Role Recognition for Multi-part Dialogue: A Combined Global and Local Approach

Wencan Luo
Department of Computer Science
University of Pittsburgh
PA 15260, USA
wencan@cs.pitt.edu

Abstract

1 Introduction

"People do not interact with one another as anonymous beings. They come together in the context of specific environments and with specific purposes" (Tischler, 1990). As an example, people involved in multi-part dialogue usually play certain roles. For example, Radio Broadcasts

Speaker role is an important cue to the structure dialogue. It can be benefit to role-based summerization (Vinciarelli, 2006), semantical coherent segmentation, information retrieval (Weng at el., 2007; Knapp and Hall, 1972), etc.

Role recognition is the task of automatically recognizing roles of participants in an interaction recording. The goal is to assign to every participant in the recording of an interaction (usually and audio recording or video recording) a role (Salamin, 2013).

In this paper, we will propose a new method for role recognition, which combines both the global and local constraints. There are two intuitions behind: firstly, during a converstion, the role of a participant doesn't change; secondly, the defined roles should be taken evenly among the participants. Take a two-person interview for example. Firstly, a interviewer is always interviewing during the conversation; Secondly, if one is the interviewer and the other must be the interviewee.

2 Related Work

Barzilay et al. (2000) exploited the lexical information (from ASR transcriptions) to identify 3 type of roles: Anchor, Journalist, Guest speakers in news broadcast.

Garg et al. (2008) identified four predefined roles for multi-part meetings. It combined lexical features and social network (SNA) based on linear model. They also extracted features from the social network (Salamin et al., 2009). Later, they proposed a graph model based on purely nonverbal vocal behavioral cues, including who talks when and how much (turn-taking behavior), and statistical properties of pitch, formants, energy and speaking rate (prosodic behavior)(Salamin et al., 2010).

Dynamic Bayesian Networks (Yaman et al., 2010) is also used in role recognition.

3 The Corpus

The corpus I will use is the AMI corpus (McCowan et al., 2005), as same as one used in (Garg et al., 2008; Salamin et al., 2009; Salamin et al., 2010).

The AMI corpus a collection of 138 meeting recordings for a total of 45 hours and 38 minutes of material in a simulated environment. In each meeting, four partici- pants play the following roles: the Project Man- ager (PM), the Marketing Expert (ME), the User Interface Expert (UI), and the Industrial Designer (ID). Each par- ticipant plays a different role, and all roles are represented in each meeting. The same person can play different roles in different meetings, and the ratio of meeting time that each role accounts for, on average, is reported in Ta-

ble 1.

Currently, the state-of-art accuracy is 67.9% on the AMI meeting corpus (Garg et al., 2008; Salamin, 2013) by combining lexical information and social network analysis.

Role	PM	ME	UI	ID
Ratio	36.6%	22.1%	19.8%	21.5%

Table 1: Role distribution.

4 Methodology

5 Timeline

Jan 12 - Jan 20

- understanding the data, know how to extract and use the data
- survey the related work regarding uncertainty prediction based on text

Jan 21 - Feb 8

- coding the rule-based methods
- extract the unigram, bigram and trigram features
- prelim results based this simple lexical features using Weka
- present the prelim results

Feb 9 - March 9

- based on the suggestions, improve the model
- develop more lexical features
- extract other features, such as syntax, semantic

March 10 - April 20

- improve the model based on new features
- anaylze the confusion matrix of the model and improve the model based on the results
- write the workshop-like paper

Acknowledgments

Do not number the acknowledgment section.

References

H. Tischler. 1990. *Introduction to Sociology*. Harcourt Brace College Publishers.

A. Vinciarelli 2006. *Sociometry based multiparty audio recordings summarization*. in 18th International Conference on Pattern Recognition, vol. 2. IEEE, 2006, pp. 11541157.

C. Weng, W. Chu, and J. Wu. 2007. *Movie analysis based on roles' social network*. in IEEE International Conference on Multimedia and Expo, pp. 14031406.

M. Knapp and J. Hall. 1972. *Nonverbal Communication in Human Interaction*. Harcourt Brace College Publishers.

R. Barzilay, M. Colins, J. Hirschberg, and S. Whittaker. 2000. *The rules behind roles: Identifying speaker role in radio broadcasts*. Proc. AAAI Conference on Artificial Intelligence & Conference on Innovative Applications of Artificial Intelligence, 679-684. AAAI Press/MIT Press.

N. Garg, S. Favre, H. Salamin, D. Hakkani-Tur, and A. Vinciarelli. 2008. *Role recognition for meeting participants:an approach based on lexical information and social network analysis*. Proceedings ACM International Conference on Multimedia, 693-696.

H. Salamin, S. Favre, and A. Vinciarelli. 2009. *Automatic role recognition in multiparty recordings: Using social affiliation networks for feature extraction*. IEEE Trans. Multimedia, vol. 11, no. 7, pp. 13731380

H. Salamin , A. Vinciarelli , K. Truong and G. Mohammadi. 2010. *Automatic role recognition based on conversational and prosodic behaviour*. Proceedings of the international conference on Multimedia, October 25-29, 2010, Firenze, Italy

H. Salamin and A. Vinciarelli. 2012. *Automatic role recognition in multiparty conversations: An approach based on turn organization, prosody and conditional random fields*. IEEE Trans. Multimedia, vol. 14, no.2, pp. 338345, 2012.

H. Salamin. 2013. *Automatic role recognition*. PhD thesis, University of Glasgow.

S. Yaman, D. Hakkani-Tuŕ, G. Tuŕ. 2010. *Social role discovery from spoken language using Dynamic Bayesian Networks*. Proc. of Interspeech, 2010.

I. McCowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner. 2005. *The ami meeting corpus*. In Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research.