将棋選手権，`https://www.apply.computer-shogi.org/wcsc35/appeal/Suisho/Suisho_Promote.pdf`（参照 2026-07-30）．

[6] 山岡忠夫：cshogi, `https://github.com/TadaoYamaoka/cshogi`（参照 2026-07-30）．

[7] 名人戦棋譜速報，`https://www.meijinsen.jp/`（参照 2026-10-16）．

[8] Wilcoxon, F.: Individual Comparisons by Ranking Methods, Biometrics Bulletin, Vol. 1, No. 6, pp. 80–83 (1945).

[9] Holm, S.: A Simple Sequentially Rejective Multiple Test Procedure, Scandinavian Journal of Statistics, Vol. 6, No. 2, pp. 65–70 (1979).

[10] Kendall, M. G. and Babington Smith, B.: The Problem of m Rankings, The Annals of Mathematical Statistics, Vol. 10, No. 3, pp. 275–287 (1939).

© 2026 Information Processing Society of Japan