

人狼ゲームにおける発話行為タグの自動付与

An automatic annotation method of speech act tag in Werewolf

平田 佑也
Yuya Hirata

稲葉 通将
Michimasa Inaba

高橋 健一
Kenichi Takahashi

広島市立大学大学院情報科学研究科知能工学専攻

Email: hirata.y@cm.info.hiroshima-cu.ac.jp

Email: inaba@hiroshima-cu.ac.jp

Email: takahasi@hiroshima-cu.ac.jp

Abstract— Werewolf is a popular game all over the world, but it's fundamental property is not clear. A previous study[2][3] analyzed log data from the WerewolfBBS[1] using speech act tags annotated manually. Giving speech act tag by hand takes an enormous amount of time and causes human errors. This paper provides an automatic annotation method of the tags.

I. はじめに

人狼ゲームは、世界各地で行われている人気のゲームであるが、ゲームの基本的性質はほとんど明らかになっていない。

先行研究では、Web 上で人狼ゲームが遊べる人狼 BBS[1]の過去ログを用いて、1 人 1 人の発言に発話内容の特徴を示した発話行為タグを手動で付与し、分析していた。しかし、手動での付与は膨大な時間を要し、人為的ミスを含む可能性がある。そこで本研究では発話行為タグの自動付与を目的とする。

II. 人狼

人狼は、人との対話でゲームが進行する。プレイヤーが人間側と人狼側に分かれ、自分の正体を隠しつつ他のプレイヤーと議論して相手の正体を探りながら自陣営側の勝利を目指すゲームである。

ゲームは夜と昼の半日単位で進行する。夜には、人狼が襲撃する対象を人間側から 1 人決定する。昼には、プレイヤー全員が議論し、処刑対象(人狼と思しきプレイヤー)を決定する。襲撃対象・処刑対象となったプレイヤーは死亡し、その後ゲームから除外される。全ての人狼を処刑することが出来れば人間側の勝利となり、人狼の数が人間の数以上になったら人狼側(狂人含む)の勝利となる。

プレイヤーはゲーム開始時に人間側か人狼側のどちらかにランダムに振り分けられる。その際、人間側のプレイヤーはどのプレイヤーがどちらの陣営に属しているかを知ることができな

い。一方、人狼側はどのプレイヤーがどの陣営に属しているかを知ることができる。人間側には、能力が何もない村人、人間でありながら人狼の味方をする狂人、特殊能力を備えた能力者の 3 つに分けられる。人狼側はどのプレイヤーがどの陣営に属しているかを知ることができるが、この特殊能力を誰が有しているのか、また誰が狂人であるのかは知ることができない。能力者には、夜の間プレイヤー 1 人の正体を占い、人間・人狼の区別を知ることができる占い師、夜の間その日の処刑者が人間なのか人狼なのかを知ることができる霊能者、夜の間プレイヤーの中から 1 人を指定し、そのプレイヤーを人狼のから守ることができる狩人、二人一組の役職であり、お互いがお互いを同じ役職だと認識できる、つまり自分以外に 1 人、人狼でないプレイヤーを知ることができる共有者がいる。

III. 人狼 BBS

人狼 BBS は Web 上で遊べる、誰もが参加可能な Web ゲームである。ここでは BBS における対話形式によってゲームが進行する(図 1)。人狼 BBS では、一般的な人狼ゲームとは一部ルールが異なるので以下に記した。

A. 対話方式

人狼 BBS には、人狼同士だけで対話を行える場がある。しかし、人間は人間側も人狼側も見聞きすることができる公の場でしか対話を行うことができない。また、死者(襲撃や処刑の対象となりゲームから除外されたプレイヤー)は墓場という死者同士で対話を行うことができる場がある。生存者には墓場での対話は見えない。

図 1 に表示されている赤い発言は、人狼同士の対話であり、白い発言は、公の場での対話である。また、灰色の発言は、墓場での死者同士の対話である。図 1 では、全ての発言が表示されているが、人狼ゲームをプレイ中は、人間側

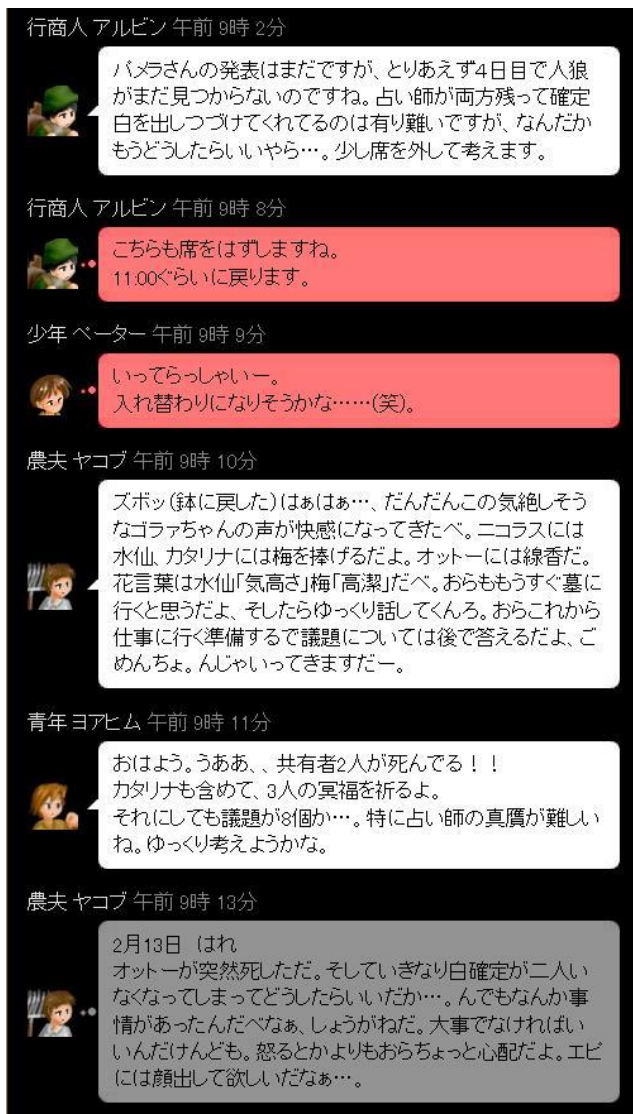


図 1 人狼 BBS

には白い発言しか見ることができない。また、人狼側には白い発言と赤い発言を見ることができ、死者は、白い発言と灰色の発言を見ることができ。一般的な人狼ゲームでは、司会役が存在する。全員が見聞きできる公の会話以外は、司会の指示によって全員が目を閉じ、該当プレイヤーのみが目を開けジェスチャー等で会話をする。図 1 のようなグループ対話の区別をしている。また、墓場は存在しないことが多い。

B. 役職構成

役割ごとの構成人数はゲーム開始時のプレイヤー数によってあらかじめ決定されている。本研究では同一のルール・人数・役職構成で行われたゲームデータを分析対象とする。なお、本研究では、プレイヤー数 16 名、役職構成は村人 7 名(1 人 NPC を含む)、人狼 3 名、占い師 1 名、狩人 1 名、霊能者 1 名、狂人 1 名、共有者 2 名である。

C. ゲームの進み方

ゲーム内の 1 日はリアルタイムの 1 日と同じ時間で進行し、約 1 週間かけて勝敗を決める。処刑と襲撃は 2 日目の夜から始まる。1 日目は、プレイヤー同士の交流を深めるために雑談を行うことが多い。また、2 日目以降のゲームの方針について話し合いも行われる。一般的な人狼ゲームでは、ゲーム内の 1 日を 5 分や 10 分といったもっと短時間に設定して行う。ゲームの流れは人狼 BBS と同様である。

IV. 発話行為タグ

本研究では人狼 BBS の過去ログを使用し、プレイヤー 1 人 1 人の発言に発話行為タグを付与する。発話行為タグは発話内容の特徴を示すものである。これを付与することにより、誰がどのような発言をしたのか統計を取ることができる。先行研究[2]より発話行為タグが 23 種作成された。しかし、先行研究[3]より勝敗に関わる重要な発話行為タグが示された。したがって本実験では重要な発話行為タグである CO(自分の役職を宣言する発言)、execution(吊りに関する発言)、fortune-telling(処刑に関する発言)の 3 つに限定する。表 1 に各発話行為タグが付与される発話例を示す。

表 1 各発話行為タグが付与される発話例

発話行為タグ	発話例
CO	(1)「CO」という単語を含む発言を選択 (2)「一人称」-「6文字以内」-「役職名」の並びを含む発言を選択
fortune-telling	(1)「占い師(占師)」-「5文字以内」-「人名」の並びを含む発言を選択 (2)「占い師(占師)」という単語に含まれている場合を除く「占」という単語を含む発言を選択
execution	(1)「吊」「処刑」「投票」という単語のいずれかを含む発言を選択

V. Support Vector Machine を用いた発話行為タグの自動付与

本実験は提案手法(TF-IDF 法を用いたキーワードによる発話行為タグの自動付与)と比較するための予備実験として行った。

すでに手動で発話行為タグが付与されている、人間側勝利 12 ゲーム、人狼側勝利 12 ゲームの計 24 ゲーム(19621 発話)のデータに対し、mecab を用いて形態素解析を行った。得られた形態素を用いて、bag-of-words により一発話ごとに単語ベクトルで表した。得られたベクトルのうち、1 ゲーム分をテストデータ、残りの 23 ゲーム分を訓練データとして、任意の発話行為タグが付与されるか否か、SVM[4]を用いた判定を行った。これを 24 回行い、平均を求めることで結果とした。

VI. TF-IDF 法

TF-IDF とは、文書中に出現した特定の単語がどの程度特徴的であるかを識別するための指標のことである。主にテキストマイニングや情報探索などの分野で利用される。TF-IDF の TF(Term Frequency)は、その文書中で特定の単語が出現した回数を表し、IDF(Inverse Document Frequency)は、コーパス全体の中でその文書を含む文書数の自然対数を表す。文書群から抽出した単語の中で、語 w の出現頻度数を $TF(w)$ 、語 w が 1 回以上出現した文書数を $IDF(w)$ 、文書群中の全文書数を N としたとき、以下の式(1)で求められるものを語 w の TF-IDF 値(TFIDF(w))とする。

$$TFIDF(w) = TF(w) \left(1 + \log \frac{N}{DF(w)} \right) \quad \text{式(1)}$$

TF-IDF による重み付けを利用したアルゴリズムを広く TF-IDF 法と呼ぶ。主にキーワード抽出や全文検索エンジンの重み付けなどに応用される。また、ベクトル空間モデルにおいて、文書間の類似度を判定する場合に、コサイン類似度を計算する際の単語の特徴ベクトルとして TF-IDF の値が利用される。

VII. キーワード

すでに手動で発話行為タグが付与されている人間側勝利 12 ゲーム、人狼側勝利 12 ゲームの計 24 ゲーム分(151 日, 19621 発話)について、TF-IDF 法によって単語の重要性を調べるに当たり、任意の発話行為タグが付与されている発話群中、特定の形態素を含む発話の出現頻度数を TF とし、24 ゲーム分の発話群中、特定の形態素を含む発話の総数を DF とする。また、24 ゲーム分の発話の総数を N とし、式(1)を用いることで TF-IDF 値を求めた。本研究では、TF-IDF 値の高い単語のうち、各発話行為タグと関係性が高いと思われるものをキーワードと呼ぶ。TF-IDF 法を用いて抽出されたキーワードを表 2 に示す。

表 2 抽出されたキーワード

発話行為タグ	抽出されたキーワード
CO	CO
fortune-telling	占
execution	吊・投票・処刑

I. TF-IDF 法を用いたキーワードによる発話行為タグの自動付与

すでに手動で発話行為タグが付与されている、人間側勝利 12 ゲーム、人狼側勝利 12 ゲームの計 24 ゲーム(19621 発話)のデータに対し、得ら

れたキーワードを用いて発話行為タグの付与条件となるルールを作成した。

A. CO についてのルール

- 「CO」という単語が含まれる発話を選択
- 「一人称」-「6 文字以内」-「役職名」の並びが含まれる発話を選択

キーワード「CO」を含む発話を調べた結果、その発話全てに発話行為タグ CO を付与されていた。更に、キーワード「CO」を含まず、CO の発話行為タグが付与されている発言を確認したところ、ほぼ全て下記の例のようになっていた。

例1)私【は】占い師だ

例2)オラ【,】占い師だ

例3)私【の役職は】占い師です

例4)おれ【さまが】占い師だ

例5)俺【さまの役職は】占い師だ

例 1~3 のような「一人称」と「役職名」の間に副詞や句読点が入っている場合、例 4 のような「一人称」+「敬称」を使用する場合、また例 5 のような前述 2 つの複合型の場合の 3 つに分けることができる。これらはいずれも「一人称」-「6 文字以内」-「役職名」の並びになっていた。

B. fortune-telling についてのルール

- 「占い師(占師)」-「6 文字以内」-「人名」の並びが含まれる発話を選択
- 「占い師」「占師」という単語が含まれている場合を除く、「占」という単語が含まれる発話を選択

キーワード「占」を含み、発話行為タグ fortune-telling が付与されている発話を調べた結果、ほぼ全てが下記の例のようになっていた。

例1)「占い師さんはペータをお願いします」
「占い師」-「副詞等」-「人名」という形で占い先を指定するもの

例2)「占い先はペータで」「占:ペータ」
「占」-「"キャラクター名"」という形で占い先を指定するもの

例3)「占い先はどうやって決めようか?」
単純な占いに関する発言であるもの

「占」という単語が含まれず、「占い師」「占師」という単語が 1.のルール以外で出てくるものは、CO や狼の襲撃、狩人（誰を守るか）に関連する発話であった。

C. execution についてのルール

1. 「吊」という単語が含まれる発話を選択
2. 「処刑」という単語が含まれる発話を選択
3. 「投票」という単語が含まれる発話を選択

キーワード「吊」「処刑」「投票」を含まず、発話行為タグ execution が付与されているものは見つからなかった。

II. 自動付与手法の違いによる比較・考察

表 4 に 3 つの発話行為タグについて Support Vector Machine を用いた発話行為タグの自動付与(以下、SVM 手法とする)の結果と TF-IDF 法を用いたキーワードによる発話行為タグの自動付与(以下、TF-IDF 手法とする)を示す。

表 4 2 手法での結果

発話行為タグ	手法	精度(%)	再現率(%)	F値
CO	SVM	55.29	37.16	0.44
	TF-IDF	79.05	97.57	0.87
fortune-telling	SVM	56.28	41.67	0.46
	TF-IDF	72.63	96.96	0.83
execution	SVM	53.54	47.12	0.49
	TF-IDF	76.37	100	0.87

SVM 手法の結果よりも TF-IDF 手法の結果の方が良い結果となった。この原因として、SVM 手法には、訓練データの不足や訓練データに人為的ミスが混入している可能性があるため、性能が低いことが考えられる。更にこれらの発話行為タグは、一般的な発話では使用頻度が少ないが、これらに関する発話をする際には必要な単語(キーワード)が存在する。そのため、実際に人が手動で発話行為タグを付与する場合と同じような付与条件(キーワードを用いたルール)を TF-IDF 手法に設定でき、良い結果が得られたのではないかと考えられる。

III. おわりに

本研究では、人狼ゲームにおける発話行為タグの自動付与を行った。本実験を通して、3 つの発話行為タグに関して、SVM 手法より TF-IDF 手法を用いた方が良い結果が得られることが分かった。

今後の課題としては、判別できる発話行為タグを拡張する必要がある。

[1] 人狼 BBS <http://ninjinix.x0.com/wolf0/>.

[2] 大島 菜央実, 稲葉 通将, 高橋 健一: “発話内容に着目した人狼 BBS の分析”, 平成 25 年度卒業論文, 広島市立大学.

[3] 稲葉 通将, 大島 菜央実, 鳥海 不二夫, 高橋 健一: “雑談ばかりしてると殺される-人狼 BBS におけるプレイヤーの発言傾向と意思決定・勝敗の分析-”, JAWS 2013, 2013.

[4] B.E.Boserl.M.Guyon and V.N.Vapnik,"A training algorithm for optimal margin classifiers,"In proceedings of the fifth annual workshop on computational learning theory,ACM,pp.144-152,1992.

問い合わせ先

〒731-3194 広島市安佐南区大塚東 3-4-1

広島市立大学大学院 情報科学研究科 知能工学専攻

平田 佑也