An experimental analysis of whispers’ effect in Werewolf BBS by relational association rules

Saki Sakaguchi
Tomonobu Ozaki
-Nihon University-
Outline

1. Introduction
2. Werewolf game
3. Purpose
4. Data
   - werewolf BBS
   - protocol
   - predicate
   - dataset
5. Result
6. Conclusion and Future work
Introduction

Complete information game

Incomplete information game

Shogi, Go

Werewolf game
Werewolf game

○ villager side player
○ werewolf side player
Werewolf game

Villager side
Detect and execute hiding werewolves

villager side player
werewolf side player
**Werewolf game**

- **Werewolf side**
  - Deceive and attack villagers

- **Villager side**

- **Players:**
  - Red circles: werewolf side player
  - White circles: villager side player
Asymmetric information

Open conversation

Secret conversation (whispers)

- villager side player
- werewolf side player
Purpose

Try to capture a characteristic relationship between contents in the secret conversation and in open conversation

Method

Extract relational association rules.

\[
\text{fact_pred}() :\text{- open_pred1()}, \text{secret_pred2()}, \ldots
\]

Head predicate: fact of real action or event

Body predicate: contents in open conversation and at least in secret conversation
Werewolf BBS

- Online BBS website.
- Participate can play text-based werewolf game.
- Get many game logs.
- The same rules in original werewolf game with a few exception.

Ex)
- sudden death
- the number of mention restriction

Werewolf BBS：http://wolfg.x0.com/index.rb
## Four types of utterances

<table>
<thead>
<tr>
<th>Log</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>White log (open)</td>
<td>All player can read and mention</td>
</tr>
<tr>
<td>Red log (secret)</td>
<td>Werewolf player can read and mention</td>
</tr>
<tr>
<td>Gray log</td>
<td>Exiled player can read and mention</td>
</tr>
<tr>
<td>Blue log</td>
<td>Monologue</td>
</tr>
</tbody>
</table>

Use **white log** and **red log** for analysis.
## Protocol

<table>
<thead>
<tr>
<th>White log</th>
<th>Red log</th>
</tr>
</thead>
<tbody>
<tr>
<td>comingout</td>
<td>question</td>
</tr>
<tr>
<td>question</td>
<td>answer</td>
</tr>
<tr>
<td>answer</td>
<td>estimate</td>
</tr>
<tr>
<td>agree</td>
<td>agree</td>
</tr>
<tr>
<td>disagree</td>
<td>disagree</td>
</tr>
<tr>
<td>vote</td>
<td>adviced</td>
</tr>
<tr>
<td>request_divined</td>
<td>want_eat</td>
</tr>
<tr>
<td>inquested</td>
<td>want_vote</td>
</tr>
<tr>
<td>divined</td>
<td>deceived</td>
</tr>
<tr>
<td>guarded</td>
<td>disrelation</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14 actions</td>
</tr>
<tr>
<td></td>
<td>13 actions</td>
</tr>
</tbody>
</table>
Predicate

White log

\[ w\text{-}question(\ Game:\text{Day}, \ Player, \ Player2) . \]
\[ w\text{-}request\_divine(\ Game:\text{Day}, \ Player, \ Player2) . \]

Red log

\[ r\text{-}question(\ Game:\text{Day}, \ Player, \ Player2) . \]
\[ r\text{-}decieve(\ Game:\text{Day}, \ Player, \ Role) . \]

Ex) \[ r\text{-}question(A:1, \ wolf1, \ wolf2) . \]

\[
\text{Pred}(\text{Game}:\text{Day}, \ N, \ Player, \ Args\ldots) \ :- \ \text{Prev\_days}(N), \\
\text{Pday is Day}-N, \\
\text{Pred}(\text{Game}:\text{Day}, \ Player, \ Args\ldots).
\]
Dataset

• Select 6 game logs
detail: Werewolf side won 3 games of 6 games and lost the rest.
   Each game has twelve villager side players,
   three werewolf side players,
   no-sudden death and at least deceiving werewolf.

• Focus on 3 actions in the head predicate
  1, attacked (27 cases)
  2, executed (39 cases)
  3, mislead_by_werewolf (53 cases)
Extract relational association rule by using Aleph

1. Data
   - Extract all rules

2. All rules
   - Select rules having red predicate

3. Rules
   - Evaluate rules with three measures

Evaluation measures
Three rule evaluation measures

- **support** : Joint probability

- **confidence** : Conditional probability

- **diff** = $P(\text{Head} | \text{Body}) - P(\text{Head})$
  
  $P(\text{Head} | \text{Body})$ : the confidence value
  
  $P(\text{Head})$ : all possible instantiation of head predicate considering alive players and their roles

  diff$>0$ : the predicate has positive effect
## Result

<table>
<thead>
<tr>
<th>Head predicate</th>
<th>The number of extracted rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>attacked</td>
<td>4702</td>
</tr>
<tr>
<td>executed</td>
<td>3048</td>
</tr>
<tr>
<td>mislead_by_werewolf</td>
<td>3094</td>
</tr>
</tbody>
</table>
## Result

Rates of diff

\[ \text{diff} = P(\text{Head}|\text{Body}) - P(\text{Head}) \]

<table>
<thead>
<tr>
<th>Predicate</th>
<th>attacked</th>
<th>executed</th>
<th>mislead_by_werewolf</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_question</td>
<td>0.97</td>
<td>0.94</td>
<td>0.61</td>
</tr>
<tr>
<td>r_answer</td>
<td>0.82</td>
<td>0.95</td>
<td>0.43</td>
</tr>
<tr>
<td>r_advised</td>
<td>0.85</td>
<td>0.95</td>
<td>0.88</td>
</tr>
<tr>
<td>r_estimate</td>
<td>0.72</td>
<td>0.71</td>
<td>0.50</td>
</tr>
<tr>
<td>r_agree</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>r_want_eat</td>
<td>0.79</td>
<td>0.86</td>
<td>0.25</td>
</tr>
<tr>
<td>r_want_vote</td>
<td>0.81</td>
<td>0.62</td>
<td>0.77</td>
</tr>
<tr>
<td>r_disrelation</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>r_deceive</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
</tr>
</tbody>
</table>
Result

- Most of the red predicates have **positive effect**.

- **werewolf prompts other players**
  to vote a target player
  by saying he/she is a werewolf in the open conversation.
Result

- Most of the red predicates have **positive effect**.

- **werewolf prompts other players to vote a target player** by saying he/she is a werewolf in the open conversation.
Result

- Most of the red predicates have **positive effect**.

- **werewolf prompts other players**
  to vote a target player
  by saying he/she is a werewolf in the open conversation.
Result

1, attacked

attacked( Game:Day, C ) :- r_want_eat( Game:Day, 0, D, C ), w_answer( Game:Day, 2, C, E ), w_estimate( Game:Day, 1, C, E, not(werewolf) ).
Result
2, executed

executed( Game:Day, C ) :-
  w_vote( Game:Day, 0, D, C ),
  r_estimate( Game:Day, 2, F, D, hunter ).
Result

3, mislead_by_werewolf

mislead_by_werewolf( Game:Day, C ) :-
    w_agree( Game:Day, 2, E, C),
    r.want.eat( Game:Day, 2, F, E ).
Conclusion

- Extract relational association rules between secret conversation and real action.

Future work

- Precise evaluation of extracted rules
- Extract condensed representations of relational association rules, and evaluate them
- Investigate propensity score matching for relational data