AIMU: Actionable Items for Meeting Understanding

Yun-Nung Chen and Dilek Hakkani-Tür
Microsoft Research, Redmond, WA, USA
y.v.chen@ieee.org, dilek@ieee.org

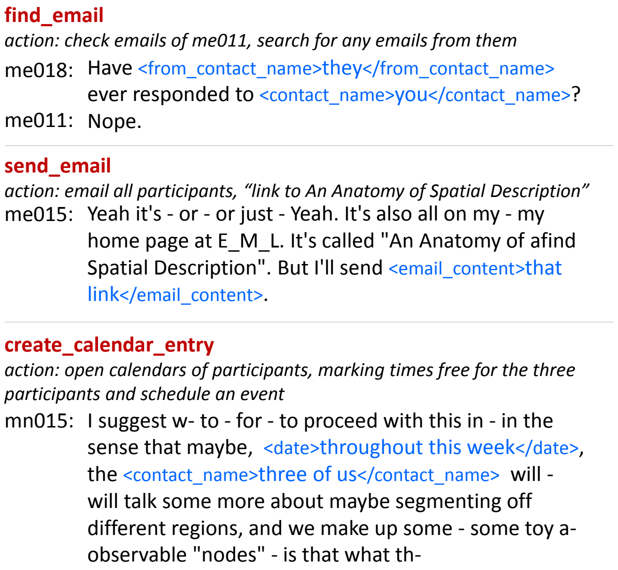
Abstract

With emerging conversational data, automated content analysis is needed for better data interpretation, so that it is accurately understood and can be effectively integrated and utilized in various applications. ICSI meeting corpus is a publicly released data set of multi-party meetings in an organization that has been released over a decade ago, and has been fostering meeting understanding research since then. The original data collection includes transcription of participant turns as well as meta-data annotations, such as disfluencies and dialog act tags. This paper presents an extended set of annotations for the ICSI meeting corpus with a goal of deeply understanding meeting conversations, where participant turns are annotated by actionable items that could be performed by an automated meeting assistant. In addition to the user utterances that contain an actionable item, annotations also include the arguments associated with the actionable item. The set of actionable items are determined by aligning human-human interactions to human-machine interactions, where a data annotation schema designed for a virtual personal assistant (human-machine genre) is adapted to the meetings domain (human-human genre). The data set is formed by annotating participants' utterances in meetings with potential intents/actions considering their contexts. The set of actions target what could be accomplished by an automated meeting assistant, such as taking a note of action items that a participant commits to, or finding emails or topic related documents that were mentioned during the meeting. A total of 10 defined intents/actions are considered as actionable items in meetings. Turns that include actionable intents were annotated for 22 public ICSI meetings, that include a total of 21K utterances, segmented by speaker turns. Participants' spoken turns, possible actions along with associated arguments and their vector representations as computed by convolutional deep structured semantic models are included in the data set for future research. We present a detailed statistical analysis of the data set and analyze the performance of applying convolutional deep structured semantic models for an actionable item detection task. The data is available at `http://research.microsoft.com/projects/meetingunderstanding/`.

Keywords: actionable item, meeting understanding, convolotional deep structured semantic model (CDSSM), embeddings.

1. Introduction

Meetings pose unique knowledge sharing opportunities, and have been a commonly accepted practice to coordinate the work of multiple parties in organizations. With the surge of smart phones, computing devices have been easily accessible during meetings and real-time information search has been a common part of regular conversations (Brown et al., 2015). Furthermore, recent improvements in conversational speech recognition suggest the possibility of automatic speech recognition and understanding on continual, in the background, audio recording of conversations (McMillan et al., 2015). In meetings, discussions could be a rich resource for identifying participants' next actions and helping them to accomplish those in a seamless way during or after the meeting, without interrupting the discussions.



find_email
action: check emails of me011, search for any emails from them
me018: Have <from_contact_name>they</from_contact_name> ever responded to <contact_name>you</contact_name>?
me011: Nope.

send_email
action: email all participants, "link to An Anatomy of Spatial Description"
me015: Yeah it's - or - or just - Yeah. It's also all on my - my home page at E_M_L. It's called "An Anatomy of afind Spatial Description". But I'll send <email_content>that link</email_content>.

create_calendar_entry
action: open calendars of participants, marking times free for the three participants and schedule an event
mn015: I suggest w- to - for - to proceed with this in - in the sense that maybe, <date>throughout this week</date>, the <contact_name>three of us</contact_name> will - will talk some more about maybe segmenting off different regions, and we make up some - some toy a- observable "nodes" - is that what th-

Figure 1: Actionable item examples in meeting corpus.

This paper proposes a novel task of actionable item detection in meetings, with the goal of providing the participants easy access to information and performing actions that a personal assistant would handle. Actionable items in meetings include discussions on scheduling, emails, action items, and search. Figure 1 shows some meeting segments from the ICSI meeting corpus (Janin et al., 2003), where actionable items and their associated arguments are annotated. A meeting assistant would then take an appropriate action, such as opening the calendars of the involved participants for the dates being discussed, finding the emails and documents being discussed, or initiating a new one. The actions could be taken during the meeting (for example, the automated assistant can display each participant their calendars when they are discussing when to meet) or after the meeting (for example, when a meeting participant commits to an action item during a meeting, s/he may be sent a note or reminder about that action item, later on). Our annotations on a subset of the ICSI meetings corpus (Janin et al., 2003) enable research on actionable item detection and associated argument extraction tasks, and the study of appropriate user interface designs for implementation of these actionable items is left as a future work.

Previous work on meeting understanding investigated detection of decisions (Bui and Peters, 2010; Fernández et al., 2008), action items (Yang et al., 2008), agreement and disagreements (Galley et al., 2004; Hillard et al., 2003), and summarization (Riedhammer et al., 2010; Xie et al., 2009; Chen and Metze, 2013). Our task is closest to detection of action items, actually, action items are considered as a subgroup of actionable items that can be actionable in the form of reminders or to do lists.

2. AIMU Data Set

The actionable items annotations are performed on a subset of the ICSI meeting corpus (Adam Janin and Wooters, 2004a; Adam Janin and Wooters, 2004b), where 22 meetings that were used as test and development data sets in previous work (Janin et al., 2003; Ang et al., 2005) are included for the actionable item detection task[1]. These include three types of weekly meetings, Bed, Bmr, and Bro, which include regular project discussions between colleagues and conversations between students and their advisers[2]. The meeting types and associated data sizes are shown in Table 1, and the meeting types are detailed as below as also described in the ICSI meeting corpus data collection documentation (Janin et al., 2003).

Name	Type	#Utt
Bed	Even Deeper Understanding	4,544
Bmr	Meeting Recorder	9,227
Bro	Robustness	7,264

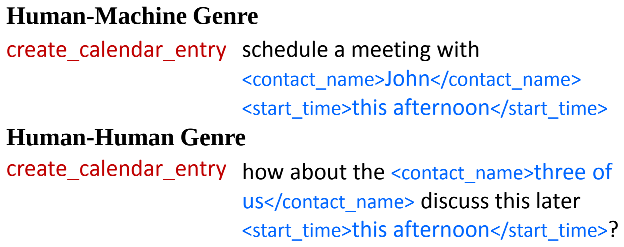
Table 1: The data set description

- Even Deeper Understanding meetings focus mainly on issues in natural language understanding and neural theories of language.
- Meeting Recorder meetings are concerned mostly with the ICSI meeting corpus data collection project, but include some discussions of more general speech research.
- Robustness meetings focus on signal processing techniques to compensate for noise, reverberation and other environmental issues in speech recognition.

3. Semantic Intent Schema

To collect actionable resources in meetings, we annotate each utterance that can trigger an actionable item with the corresponding intent and associated arguments (i.e., slot-fillers).

Although utterances in the human-human genre are more casual and include conversational terms, some intents and the terms related to the actionable item, such as dates, times, and participants are similar in terms of form and content to the ones in human-machine genre. Figure 2 shows utterance examples with the same intents, create_calendar_entry, from different genres (human-machine v.s. human-human). Therefore, we apply the semantic intent schema for an intelligent assistant to meeting conversations.



Human-Machine Genre

create_calendar_entry schedule a meeting with <contact_name>John</contact_name> <start_time>this afternoon</start_time>

Human-Human Genre

create_calendar_entry how about the <contact_name>three of us</contact_name> discuss this later <start_time>this afternoon</start_time>?

Figure 2: Utterance examples with the same action.

We chose five domains that may contain actions triggered by utterances in meetings, where there are total 10 intents in these domains. Table 2 shows the detailed schema and description.

4. Annotation Agreement

Randomly selected two meetings were annotated by two annotators, and we tested the agreement for three meeting types using Cohen's Kappa coefficient (Cohen and others, 1960). The agreements are shown in Table 3.

- Actionable Utterance Agreement
  We treat 10 defined actions as positive and others as negative (binary) to compute the actionable utterance agreement. The average agreement about whether an utterance includes an actionable item is 0.644.
- Action Type Agreement
  To deeply analyze action types cross annotators, we compute the agreement on the actionable utterances that both annotators agree with (positive). The average agreement is 1.000, indicating that both annotators always decide on the same action if they both agree that there is an action in this utterance. It also suggests that most actions may not be ambiguous.
- Overall Agreement
  We treat 10 actions and others as total 11 considered labels, and the average agreement about the annotations is 0.673, showing that the actionable items are consistent across human annotators.

5. Statistical Analysis

Actionable items were manually annotated based on the designed schema. There are total 318 turns annotated with actionable items, which account for about 1.5% of all of the turns. The detailed numbers are shown in Table 4, where the number of actionable utterances in Bed meetings are much more than ones in Bmr and Bro meetings. It suggests that the number of actionable utterances may depend on the meeting type. For example, a project status update meeting usually contains more actionable items about reminders, so there is a difference across meeting types and across meeting groups.

[1] The data is available at http://research.microsoft.com/projects/meetingunderstanding/.

[2] Bed (003, 006, 010, 012), Bmr (001, 005, 010, 014, 019, 022, 024, 028,030), Bro (004, 008, 011, 014, 018, 021, 024, 027)

Domain	Action		Description
Calendar	Intent	find_calendar_entry	check the calendar of participants to find a specific event
		create_calendar_entry	create a new event in the opening period of participants' calendars
		open_agenda	check the meeting agenda
		add_agenda_item	create a new entry for the meeting agenda
	Argument	contact_name	owners of the targeted calendars
		start_date, end_date	exact date or generic descriptions like "tomorrow" or "yesterday"
		start_time, end_time	exact time or generic descriptions like "afternoon"
		entry_type	e.g. "meeting", "talk", "discussion"
		title	the meeting goal, e.g. "discussion on finite state automaton"
		absolute_location	exact location
		implicit_location	implicitly described location
		agenda_item	the content of the agenda item
Reminders	Intent	create_reminder	create a reminder of participants
	Argument	contact_name	the person who should get the reminder or the speaker when note taking
		reminder_text	the reminder content, also could be a referral
		start_date	targeted date/meeting reference; exact time or generic descriptions
		start_time	exact time or generic descriptions
Communication	Intent	send_email	initialize an email to someone
		find_email	search the specific content from email
		make_call	dial a phone call to someone
	Argument	contact_name	person/people who will be contacted
		email_subject	what the email is about
		email_content	what to include in the email (could also be a description, such as "the paper")
		from_contact_name	the email sender (could be the speaker of the utterance)
OnDevice	Intent	open_setting	launch the setting of devices (e.g. computer, projector)
	Argument	setting_type	the type to modify
Search	Intent	search	retrieve information through the search engine
	Argument	query_term	word sequence to search for

Table 2: The description of the semantic intent schema for meetings

Agreement	Bed003	Bed010	Average
Actionable	0.699	0.642	0.644
Type	1.000	1.000	1.000
Overall	0.70	0.646	0.673

Table 3: Annotation agreement during different settings.

Meeting	#Utt	#Utt w/ Actions	Percentage
Bed	4,544	192	4.2%
Bmr	9,227	116	1.3%
Bro	7,264	110	1.5%
Total	21,035	318	1.5%

Table 4: The annotation statistics

Figure 3 shows actionable item distribution in the meeting corpus, where it can be found that different types of meetings contain slightly different distribution of actionable items, but some actions frequently occur in all meetings, such as create_single_reminder and find_calendar_entry. Some actions such as open_setting and make_call rarely appear in all meetings.

6. Intent & Utterance Embeddings

In addition to the data set, we also include trained intent and utterance embeddings into the data set, which can provide additional features for researchers to utilize or further explore. We perform convolutional deep structured semantic models (CDSSM) to train vector embeddings for this task (Huang et al., 2013; Shen et al., 2014; Chen et al., 2015a), where with all utterance-action pairs, a set of parameters is optimized based on an objective. The objective is to minimize the distance between utterance embeddings and the corresponding intent embeddings in the continuous latent space, so that the vector embeddings can be used as features to represent utterances and intents. The models with two directions for optimization are performed, a predictive model and a generative model (Chen et al., 2015a). The predictive model estimates the probability of each action given an utterance, while the generative model estimates the probability of generating an utterance based on an action.

The CDSSM embeddings have been used as features for training task-specific models (Belinkov et al., 2015; Gao et al., 2014; Chen et al., 2014; Chen and Rudnicky, 2014;

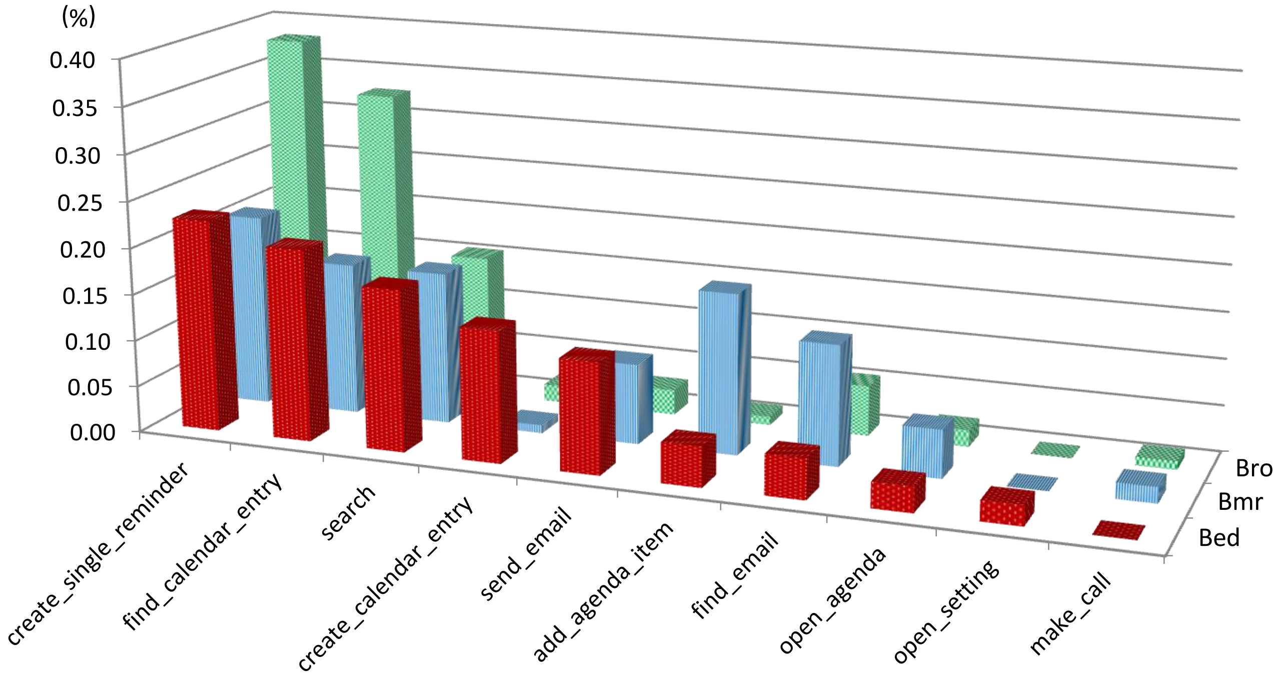


Figure 3: Action distribution for different types of meetings.

Chen et al., 2015b). Here we train an SVM with RBF kernel using the provided embeddings, and show the actionable item detection performance in Table 5. Comparing with baseline lexical features and other semantic embeddings (Le and Mikolov, 2014), the CDSSM embeddings perform better. The performance can be viewed as a starting point of the actionable item detection task. The CDSSM embeddings can be combined with other features and may improve the current state-of-the-art performance (Chen et al., 2016).

Feature		AUC
Baseline	N-gram ($N = 1, 2, 3$)	52.84
	Paragraph Vector from doc2vec	59.79
Proposed	CDSSM: Predictive	64.33
	CDSSM: Generative	65.58
	CDSSM: Bidirectional	**69.27**

Table 5: Actionable item detection performance on the area of the precision-recall curve (AUC) (%).

7. Conclusions

This paper illustrates a novel task, actionable item detection. We publish a data set containing 22 ICSI meetings along with annotated actions, associated arguments, and vector representations for this goal. The annotation utilizes an adapted semantic intent schema based on the design of conversational agents. The detailed statistical analysis of the data set and the baseline performance of currently provided vector representations are presented and suggests a research direction for exploration.

8. Acknowledgements

Place all acknowledgements (including those concerning research grants and funding) in a separate section at the end of the article.

9. Bibliographical References

Ang, J., Liu, Y., and Shriberg, E. (2005). Automatic dialog act segmentation and classification in multiparty meetings. In *Proceedings of 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1061–1064.

Belinkov, Y., Mohtarami, M., Cyphers, S., and Glass, J. (2015). Vectorslu: A continuous word vector approach to answer selection in community question answering systems. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval*, volume 15.

Brown, B., McGregor, M., and McMillan, D. (2015). Searchable objects: Search in everyday conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 508–517. ACM.

Bui, T. H. and Peters, S. (2010). Decision detection using hierarchical graphical models. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 307–312. Association for Computational Linguistics.

Chen, Y.-N. and Metze, F. (2013). Multi-layer mutually reinforced random walk with hidden parameters for improved multi-party meeting summarization. In *Proceedings of 14th Annual Conference of the International Speech Communication Association*. ISCA.

Chen, Y.-N. and Rudnicky, A. (2014). Dynamically supporting unexplored domains in conversational interactions by enriching semantics with neural word embeddings. In *Proceedings of 2014 IEEE Spoken Language Technology Workshop*, pages 590–595. IEEE.

Chen, Y.-N., Wang, W. Y., and Rudnicky, A. I. (2014). Leveraging frame semantics and distributional semantics for unsupervised semantic slot induction in spoken dialogue systems. In *2014 IEEE Spoken Language Technology Workshop*, pages 584–589. IEEE.

Chen, Y.-N., Hakkani-Tur, D., and He, X. (2015a). Detecting actionable items in meetings by convolutional deep structured semantic models. In *Proceedings of 2015 IEEE Workshop on Automatic Speech Recognition and Understanding*. IEEE.

Chen, Y.-N., Wang, W. Y., and Rudnicky, A. I. (2015b). Learning semantic hierarchy with distributional representations for unsupervised spoken language understanding. In *Proceedings of The 16th Annual Conference of the International Speech Communication Association*, pages 1869–1873. ISCA.

Chen, Y.-N., Hakkani-Tür, D., and He, X. (2016). Zero-shot learning of intent embeddings for expansion by convolutional deep structured semantic models. In *Proceedings of 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Cohen, J. et al. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Fernández, R., Frampton, M., Ehlen, P., Purver, M., and Peters, S. (2008). Modelling and detecting decisions in multi-party dialogue. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 156–163. Association for Computational Linguistics.

Galley, M., McKeown, K., Hirschberg, J., and Shriberg, E. (2004). Identifying agreement and disagreement in conversational speech: Use of bayesian networks to model pragmatic dependencies. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 669. Association for Computational Linguistics.

Gao, J., Pantel, P., Gamon, M., He, X., Deng, L., and Shen, Y. (2014). Modeling interestingness with deep neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*.

Hillard, D., Ostendorf, M., and Shriberg, E. (2003). Detection of agreement vs. disagreement in meetings: Training with unlabeled data. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003–short papers-Volume 2*, pages 34–36. Association for Computational Linguistics.

Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., and Heck, L. (2013). Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 2333–2338. ACM.

Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., and Wooters, C. (2003). The ICSI meeting corpus. In *Proceedings of 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE.

Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning*, pages 1188–1196.

McMillan, D., Loriette, A., and Brown, B. (2015). Repurposing conversation: Experiments with the continuous speech stream. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*, pages 3953–3962. ACM.

Riedhammer, K., Favre, B., and Hakkani-Tür, D. (2010). Long story short–global unsupervised models for keyphrase based meeting summarization. *Speech Communication*, 52(10):801–815.

Shen, Y., He, X., Gao, J., Deng, L., and Mesnil, G. (2014). Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 373–374. International World Wide Web Conferences Steering Committee.

Xie, S., Hakkani-Tür, D., Favre, B., and Liu, Y. (2009). Integrating prosodic features in extractive meeting summarization. In *IEEE Workshop on Automatic Speech Recognition & Understanding*, pages 387–391. IEEE.

Yang, F., Tur, G., and Shriberg, E. (2008). Exploiting dialogue act tagging and prosodic information for action item identification. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4941–4944. IEEE.

10. Language Resource References

Adam Janin, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and Chuck Wooters. (2004a). *ICSI Meeting Speech*. Linguistic Data Consortium, LDC2004S02, ISLRN 723-437-529-684-1.

Adam Janin, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and Chuck Wooters. (2004b). *ICSI Meeting Transcripts*. Linguistic Data Consortium, LDC2004T04, ISLRN 295-380-961-299-0.