

Retrieving Abstract Information from Multi-Party Conversations

Surangika Ranathunga

*Department of Computer Science and Engineering, University of Moratuwa
Katubedda, Sri Lanka*

surangika@cse.mrt.ac.lk

Abstract— This paper presents the use of dialog acts and addressee information in extracting domain-specific abstract information from multi-party conversations. Identification of dialog acts and addressee information is done in a domain-independent manner. This information is then used with a domain-specific rule set for identifying abstract information in chat conversations. To retrieve abstract information from a conversation in a new domain, the only requirement is to define a new rule set. This eliminates the need of domain-specific corpora. This paper also presents an improved version of a commonly used addressee recognition algorithm.

Keywords— natural language processing, dialog act recognition, addressee recognition.

I. INTRODUCTION

With the advancement of technology, human conversations take place in many ways other than face-to-face conversations. These include methods such as telephone conversations, video conferencing, typed conversations via online chat applications, and even communicating inside virtual worlds via avatar representations.

Retrieving abstract information, which we hereafter refer to as ‘communication actions’, from these conversations is useful in many scenarios such as in meeting summarizations, and implementing intelligent agents that converse with humans, e.g. as call centre agents. Moreover, abstracting chat conversations is important because of the causal relationship they have with other human actions. For example, in a football game, a player might pass the ball to another player upon his verbal request (e.g. pass me the ball, Bob). Therefore, for applications such as video summarization or analysis that require understanding of both human actions and communication, communication action recognition becomes a pre-requisite.

As the name implies, a multi-party conversation involves more than two participants. When the number of participants in a conversation increases, communication action recognition becomes difficult mainly because it is difficult to automatically understand who is talking to whom. The most common approach to identify communication actions is by using a machine learning technique on a corpus manually tagged with communication actions [11]. The problem with this approach is that for each domain, it requires a corpus annotated with communication actions applicable to that domain. In contrast, this paper presents the use of domain-independent dialog act and addressee information related to participant communications in identifying communication actions.

This approach employs three interconnected components for communication action recognition. A machine learning approach has been employed for dialog act recognition, and

rule-based algorithms are used for addressee recognition and communication action recognition. The intuition is to make use of domain-independent language understanding techniques for dialog act and addressee recognition. Identification of communication actions is done in a domain-specific manner, based on the output of the first two steps, along with the key words and constraints already available for a given domain. Because of the domain-independent nature of the dialog act recognition step, the classifier trained with a corpus belonging to one domain can be easily used for another domain. In order to identify communication actions in a different domain, this trained classifier and the addressee recognition algorithm can be reused. A domain-specific rule set should be generated only for the last component. The experiment results show that this approach gives a good overall accuracy for communication action recognition.

As the initial phase, we only focus on typed-chat conversations. To be more specific, the chat data used in this work were retrieved from a medical training scenario that took place inside the Otago Virtual Hospital (OVH) system [2].

The rest of the paper is organized as follows. Section II discusses related work. Section III discusses the virtual world scenario that provided the data for our experiments, and the composition of the corpus. Section IV, V and VI discuss the three steps of our communication recognition approach. Section VII discusses the experimental results and finally, Section VIII concludes the paper.

II. RELATED WORK

Dialog act recognition is an important sub-area of natural language processing, and is an important step in understanding spontaneous dialog [15]. Traditional dialog act recognition has mainly focused on face-to-face conversations or telephone conversations [15]. Wu et al. [18] applied dialog act recognition techniques to chat conversations taking place in the Internet relay chat (IRC) channels. They introduced the term ‘posting act tagging’ to refer to the dialog act recognition task for chat room data. In recent research, Forsyth and Martell [6] applied dialog act recognition techniques to data collected from public chat rooms.

Machine learning approaches have been a common choice for dialog act recognition. Training a classifier to identify dialog acts involves several steps:

- Identification of a labelling scheme, i.e., the dialog act tag set
- Selecting a classifier that provides good accuracy and performance
- Identification of a feature set that gives the optimal classifier performance for the selected dialog act tag set

Ample research can be found that discusses each of these steps. There are dialog act tag sets defined for different applications [15], while some have presented domain-independent dialog act labelling systems, such as the DAMSL tag set [4]. As for online chat classification, Wu et al. [18] have introduced a set of 15 dialog acts. Naïve Bayes classifier, Support Vector Machines (SVM), Maximum Entropy, Decision Trees and Conditional Random Fields (CRF) are some of the classification techniques applied for dialog act recognition. The final step of dialog act recognition is to identify a suitable feature set that will give the highest accuracy for the classifier. For this step also, the related research reports many different possible feature selections. These include sentence length, word ngrams, features of utterances in the vicinity, and specific cue phrases [17].

In the related literature, the intended recipient or the receiver of an utterance is termed an addressee. According to Goffman [7], the targeted addressee of a piece of utterance is “oriented by the speaker in a manner that suggests that his words are particularly for them, and that some answer is therefore anticipated from them” rather than from the other participants active in the conversation, or any over hearer.

Just like in many fields related to natural language processing, machine learning approaches have been a natural choice for addressee recognition. As for face-to-face group conversations, lexical (e.g. the number of words in the message content), discourse (e.g. the dialog act), and visual (e.g. the gaze direction) features have been used to train classifiers using the AMI human meeting corpus [13], [8], [1]. For human-agent communication, Baba et al. [1] made use of acoustic and visual queues retrieved from a video stream to enable a virtual agent to identify when it is being addressed by two other participants in real world. In similar research, Katzenmaier et al. [9] made use of head pose and speech-related features to implement addressee recognition by a simulated robot.

Related work for communication action recognition can be found in the area of meeting summarization, detecting meeting decisions in particular [5], [3]. More relevant to our work is the work by Purver et al. [12], which identifies action terms (public commitments to perform a given task). All this research has used either the AMI meeting corpus or the ICSI meeting corpus, which are publicly available popular corpora containing spoken dialog. Therefore the decision identification considered situations where only participant communication is taking place. In contrast, in our work, the concept of communication actions was introduced with the understanding that spoken dialog can co-exist with other participant actions such as doctor taking history of patient while carrying out different tests, or player communication in a football match. In this respect, the work of Orkin and Roy [11] is of interest. They have presented a classifier that abstracts communication messages exchanged in a virtual environment as \langle speech act; propositional content; referent \rangle triples. This abstraction had been done in such a way that the abstracted information can be used alongside other actions of participants inside the virtual environment. However, none of this research has focused on using addressee information in abstracting these conversations, although it has been identified as an important step in the process [12].

III. SCENARIO AND CORPUS

The Otago Virtual Hospital Simulation (OVH) resembles the Emergency Department (ED) in a New Zealand hospital. It is a “virtual hospital in which medical students, playing the role of junior doctors, solve open-ended clinical cases” [2]. Currently, the simulation runs two training scenarios — the UTI patient scenario and the accident victim scenario.

From these two scenarios, the UTI patient scenario is more complex. It involves an elderly lady suffering from a urinary tract infection. She has been brought into the Emergency Department, and the doctors’ responsibility is to diagnose the illness and prescribe medicine. In the accident victim scenario, a motorcyclist who has met with a road accident has been brought into the ED. He is unconscious, so he cannot communicate with the doctors. The doctors’ responsibility is to observe the patient’s oxygen level, breathing, etc. using the provided medical equipment, and give him initial treatments. The participants in these virtual hospital scenarios converse with each other through the public chat channel of the virtual world.

The OVH chat corpus contains 29 log files, and 5978 chat messages. These log files were provided in two phases. Log files of phase II were the result of formal runs of the urinary tract infected (UTI) patient scenario using medical students. In contrast, log files from phase I correspond to both the UTI patient scenario, and the accident victim scenario. These scenario runs have been conducted as test and demonstration runs. In addition, logs from phase I contained chat posts related to general conversations of simulation implementers during system testing. Therefore, when taken as a whole, the OVH chat corpus is not domain-specific.

The corpus was divided into a training set and a test set. In order to allow for the possible requirement of analysing correlations among different chat posts, the test set was grouped into blocks. Six blocks from the chat corpus were selected (1–100, 1000–1099, ..., 5000–5099). This resulted in 600 chat posts for the test set, accounting for 10% of the corpus.

IV. DIALOG ACT RECOGNITION

Selecting the correct dialog act tag (dtag) set is crucial in successful recognition of dialog acts. The number of dialog acts to be used depends on the size of the training corpus, and the type of analysis expected to be done on the corpus.

For the current analysis, a set of 14 dialog acts was identified. This tag set is a subset of the tag set by Wu et al. [18] defined for online chat communications, and has three additional dtags required for further analysis of the corpus. These are highlighted in Table I.

TABLE I
DIALOG ACT TAG SET

Tag	Example
Accept	True
Apology	Sorry I’m bit late
Conventional closing (bye)	See you soon
Continuer	And he did not come
Emotion	Oh dear :(
Conventional opening (greet)	Nice to meet you
No-Answer (nAnswer)	No, he is not like that
Other	*****
Request	Can I have some water?

Statement	I am not feeling well
Thank	Oh thank you dear
Wh-Question	What is your name?
Yes-Answer (yAnswer)	Yes, that could be it
Yes-no-Question (ynQuestion)	Do you have a cough?

The corpus was manually tagged using these dtags. Each utterance was assigned only one dtag. If more than one sentence was present, the dtag that is most relevant in that context was selected.

In selecting the feature set, the feature set provided by Forsyth and Martell [6] for online chat classification was selected as the starting point. Table II shows this feature set, and the rationale for selecting each feature. From these, the first 12 features are from the feature set given by Forsyth and Martell [6]. After experimenting with different combinations of new features, the last six features were newly added.

TABLE II
FEATURE SET

Feature	Rationale
Total number of words in a post	Indicates short answers, and Emotions
First word is a conjunction, preposition, or ellipse (POS tag of 'CC', 'IN', or ':')	Indicates a Continuer
Contains part of speech (POS) tags 'WRB'(Wh-adverb) or 'WP'(Wh-pronoun)	Indicates a wh-Question
Has an emotion abbreviation	Indicates an Emotion
Has an emoticon (e.g. :(, :P)	Indicates an Emotion
Has a greeting word (e.g. hello, hi)	Indicates a Greet
Has a bye word (e.g. bye, cya)	Indicates a Bye
Has 'yes' or yes variant (e.g. yes, yeah)	Indicates an yes-Answer
Has 'no' or no variant (e.g. no, nopes)	Indicates a no-Answer
Contains question mark	Indicates a question
Contains exclamation mark	Indicates an Emotion tag
A word in the sentence is in upper case and has a word length > 1	Indication of short answers s.a. OK
Contains an accept word (e.g. ok, sure)	Indicates an Accept
Has a repetition of a character for 3 or more times	Indicates an Other tag
One of the first three words of the sentence contains a question word	Indicates a Question
Contains a request word (e.g. please, can)	Indicates a Request
Contains an apology word (e.g. sorry)	Indicates an Apology
Contains a thanking word (e.g. thanks)	Indicates a Thank

V. ADDRESSEE RECOGNITION

Machine learning techniques for addressee recognition discussed in Section II have assumed that the number of participants and their vocatives are fixed throughout the

corpus. However, different chat files in our scenarios contained different numbers of participants and their vocatives were not the same. Therefore, for the addressee recognition, a rule-based algorithm was employed.

To start with, two existing algorithms for addressee recognition [16], [10] were investigated. In Traum's [16] algorithm, the addressee and speaker information in the previous message, along with vocatives, are used to identify the addressee. The algorithm of Knott and Vlugter [10] also uses a similar approach to Traum's, however it is capable of handling group addressing. Both these algorithms make two assumptions: that an explicit addressee can be correctly identified using a vocative, and that the addressee identified in the previous message is always correct. This led to a low-level addressee recognition accuracy for our chat data.

Therefore, based on these two basic algorithms, Algorithm 1 was implemented. The input to the algorithm is the chat message. The algorithm returns the identified text value corresponding to the addressee or the set of addressees. Here, sp, slm and alm refer to the speaker in the current message, speaker in the last message, and the addressee of the last message, respectively.

As in the aforementioned two algorithms, the first step of Algorithm 1 (line 2) is to identify any vocative terms that refer to explicit addressees. However, vocative term identification is improved using three techniques. (1) If a participant name has more than one part, both name parts are separately considered. When a participant name is detected, the three words that immediately follow the name are checked for the existence of any other vocative, thus leading to the identification of multiple addressee terms. (2) Each vocative is considered eligible for being an addressee term only if it is not immediately preceded or succeeded by certain Part-of-Speech tags. These Part-of-Speech tags have a very low probability of occurring immediately before or after (respectively) a vocative. The former are PRP\$, (possessive pronouns), TO (the word 'to'), and WDT (wh-determiner). Similarly, Part-of-Speech tags for the latter are VBZ (third person present tense verb) and the Part-of-Speech tag that stands for possessive ending.

In conversations, participants sometimes tend to repeat what the previous participant said. Therefore, in the second step of the algorithm (line 4), a similarity measure between the current message content and the set of previous message content is computed, with respect to noun word frequency. Words with Part-of-Speech tags NN (singular noun), NNS (plural noun), NNP (singular proper noun), and NPS (singular plural noun) are considered as noun terms. The frequency of each noun word in the current message content is calculated. From these, the set of nouns (Ns) that have an occurrence frequency (nf) greater than some threshold (tf) is selected. From the set of preceding chat messages, the most recent message that contains any of the nouns in Ns is selected. Then the speaker of this message is set as the addressee of the current message.

The third and fourth rules in the algorithm (lines 6-10) are similar to the algorithm of Traum. As can be seen, addressee recognition based on vocative terms and noun word frequency is given higher priority than speaker and addressee information in the previous message. This removes many of the false negatives of simply using previous message information.

This algorithm is domain-independent. It is possible to make use of domain-specific information to improve the accuracy of the algorithm. Such an improvement is reported elsewhere [14].

Algorithm 1 Outline of addressee recognition algorithm

```

1: addressee = ""
2: if message content has an explicit vocative(s) satisfying
   certain conditions then
3:   addressee = vocative(s)
4: else if there exists a previous message (PM) with
   content containing a noun with  $nf > tf$  AND  $sp \neq sp$  of PM
   then
5:   Set the addressee to  $sp$  of PM
6: else if  $sp = slm$  then
7:   addressee =  $alm$ 
8: else if  $slm \neq sp$  then
9:   addressee =  $slm$ 
10: end if
11: if addressee has not been set then
12:   addressee = 'ALL'
13: end if

```

VI. IDENTIFYING DOMAIN SPECIFIC COMMUNICATION ACTIONS

In this work, only communication actions of doctors in the UTI patient scenario are identified. Table III shows the communication actions that are currently defined.

TABLE III
COMMUNICATION ACTIONS OF DOCTORS

Communication Action	Example
Doctor greeting the patient	Hello Gertrude
Doctor introducing himself to patient	I am doctor Ben
Doctor taking history	Are you a smoker?
Doctor informing about test	I will have a look at your tummy now
Doctor informing illness	I think you have got an infection
Doctor informing treatment	I am going to give you some antibiotics
Other	When an utterance does not fall into any of the above categories

Algorithm 2 is used to identify these communication actions. It makes use of the dialog acts and addressee information identified in the previous two steps. To start with, an utterance is processed by the algorithm only if its speaker is a doctor, and the addressee is the patient. An utterance is labelled as *greet patient* if the associated dtag is 'Greet'. If the name of a doctor or the terms 'doctor' and 'doctors' is preceded by an introduction phase (e.g. 'I am', 'my name is'), the utterance is identified as the communication action *introduce to patient*. If the dtag of an utterance is either 'Statement', or 'Request', and it contains a phrase describing a test, this utterance refers to the *inform test* communication action. Similarly, if the utterance has the dtag 'Statement', and it has a phrase describing the illness or possible treatment, then this utterance is categorized as *inform illness*, or *inform*

treatment, respectively. An utterance tagged as a 'ynQuestion' or 'whQuestion' is identified as *take history*. Finally, if the utterance is tagged as 'Continuer', the communication action assigned to the previous utterance of the same speaker is assigned. An utterance that has not been associated with a communication action by the above rules is marked as other.

Note that unlike in dialog act recognition, it is possible to assign more than one communication action to an utterance. The test terms, illness terms, and treatment terms used in Algorithm 2 were extracted from the patient role-playing script, names and functionalities of the objects in the simulation, and types of object clicking actions allowed to be performed by a doctor.

Algorithm 2 Outline of the algorithm for communication action recognition

```

1: string com action = ""
2: if speaker is the doctor AND addressee is the patient
   then
3:   if dtag = 'Greet' then
4:     com action = 'greet patient'
5:   end if
6:   if name of the doctor is preceded by one of the
   introduction terms then
7:     com action = com action + ',' + 'introduce to
   patient'
8:   end if
9:   if (dtag = 'Statement' OR dtag = 'Request') AND
   utterance contains a test term then
10:    com action = com action + ',' + 'inform test'
11:   end if
12:   if dtag = 'Statement' then
13:     if Utterance contains an illness term then
14:       com action = com action + ',' + 'inform illness'
15:     end if
16:     if Utterance contains a treatment term then
17:       com action = com action + ',' + 'inform
   treatment'
18:     end if
19:   end if
20:   if dtag = 'ynQuestion' OR dtag = 'whQuestion' then
21:     com action = com action + ',' + 'take history'
22:   end if
23:   if dtag = 'Continuer' then
24:     com action = com action of the speaker's
   previous utterance
25:   end if
26: end if
27: if no communication action has been identified then
28:   com action = 'other'
29: end if

```

VII. RESULTS

For dialog act recognition, 10-fold cross validation [19] was carried out on the training set using the Naïve Bayes classifier [20]. The average accuracy of the selected feature set was 0.72612. In contrast, the base line test that used the Bag of Words (BOW) feature set provided an average accuracy of only 0.227. The mean and standard deviation for

precision, recall and F-Score of each dtag are given in Table IV.

TABLE IV
MEAN AND STANDARD DEVIATION FOR PRECISION, RECALL, AND F-SCORE
IN 10-FOLD CROSS VALIDATION

Dtag	Precision		Recall		F-Score	
	Mean	Std	Mean	Std	Mean	Std
Greet	0.88	0.096	0.851	0.104	0.859	0.069
whQuestion	0.777	0.131	0.751	0.068	0.758	0.077
Statement	0.73	0.036	0.826	0.028	0.774	0.019
Accept	0.604	0.117	0.533	0.149	0.559	0.125
ynQuestion	0.826	0.062	0.716	0.031	0.766	0.03
Thank	0.926	0.088	0.883	0.095	0.899	0.057
Request	0.413	0.132	0.252	0.099	0.31	0.112
Continuer	0.664	0.113	0.645	0.076	0.653	0.09
yAnswer	0.651	0.155	0.712	0.097	0.671	0.11
nAnswer	0.657	0.089	0.759	0.1	0.701	0.082
Emotion	0.758	0.315	0.478	0.277	0.547	0.286
Other	0.646	0.44	0.261	0.262	0.378	0.292
Bye	1	0	0.323	0.263	0.577	0.218
Apology	0.885	0.106	0.813	0.135	0.843	0.101

The classifier was then trained using the entire training corpus, and was tested on the test set to validate the result received in 10-fold cross validation. This gave an accuracy of 0.722. The precision, recall and F-Score of each dtag for the test set are given in Table V.

TABLE V
PRECISION, RECALL, AND F-SCORE VALUES FOR THE TEST SET

Dtag	Precision	Recall	F-Score
Greet	0.958	0.742	0.836
whQuestion	0.8	0.727	0.762
Statement	0.733	0.83	0.778
Accept	0.429	0.652	0.517
ynQuestion	0.839	0.768	0.802
Thank	1	0.778	0.875
Request	0.25	0.222	0.235
Continuer	0.667	0.533	0.593
yAnswer	0.694	0.926	0.794
nAnswer	0.581	0.545	0.563
Emotion	None	0	None
Other	0	0	0
Bye	None	0	None
Apology	1	0.8	0.889

Chat logs received in phase II of UTI scenario contained only three participants. In contrast, chat logs received in phase I for the UTI patient scenario contained 5-7 participants. Due to these different compositions in the two phases, the results of addressee recognition for each phase are separately reported.

Table VI reports the results of the addressee recognition step for phase I training data. The addressee recognition algorithm given by Traum was used as the baseline test case.

TABLE VI
ACCURACY COMPARISON OF ADDRESSEE RECOGNITION ALGORITHMS FOR
PHASE I TRAINING SET

log file	participant count	Traum's Alg.	Alg. 1
g1	5	0.822	0.839
g2	7	0.655	0.694
g3	5	0.689	0.727
g4	7	0.476	0.481
g5	6	0.644	0.685

Table VII shows the McNemar's significance test values for the five log files, and the significance test value for all the log files taken as a whole. Three out of five log files show a significant increase in accuracy ($p < 0.05$) for Algorithm 1 over Traum's algorithm.

TABLE VII
MCNEMAR'S SIGNIFICANCE TEST VALUES FOR PHASE I TRAINING SET

Log File	Significance value for accuracy increase for Alg. 1 over Traum's Alg.
g1	0.248
g2	0
g3	0.003
g4	0.074
g5	0.008
all	0

Table VIII reports the results of the addressee recognition step for phase I test data.

TABLE VIII
ACCURACY COMPARISON OF ADDRESSEE RECOGNITION ALGORITHMS FOR
PHASE I TEST SET

Log file	Traum's Alg.	Alg. 1
g1	0.74	0.78
g3	0.53	0.58
g5	0.469	0.484

Table IX shows the accuracy of addressee recognitions on the training set of phase II.

TABLE IX
ACCURACY COMPARISON OF ADDRESSEE RECOGNITION ALGORITHMS FOR
PHASE II TRAINING SET

Log File	Traum's Alg.	Alg. 1
m1	0.941	0.949
m2	0.924	0.93
m3	0.975	0.975
m4	0.948	0.948
m5	0.919	0.919
m6	0.969	0.969
m7	0.971	0.971
m8	0.948	0.961
m9	0.981	0.981
m10	0.981	0.987
m11	0.951	0.96
mean	0.955	0.959
stdev	0.022	0.021

As can be seen, the accuracy of Algorithm 1 compared with Traum's algorithm is not significantly higher. To conclude, if a chat conversation has more participant interactions, our algorithm performs better than Traum's algorithm.

When observing the conversations in phase I, it was evident that sometimes the participants have formed subgroups in different locations inside the OVH simulation. However, all this chat is communicated in the same chat channel, making these sub group chats interleaved with each other. This has affected the accuracy of the addressee recognition algorithm. In order to correctly identify addressee in such situations, position information of participants should be considered.

Accuracy of communication action recognition was 0.961, 0.96 and 0.99 for the training set, phase I test set and phase II test set, respectively.

A. Inter-Rater Reliability

In order to validate the manual annotation of the chat corpus done by the author, an inter-rater reliability (IRR) test was conducted. Three blocks of 200 chat posts (i.e., 10% of the corpus) were selected. Six individuals who had no involvement with this research were selected. Each block was given to two participants, so that this 10% of the corpus was tagged by three annotators (the author, and two participants).

Cohen's Kappa value for dialogue act tagging, addressee tagging, and communication action tagging are 0.70 (acceptable), 0.90 (good), and 0.85 (good), respectively.

VIII. CONCLUSIONS

This paper presented the use of dialog act and addressee information in identifying communication actions in multi-party conversations. Here, dialog act recognition and addressee recognition is carried out in a domain-independent manner. Output of these is then used by a domain-specific rule set to identify the communication actions in a given domain. Therefore in order to identify communication actions in a new domain all that is required is to come up with this domain-specific rule set. This technique eliminates the need of domain-specific corpora for communication action recognition. This paper also presented an improved version of a commonly used addressee recognition algorithm. This improved version proved to be performing better than the original one, when the number of participants was higher. The dialog act recognition step gave an accuracy of 0.722. By employing a more sophisticated classifier, this accuracy could be improved.

Although this paper focused on virtual world chat conversations, we see no barrier in applying the same techniques for identifying face-to-face human conversations. However, this requires a speech-to-text module that first provides the textual representation of spoken conversations. Moreover, if the conversation participants are spatially distributed, their position information has to be collected and used in addressee recognition. The current system lacks this capability. As additional future work, we are exploring how communication action recognition can be done for other languages such as Sinhala.

ACKNOWLEDGMENT

The author would like to thank (1) Professor Stephen Cranefield and Professor Martin Purvis at the Department of Information Science, University of Otago for their guidance in her PhD research, out of which this paper was produced, (2) the OVH team - Phil Blyth, Judith Swan, Swee-Kin Loke, and Meylissa Roy, for allowing to work in the Otago Virtual Hospital (OVH) simulation, and for providing the chat data, and (3) Angrosh Mandya, Chandima Daskon, Christopher Frantz, Duminda Kurupparachchi, Kalpana Kathriarachchi, Rasika Withanawasam and Sharmila Savarimuthu for annotating the OVH chat corpus.

REFERENCES

- [1] N. Baba, H.-H. Huang, and Y. Nakano. Identifying utterances addressed to an agent in multiparty human-agent conversations. In H. Vilhjálmsson, S. Kopp, S. Marsella, and K. Th'orisson, editors, *Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, pages 255–261. Springer, 2011.
- [2] P. Blyth, S.-K. Loke, and J. Swan. Otago Virtual Hospital: Medical students learning to notice clinically salient features. In *Curriculum, Technology & Transformation for an Unknown Future. Proceedings ascilite Sydney 2010*, pages 108–112, 2010.
- [3] T. H. Bui, M. Frampton, J. Dowding, and S. Peters. Extracting decisions from multi-party dialogue using directed graphical models and semantic similarity. In *the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue*, pages 235–243. Association for Computational Linguistics, 2009.
- [4] M. G. Core and J. F. Allen. Coding dialogs with the DAMSL annotation scheme. In *Working Notes of the AAIL Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, 1997.
- [5] R. Fernández, M. Frampton, P. Ehlen, M. Purver, and S. Peters. Modelling and detecting decisions in multi-party dialogue. In *the 9th SIGdial Workshop on Discourse and Dialogue*, pages 156–163. Association for Computational Linguistics, 2008.
- [6] E. N. Forsyth and C. H. Martell. Lexical and discourse analysis of online chat dialog. In *Proceedings of the First IEEE International Conference on Semantic Computing*, pages 19–26. IEEE Computer Society Press, 2007.
- [7] E. Goffman. Replies and responses. In *Forms of Talk*, pages 5–77. University of Pennsylvania Press, 1981.
- [8] N. Jovanović. *To whom it may concern: Addressee identification in face-to-face meetings*. PhD thesis, University of Twente, 2007.
- [9] M. Katzenmaier, R. Stiefelwagen, and T. Schultz. Identifying the addressee in human-human-robot interactions based on head pose and speech. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 144–151. ACM, 2004.
- [10] A. Knott and P. Vlugter. Multi-agent human-machine dialogue: Issues in dialogue management and referring expression semantics. *Artificial Intelligence*, 172(2-3):69–102, 2008.
- [11] J. Orkin and D. Roy. Semi-automated dialogue act classification for situated social agents in games. In F. Dignum, editor, *Agents for Games and Simulations II*, volume 6525 of *Lecture Notes in Computer Science*, pages 148–162. Springer, 2011.
- [12] M. Purver, J. Dowding, J. Niekrasz, P. Ehlen, S. Noorbaloochi, and S. Peters. Detecting and summarizing action items in multi-party dialogue. In *the 8th SIGdial Workshop on Discourse and Dialogue*, pages 18–25. Association for Computational Linguistics, 2007.
- [13] M. Purver, R. Fernández, M. Frampton, and S. Peters. Cascaded lexicalised classifiers for second-person reference resolution. In *the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue*, pages 306–309. Association for Computational Linguistics, 2009.
- [14] S. Ranathunga. *Improving awareness of intelligent virtual agents*. PhD thesis, University of Otago, 2013.
- [15] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Ess-Dykema, and M. Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373, 2000.
- [16] D. Traum. Issues in multiparty dialogues. In F. Dignum, editor, *Advances in Agent Communication*, volume 2922 of *Lecture Notes in Computer Science*, pages 1954–1954. Springer, 2004.
- [17] D. Verbree, R. Rienks, and D. Heylen. Dialogue-act tagging using smart feature selection: Results on multiple corpora. In B. Raorke,

- editor, *First International IEEE Workshop on Spoken Language Technology*, pages 70–73. IEEE Computer Society Press, 2006.
- [18] T. Wu, F. M. Khan, T. A. Fisher, L. A. Shuler, and W. M. Pottenger. Posting act tagging using transformation-based learning. In T. Y. Lin, S. Ohsuga, C.-J. Liao, X. Hu, and S. Tsumoto, editors, *Foundations of Data Mining and Knowledge Discovery*, volume 6 of *Studies in Computational Intelligence*, pages 319–331. Springer, 2005.
- [19] Bishop, Christopher M. *Pattern recognition and machine learning*. Vol. 1. New York: springer, 2006.
- [20] Murphy, Kevin P. "Naive Bayes classifiers." *University of British Columbia*(2006).