

修士論文審査

解説文中の指し手表現を考慮した
将棋解説文の自動生成モデルの獲得

亀甲 博貴

指導教員：鶴岡慶雅 准教授

2015.02.06

自然言語生成

• 自然言語処理研究の課題のひとつ

0	3	6	9	12	15	18	21	24	
N	NW	W	W	S	S	SW	S	SW	
3	3	4	5	7	10	9	7	5	...
15	13	16	18	20	21	19	17	16	
⋮									



.....昼から夕方にかけて
強い南風が吹く。.....

[Sripada+, 2003]



これは白い布製の靴です。
滑り止めがあり.....

[Kiros+, 2014]

自然言語生成の課題

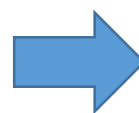
•何を話すか

- 観測値の要点
- 未来の予測
- etc.

•どうやって話すか

- 定型文
- テンプレート文
- 単語の組み合わせ
- etc.

内容があいまい
多様な表現を要する



将棋の解説文
に着目

将棋の解説文



矢倉模様の出だし
になりました。



プロ棋士

将棋：強い

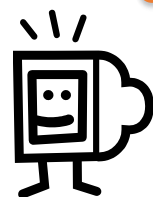
言葉：話せる

ファンには難解な将棋の理解を助ける

将棋の解説文生成



矢倉模様の出だし
になりました。



コンピュータプログラム
将棋：強い（プロに比肩）
言葉：話せる？

教師あり機械学習による生成モデルの獲得

発表の流れ

- 背景・目的
- 戦型に限定した解説文生成
- 指し手表現を考慮した解説文生成
- まとめ

将棋の解説文



(△7二銀まで)

現局面の説明

後手は美濃囲いに構えた。

未来の予測

ここでは▲6八銀と上がるのが
 有土相さわていす。▲7七角
特に戦型に言及するものに注目
 から八照を日相9のは下戦負け
 になりやすい。

データベース

これで前例と離れた。

盤外の様子

(先手) はさかんにお茶を
 飲んでいる。

戦型に言及する文の選択

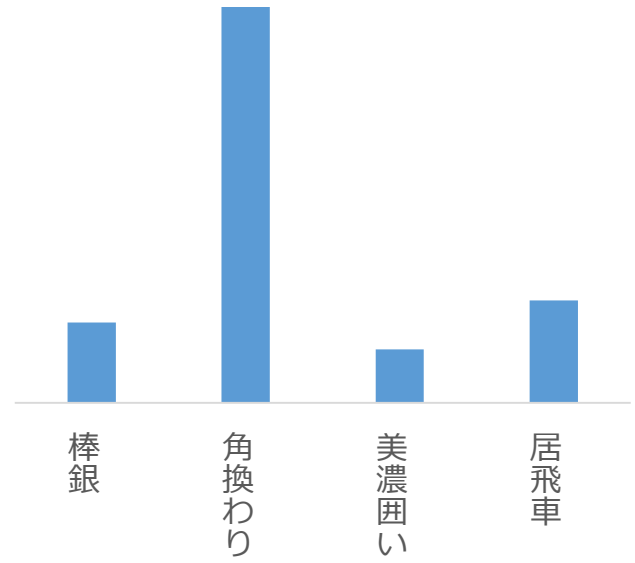
- 戦型に言及する文のみを抽出
 - 文中の各単語の頻度を元に分類
 - ex.) “構えた”などは関係しそう
 - 平均化パーセプトロンで学習
 - 能動学習で部分的にラベル付け
- 10,703文（/ 約340,000文）の抽出
 - 約8割は正しく戦法に言及

生成モデルの全体像

	9	8	7	6	5	4	3	2	1	
香	桂	銀	金	王		銀	桂	香		一
	飛					金				二
歩		歩	歩	歩	歩	歩	歩	歩		三
	歩					歩				四
							歩			五
		歩								六
歩	歩		歩	歩	歩	歩		歩		七
	留	金						飛		八
香	桂	銀		玉	金	銀	桂	香		九

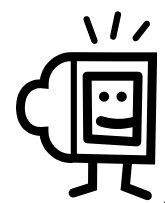
▲なし
▲なし

単語予測



言語モデル

文生成



一手損角換わり
となった。

単語予測

- 解説文中に現れうる単語の予測

The diagram illustrates a word prediction process. On the left is a crossword puzzle grid with columns numbered 1-9 and rows numbered 1-9. The grid contains various Japanese characters, including '桂', '銀', '玉', '金', '飛', '歩', and '香'. A large black box labeled '単語予測' (Word Prediction) is positioned over the grid. To the right of the box is a table with two columns: the predicted word and its score.

角	0.991
損	0.983
手	0.966
一	0.701

□ 評価関数を用いる特徴

- 単語が現れるか否かの分類器
- 本研究では3層パーセプトロン

文生成

- 言語モデルを用いた文生成

p : 局面 $S = w_1, w_2, \dots, w_n$: 文

$$P(S_n | p)$$

$$= \prod_i P(w_i | p, w_1, w_2, \dots, w_{i-1})$$

本研究では直前2単語

$$\times P(\text{length}(S) = n)$$

が最大になる S を求める

対数線形言語モデル

$$P(w_i | attr) = \frac{\exp\left(W_{w_i}^T \phi(attr)\right)}{\sum_{j \in V} \exp\left(W_{w_j}^T \phi(attr)\right)}$$

$attr$: 用いる特徴

直前の単語, 予測した単語, etc.

$\phi(attr)$: 特徴ベクトル

V : 語彙

生成された解説文の例

与えられた局面

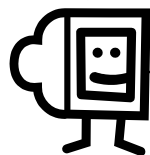


(▲7八銀まで)

現れる単語の予測結果

美濃	0.809	左	0.747
に	0.560	は	0.538

生成文



左美濃に構えた。

人間による解説文：後手の穴熊に対して先手は美濃囲いに囲った。

システムの生成文と人間による実際の解説文

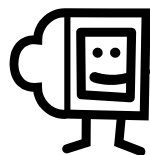
生成された解説文の例



(▲6八銀まで)

揃	0.613	条件	0.600
矢倉	0.531	に	0.530

生成文



矢倉になりそうだ。

人間による解説文：これで（先手）は左美濃に組むことはなくなった。

生成文の意味上・文法上の精度

		意味			
		Good	↔	Bad	小計
文法	Good	99	46	44	189
	↓	0	0	6	6
	Bad	0	0	5	5
小計		99	46	55	200

文法上は正しい文が多く生成できた

意味上の精度は
約5割と向上の余地あり

発表の流れ

- 背景・目的
- 戦型に限定した解説文生成
- 指し手表現を考慮した解説文生成
- まとめ

解説文中の指し手表現

△ 1 四玉 ▲ 3 五飛成
△ 同馬 ▲ 同角
△ 3 六飛 ▲ 4 三角
△ 3 五飛 ▲ 2 六金

指し手表現 (自然言語表現)
実際の指し手 (状態空間)
(▲ 2 四歩まで)

の対応付けが必要
(▲ 2 六金まで)

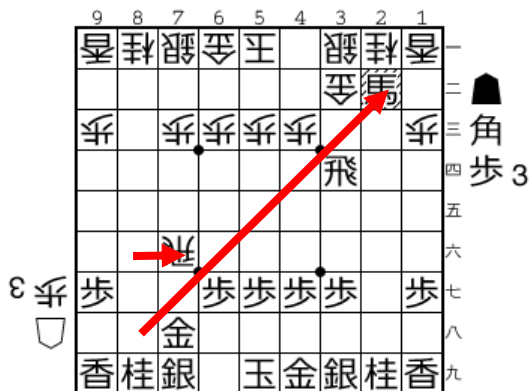
解説文

△ 1 四玉なら▲ 3 五飛成が好手という。以下△ 同馬▲ 同角△ 3 六飛▲ 4 三角で△ 3 五飛は▲ 2 六金と押さえる。

指し手表現を含む文の対応付け

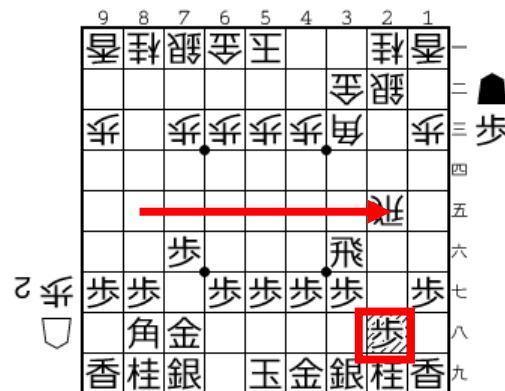
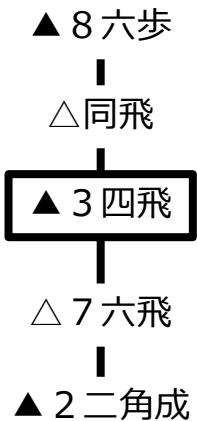
- **解説木**と**候補木**の概念を導入
 - 解説木：指し手表現と状態空間の対応関係を表現するツリー図
 - 候補木：解説木の候補

解説木



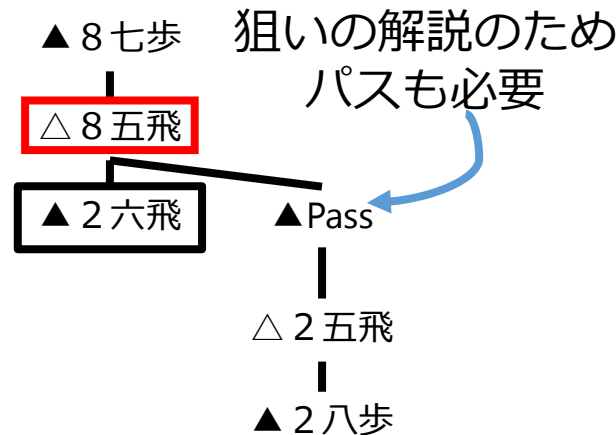
解説文

ここで△7六飛のマネ将棋は▲2二角成に取る駒がなく、終わってしまう。



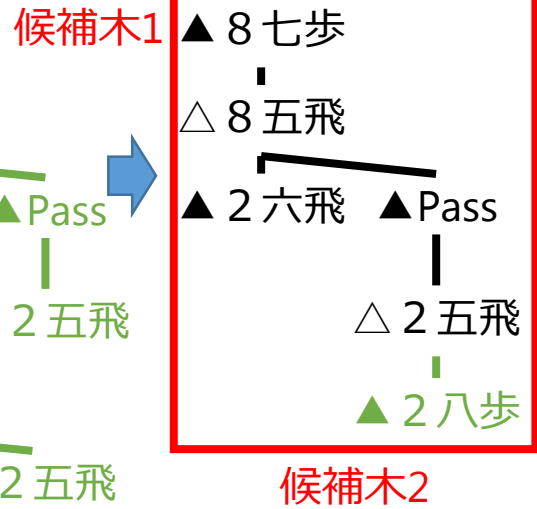
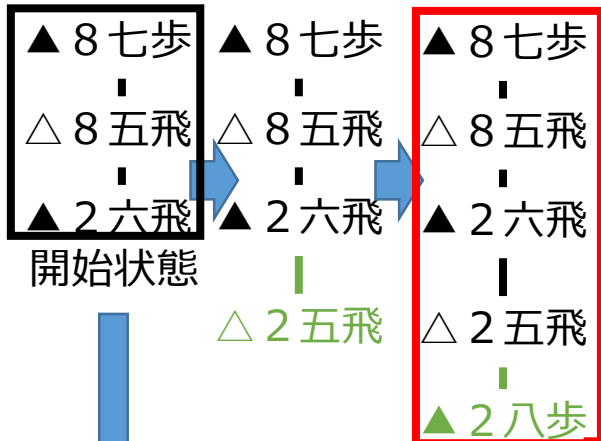
解説文

△2五飛と回られては▲2八歩と使わされてしまう。



合法手を繋げた木
→解説木

候補木



- 棋譜
- ▲ 3六飛
 - △ 2二銀
 - ▲ 8七歩
 - △ 8五飛
 - ▲ 2六飛



解説文

△ 2五飛と回られては▲ 2八歩と使わされてしまう。

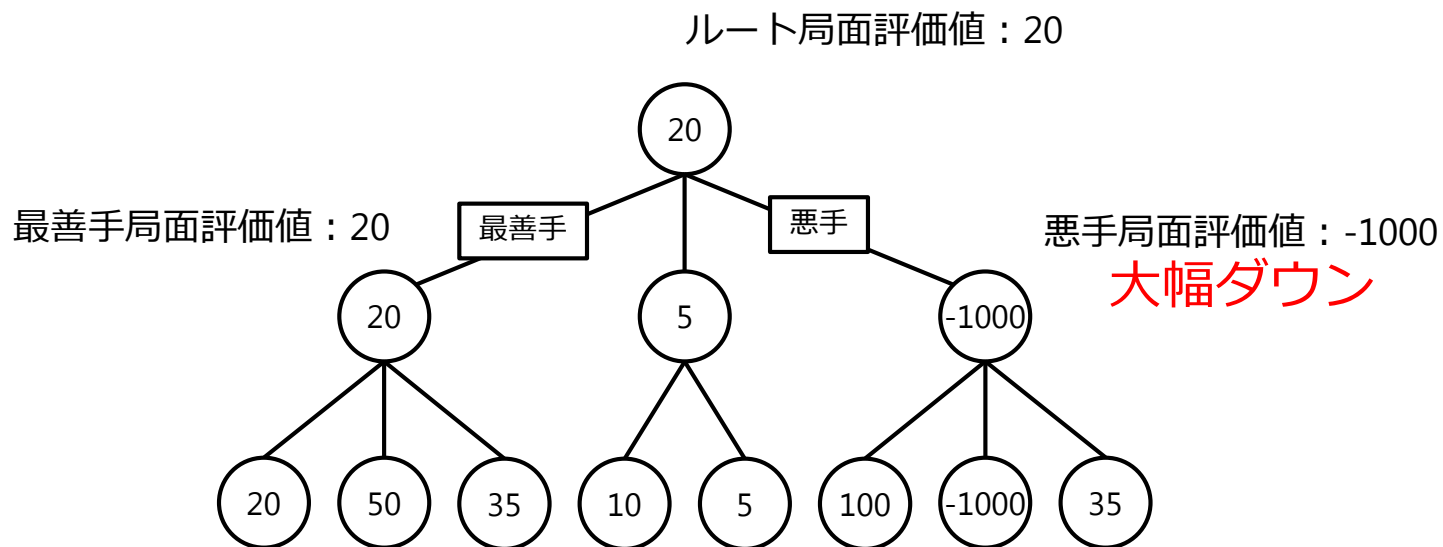
解説木になりうる木
→ 候補木

候補木の列挙 → 解説木の選択

解説木の選択

- 将棋プログラムの評価値を元に選択
 - 解説文中に現れる指し手は基本的には悪手ではない指し手
 - 解説のために悪手を提示することもある
 - 手順中に悪手を1度は認める
 - ただし悪手が連続することは稀

悪手の判別



指し手前後の**評価値変動幅** = 指し手の悪手度
を基準に解説木選択
木全体の悪手度が低い候補木

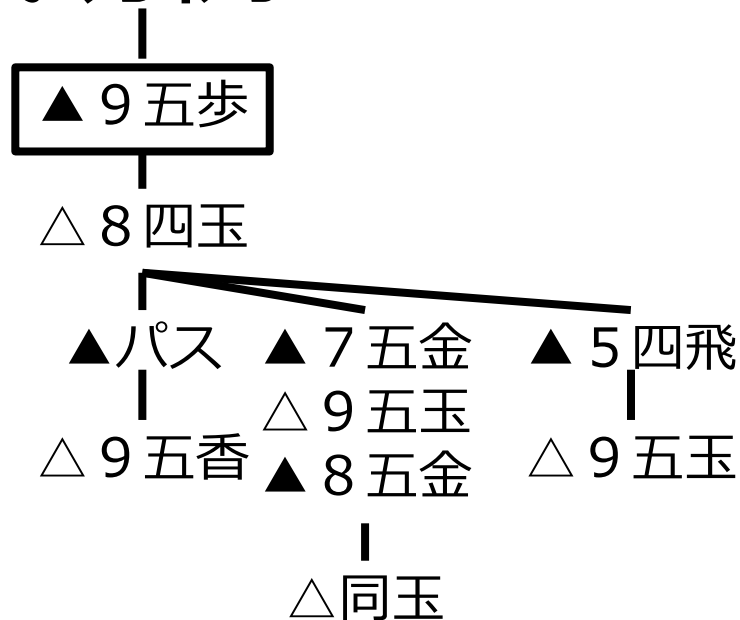
文生成への適用



解説文
 △同飛▲8三角に△6四飛として、飛車を活用しやすいように逃げる手も有力だ。～～

対象の指し手から句点まで、あるいは次の指し手までを対象とする
 ある程度長い文（単語数15以上）に限定

解説木生成の成功例



解説文

(略) △8四玉が次の一手のような絶妙手。先手玉には△9五香の詰めろがかかり、後手玉は▲7五金△9五玉▲8五金と追われても、△同玉が逆王手となる。また▲5四飛の王手も△9五玉と逃げて後手の勝ち。(略)

解説木の生成

- 43,408 / 52,703 局面 (82.4%) について候補木の生成に成功
- 候補の生成に成功した150局面中119局面 (79.3%) は正しい木の選択に成功

生成に失敗した局面の解析結果

失敗した局面のうち150局面について調査

誤りの種類	局面数
候補数による打ち切り	68
符号などの記載誤り	23
他の対局	1
指し手符号によらない指し手表現	16
遠い過去の指し手	15
手持ちにない駒	13
駒そのものを示す符号	6
不可能な指し手	4
王手放置	4

手法の改善が必要

生成に失敗した局面の解析結果

- 指し手符号によらない指し手表現
 - 例：「銀を上げて」 「駒を取り合って」
 - ルールの設計または意味の理解が必要
 - ルールの設計は高コストかつ網羅することは困難
 - 意味の理解は現状の技術では難しい

解説木の選択

- 将棋プログラム「激指」を用いて選択
 - 探索深さ8（1手進めた局面で7）

同一の130局面についての生成精度と計算時間

探索深さ	4	8	12
正解率	70%	74%	73%
生成時間	1.2分	3.9分	45分

- プログラムと解説者の意見の不一致
- 悪手を含まない木は除外できない

生成結果の例

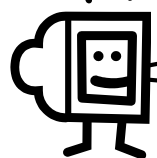


入力指し手

指し手：△8五歩

単語	値	単語	値
—	0.909	—	0.909
角	0.770	換わ	0.754
突	0.677		

単語予測の結果



△8五歩と突 **生成文** と一手損角換わりになる

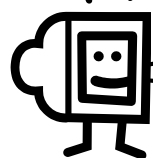
一手損ではなくノーマル角換わりになる形だが角換わりではある歩を「突く」という動作が獲得できている

生成結果の例

	9	8	7	6	5	4	3	2	1	
	香	桂	銀	金	王		銀	桂	香	一
		飛					金	角		二
	歩		歩	歩	歩	歩	歩	歩	歩	三
		歩								四
										五
つ			歩					歩		六
ね	歩	歩		歩	歩	歩	歩		歩	七
口		角	金					飛		八
	香	桂	銀		玉	金	銀	桂	香	九

指し手：△ 3 四歩

単語	値	単語	値
な	0.995	取り	0.966
角	0.930	横歩	0.854



△ 3 四歩は横歩取りになり、横歩取りになっている

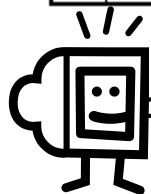
局所的には「横歩取りになる」ことが獲得できている
2単語前までしか見ておらず同じフレーズを繰り返す
言語モデルの改善が必要

生成結果の例

	9	8	7	6	5	4	3	2	1	
	王	桂	銀	金	飛			桂	香	一
	香				▲					二 ▲
	歩	歩	歩	歩		歩	馬	歩	歩	三 なし
					銀	歩				四
			歩	歩			歩			五
つ			歩		銀	歩				六
の	歩	歩	銀	金	歩	歩			歩	七
口		角	玉					飛		八
	香	桂		金				桂	香	九

指し手 : ▲ 6 六銀

単語	値	単語	値
い	0.999	銀	0.890
上が	0.158		



▲ 6 六銀と銀を上げていこうということだろう

「銀を上げる」という動作は取れている
 しかしその狙い（中央に利きを集める）の解説には至っていない
 探索の結果（5五の地点の争い）などの情報が必要

発表の流れ

- 背景・目的
- 戦型に限定した解説文生成
- 指し手表現を考慮した解説文生成
- **まとめ**

まとめ

- ルールベースの手法と探索結果を用いて解説木の獲得にある程度成功した
 - より高度なルールの設計による精度向上が期待できる
- 戦型への言及，駒の動きの獲得には成功したが狙いの解説には至らなかった
 - **探索木の情報**などより多くの情報を用いた生成が必要

発表文献

- 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年11月.
- 亀甲博貴, 森信介, 鶴岡慶雅. **将棋解説文のグラウンディングのための指し手表現と局面状態の対応付け**. 第19回ゲームプログラミングワークショップ, pp.202-209, 2014年11月.
*優秀論文賞受賞
- 亀甲博貴, 浦晃, 三輪誠, 鶴岡慶雅, 森信介, 近山隆. **将棋解説の自動生成のための局面からの特徴語生成**. 第18回ゲームプログラミングワークショップ, pp.36-43, 2013年11月.
*研究奨励賞受賞
- 亀甲博貴, 三輪誠, 鶴岡慶雅, 森信介, 近山隆. **ロジスティック回帰による言語モデルを用いた将棋解説文の自動生成**. 言語処理学会第20回年次大会, pp.943-946, 2014年3月.