

人狼ゲームにおける 能力者名乗り出プレイヤー・能力行使結果の同定*

林 友超[†] 小林 拓貴[†] 宇津呂 武仁[†]

[†]筑波大学大学院 システム情報工学研究科

1 はじめに

本論文では、人狼ゲームにおける自然言語理解モデルについての研究を行う。人狼ゲームにおける自然言語理解タスクの中でも特に重要な部分は、(1) 能力者(占い師・霊能者)と名乗り出たプレイヤー、(2) (1)のプレイヤーによる「占い発言」および「霊能発言」、(3) それらの発言において、(3-1) 占い・霊能対象となっているプレイヤー、(3-2) 占い・霊能結果(人狼か人間)、を同定するタスクである。

例えば、文献 [1] においては、人狼ゲームに参加する初心者プレイヤーをサポートする支援システム、および、人狼ゲームをプレーする AI プログラムを開発するための要素技術研究として、テキスト版人狼ゲームのプレーログを情報源として、人狼役職プレイヤーを推測するためのセオリーをマイニングする手法を提案している。具体的には、文献 [1] では、本物の「占い師」と、「占い師」と名乗り出た人狼または狂人の発言中の「占い発言」を用いてセオリーをマイニングする。しかし、このマイニング手法を大規模に適用するためには、(1) 「占い師」、「霊能者」と名乗り出た本物および偽物(人狼または狂人)のプレイヤー、(2) (1)のプレイヤーによる「占い発言」、「霊能発言」、および、(3) それらの発言において、(3-1) 占い・霊能対象となっているプレイヤー、および、(3-2) 占い・霊能結果が特定された大規模事例が必要不可欠である。

そこで、本論文では、任意の役職のプレイヤーを模倣する AI における自然言語理解タスクの要素技術として、参照可能な情報が最も少ない村人プレイヤー視点のもとで上述の(1)、(2)、(3-1)、(3-2)を同定するタスクを実現する。

2 人狼サーバ・ログデータ

本論文では、掲示板型インターネット人狼ゲームサイトである人狼 BBS から収集可能なゲームログのテキストデータを研究資源とする。人狼 BBS におけるゲーム参加者数は、一回のゲームあたり 9 人から 15 人である。参加者数 15 人の場合の役職内訳は、占い師、霊能者、狩人、狂人各 1 人、人狼 3 人、村人 8 人である。

3 能力者名乗り出プレイヤー及び能力行使結果の同定タスク

本論文のタスクは、以下の 4 つから構成される。

タスク 1: 能力者名乗り出プレイヤーの同定タスク

タスク 1 では、全プレイヤーの一日目から三日目までの発言集合¹に基づいて、各プレイヤーに対して、「占い師であると名乗り出」、「霊能者であると名乗り出」、「占い師・霊能者と名乗り出ていない」のいずれであるかを判定する。タスク 1 の参照用事例 4 例を表 1 に示す。

タスク 2: 能力者名乗り出プレイヤーによる占い・霊能発言の同定タスク

タスク 2 においては、タスク 1 で「占い師であると名乗り出」または「霊能者であると名乗り出」と判定されたプレイヤーの 2 日目からの発言を対象として、各発言に対して、占い・霊能発言であるか否かを判定する²。タスク 2 の参照用事例を表 2 に示す。

タスク 3-1: 能力者名乗り出プレイヤーによる占い・霊能発言の対象プレイヤー名の同定タスク

タスク 3-1 においては、タスク 2 で「占い・霊能発言である」と同定された発言に対して、占い・霊能発言

*Identifying Players Revealing Oneselves as Seer/Medium Players and Seer/Medium Results in Werewolf Game

[†]Youchao Lin, Hiroki Kobayashi, Takehito Utsuro, Graduate School of Systems and Information Engineering, University of Tsukuba

¹人狼 BBS では、占い師名乗り出発言は二日目までに、霊能者名乗り出発言は三日目までに出される場合が多いため、本タスクでは三日目までの発言を対象とする。

²占い師、霊能者共に、特殊能力を行使できるのは 2 日目からであるため、タスク 2 においても、2 日目からの発言を対象とする。

の対象となったプレイヤーを同定する。タスク 3-1 の参照用事例を表 2 に示す。

タスク 3-2: 能力者名乗り出プレイヤーによる占い・霊能結果の同定タスク

タスク 3-2 においては、タスク 2 で「占い・霊能発言である」と同定された発言に対して、占い・霊能結果の内容が、「人間」・「人狼」のいずれであるかを同定する。

3.1 分類器の設計

本論文では、人手で作成した規則を入力とする分類器 (SVM あるいは CNN), 分散表現を入力とする分類器 (CNN), 分散表現と規則を併用してそれを入力とする分類器 (CNN), の 3 種類を設計した。

分類器が SVM³ の場合、人手で作成した規則の照合状況を二値パターン化し、SVM の素性として用いる。

分類器が CNN の場合、人手で作成した規則の照合状況の二値パターンを入力する場合、分散表現を入力する場合、および、分散表現・規則を併用する場合、の三通りを評価する⁴。規則の照合状況の二値パターンを入力する場合は、SVM の場合と同様である。分散表現を入力とする場合は、文字単位の分散表現⁵を用い、モデルの訓練時に分散表現の更新を行わない設定とした⁶。規則の照合状況および分散表現を併用する場合は、文の文字列の末尾に規則の照合状況の二値化文字列を結合した後、分散表現を入力とする場合と同様の手順とした。

4 各タスクの同定手法と評価結果

4.1 手法

タスク 1 では、1~3 日目までに発言した全プレイヤーを対象として「占い師と名乗り出た」、「霊能者と名乗り出た」、「占い師・霊能者のどちらにも名乗り出していない」のいずれであるかを同定する。使用する規則は 73 個 (図 1) あり、「占い師・霊能者である」と名乗り出る発言、および、「占い師・霊能者ではない」と名乗

³<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

⁴実装は python のライブラリである Pytorch で行ない、活性化関数は ReLu とした。人手で作成した規則の照合状況の二値パターンを入力する場合は、畳み込み層 1 層、マックスプーリング層 1 層、全結合層 2 層とし、チャンネル数 40, epoch 数 50, 学習率 0.001 とした。その他の場合は、文献 [4] に従い、畳み込み層 1 層、マックスプーリング層 1 層、全結合層 1 層とし、チャンネル数 1, epoch 数 100, 学習率は 0.001~0.0001 (徐々に減少) とした。

⁵日本語 Wikipedia 本文テキストを用いて訓練した FastText [2](<https://fasttext.cc/docs/en/pretrained-vectors.html>) 版の分散表現。

⁶以下の設定のうち最も高性能の組み合わせを用いた結果である。(1) 入力文の分割単位: 文字または単語, (2) 分散表現の訓練事例: 日本語 Wikipedia 本文テキストのみ, または, 日本語 Wikipedia 本文テキストと人狼 BBS646 ゲーム分のテキストの混合集合, (3) モデル訓練時分散表現の更新の有無: 「あり」または「なし」。

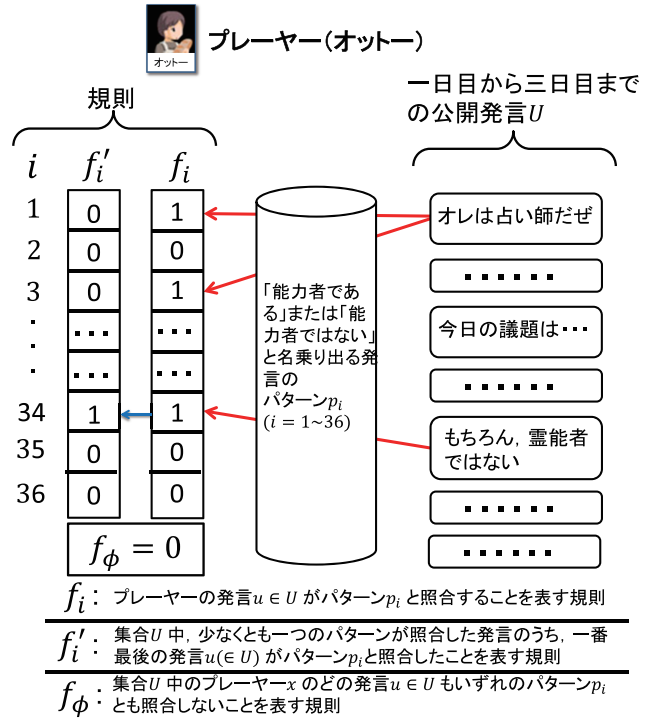


図 1: タスク 1 におけるプレイヤーの名乗り出状況素性化のための規則群

り出る発言の照合規則 36 個、それらの 36 種類の規則の複数に照合した場合、「どの規則が一番最後に照合した規則か」を表す規則 36 個、および、「どの規則にも照合しなかった」を表す 1 個である。

タスク 2 では、タスク 1 で同定した「占い師・霊能者と名乗り出た」プレイヤーの各発言に対して、占い・霊能発言であるか否かを判定する。使用する規則は、発言に“【】”含むか否か、“【】”で囲まれた発言が「確認」あるいは「把握」を含むか否か、等の規則 6 個である。

タスク 3-1 では、タスク 2 で同定した「占い・霊能結果」を含む発言に対して、占い・霊能結果の対象となるプレイヤー (19 種類のアバター名のうちの一つ) を同定するタスクである。使用する規則は、19 種類のアバター名ごとに、そのプレイヤーが生存しているか、発言にアバター名の呼び方を含んでいるか、等全 4 種類に加えて「いずれの規則にも照合しない」の 1 個の合計 77(=4 × 19 + 1) 個である。

タスク 3-2 では、タスク 2 で同定した「占い・霊能結果」を含む発言に対して、占い・霊能結果の内容が、「人間」・「人狼」のいずれであるかを同定する。使用する規則は、人狼 BBS において「人間」を表す 3 種類の単語の照合規則、「人狼」を表す 3 種類の単語の照合規則、「いずれの規則にも照合しない」を表す規則 1 個の合計 7 個である。

表 1: タスク 1 の参照用事例の例

発言者・役職・発言日	リーザ・占い師・1日目	カタリナ・狂人・1日目	パメラ・人狼・1日目	オットー・村人・1日目
1~3日为中心的な役割を果たした発言	【リーザは占いができるの!】おにいちゃんの役に立ちたくて...	... あたしは【占い師でも霊能者でもない】じよー...	... 【霊能者CO、霊能者】...	... 【僕は占いも出来ないし、幽霊も見えないよ】...
能力者名乗り出状況の値	占い師であると名乗り出	占い師・霊能者であると名乗り出していない	霊能者であると名乗り出	占い師・霊能者であると名乗り出していない

表 2: タスク 2, タスク 3-1, タスク 3-2 の参照用事例の例

発言者・役職・発言日	リーザ・占い師・3日目	カタリナ・狂人・4日目	パメラ・人狼・4日目
発言	... 【パメラさんは人間】...	... 羊は、消去法で狼かなあ...	... 羊狼だと違和感...
タスク 2: 「能力者名乗り出プレイヤーによる占い・霊能発言であるか否か」の値	占い・霊能発言である	タスク 1 の値が「占い師・霊能者であると名乗り出していない」であるため、タスク 2, タスク 3-1, タスク 3-2 の対象にならない	占い・霊能発言ではない
タスク 3-1: 「能力者名乗り出プレイヤーによる占い発言の対象プレイヤー名」の値	パメラ		タスク 2 の値が「占い・霊能発言ではない」であるため、タスク 3-1 及びタスク 3-2 の対象にならない
タスク 3-2: 「能力者名乗り出プレイヤーによる占い・霊能結果」の値	人間		

4.2 評価

各モデルの評価方法として、再現率・適合率曲線を用いた。各タスクの評価結果をそれぞれ図 2, 図 3, 図 4, 図 5 に示す。

タスク 1 の評価結果 (図 2) においては、CNN(規則), SVM(規則) が最も性能が高く、CNN(分散表現) が最も低い性能となった。CNN(分散表現), CNN(分散表現+規則) の性能が低い理由としては、一日目から三日目までの最大 60 発言を一つの文として入力するため、入力文が長くなり、分散表現を用いたモデルの性能が低くなったためであると考えられる。

タスク 2 の評価結果 (図 3) においては、CNN(分散表現+規則) が全体的に最も性能が高く、CNN(分散表現) は、低再現率では SVM(規則), CNN(規則) よりも高い適合率となった。タスク 2 においては、分散表現と規則の併用により、「分散表現のみ」・「規則のみ」の性能を改善することができた。

タスク 3-1 の評価結果 (図 4) においては、CNN(規則), SVM(規則) が最も性能が高く、CNN(分散表現), CNN(分散表現+規則) は共に低い性能となった。タスク 3-1 は 19 値分類であり各クラス毎の訓練データ数が少ないため、分散表現を用いる場合に訓練事例数が不足して性能が低くなったと考えられる。

タスク 3-2 の評価結果 (図 5) においては、再現率最大では、SVM(規則), CNN(規則), CNN(分散表現+規則) がほぼ同等に高適合率となったが、再現率最大以外では、CNN(規則), CNN(分散表現+規則) の適合率が高くなった。

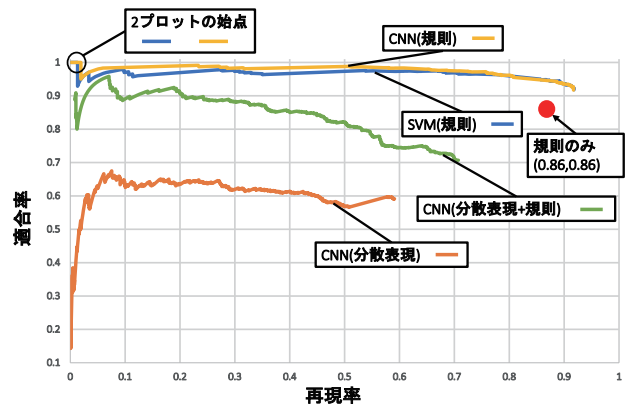


図 2: タスク 1 の評価結果 (全体)

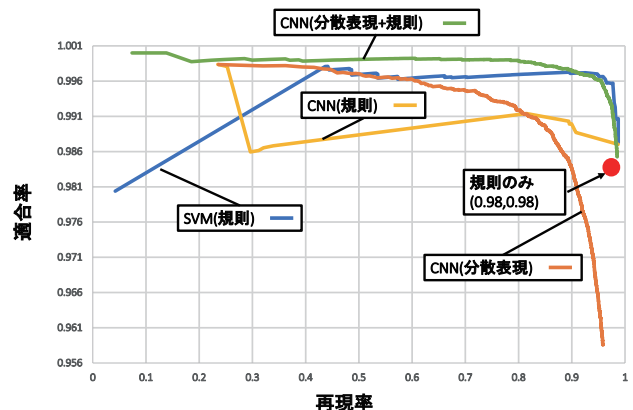


図 3: タスク 2 の評価結果 (全体)

5 全タスク群の直列適用評価

4 節では、各タスクの入力には誤りが含まれないことを仮定し、各タスクが相互に独立に適用できる場合に限定して、評価を行った。これに対して、本論文では、4 節の評価において最も性能の高かった CNN を

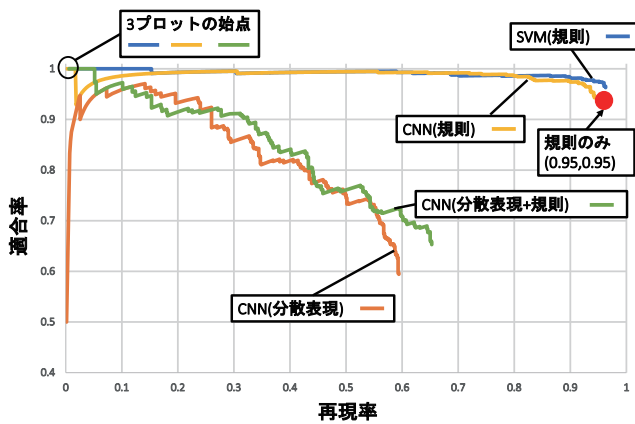


図 4: タスク 3-1 の評価結果 (全体)

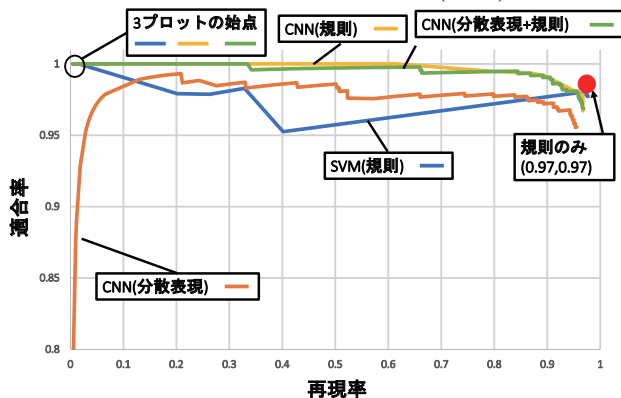


図 5: タスク 3-2 の評価結果 (全体)

用いて、以下の手順により、タスク 1, タスク 2, タスク 3-1, タスク 3-2 の直列適用 (本論文では「タスク 1+2+3」と定義する) を行い、その性能を評価した。

- (1) タスク 1 において、占い師・霊能者と名乗り出たと予測されたプレイヤーの全発言をタスク 2 の入力とする。
- (2) タスク 2 において、能力行使結果発言であると予測された全発言をタスク 3-1, タスク 3-2 の入力とする。

タスク 1(のうち、占い CO・霊能 CO プレイヤー同定結果), タスク 2(のうち、占い・霊能発言同定結果), タスク 3-1(全体), タスク 3-2(全体), および、タスク 1+2+3 の評価結果を図 6 に示す。評価結果においては、CNN の softmax 関数の出力値に下限値を設け、この下限値を変化させて再現率・適合率曲線をプロットした。この評価結果においては、各タスク単独の性能と比較して、全タスクを直列適用することにより、各タスク単独での再現率・適合率が積算することにより、タスク 1+2+3 の性能が低下していることが分かる。

6 関連研究

人狼ゲーム AI の開発において、文献 [6] では、第一人狼知能大会の対戦ログを用いて、大会に出場した

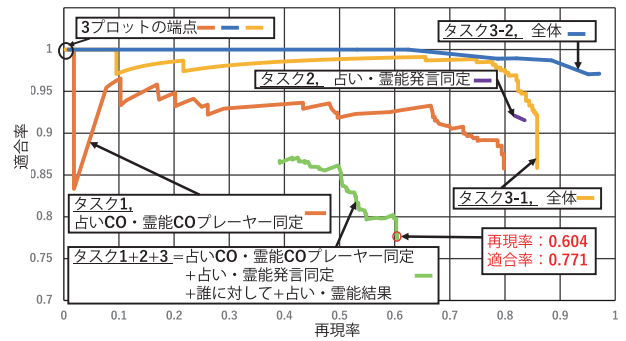


図 6: 各タスク (1・2・3-1・3-2) ごと及びタスク 1+2+3 の評価結果

エージェント (AI) の行動が勝敗への影響を与えるかの分析を行った。また、同じ対戦ログに対して、文献 [3] では、CO 状況, 占い結果, 投票変更数などを素性として、SVM 分類器で人狼エージェントの推定精度について評価を行った。それを踏まえて SVM 分類器を実装したエージェントも設計した。これらの研究はエージェント同士のプレーログのみを用いているため、本研究で扱っている人間同士のプレーログという点で異なる。

7 おわりに

本論文では、全プレイヤーの発言内容を自然言語解析した結果を素性とする分類器学習によって、「占い師」・「霊能者」と名乗り出たプレイヤーの同定、および、そのプレイヤーの能力行使結果の同定を行う手法を提案した。今後は、本研究で同定した結果に対して、制約充足により役職絞り込み [5] を行うこと、および、セオリーマイニング [1] を行うことが挙げられる。

謝辞

本研究を行うにあたり、人狼 BBS のデータ使用を許可していただいた ninjin 氏に感謝する。

参考文献

- [1] 板東勇樹, 呉双, 林友超, 宇津呂武仁. 人狼ゲームログからの狼役職絞り込みセオリーのマイニング. 第 31 回人工知能学会全国大会論文集, 2017.
- [2] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. *Transactions of ACL*, Vol. 5, pp. 135–146, 2017.
- [3] 梶原健吾, 鳥海不二夫, 稲葉通将, 大澤博隆, 片上大輔, 篠田孝祐, 松原仁, 狩野芳信. 人狼知能大会における統計分析と SVM を用いた人狼推定を行うエージェントの設計. 第 30 回人工知能学会全国大会論文集, 2016.
- [4] Y. Kim. Convolutional neural networks for sentence classification. In *Proc. EMNLP*, pp. 1746–1751, 2014.
- [5] 林友超, 呉双, 板東勇樹, 宇津呂武仁. 制約充足手法を用いた人狼ゲームの役職絞り込み方式. 第 31 回人工知能学会全国大会論文集, 2017.
- [6] 鳥海不二夫, 稲葉通将, 大澤博隆, 片上大輔, 篠田孝祐. 人狼知能大会におけるエージェントの行動分析. 情報処理学会研究報告, Vol. 2016–EC–41, pp. 1–8, 2016.