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

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

Abstract

We proposed a new model to do role recognition for multi-part dialogue. It relies on two observations. Firstly, a speaker's role doesn't change during a conversation; secondly, all the defined roles must be assigned to the speakers. In this way, a combined global and local model is proposed.

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. A broadcast can have an anchor, a guest or a journalist; An interview has an interviewee and an interviewer; a debate has an opponent, an antagonist and a judge.

Speaker role is an important cue to the structure of a dialogue. Identification speaker role can be benefit to role-based summarization (Vinciarelli, 2006), semantically coherent segmentation, information retrieval (Weng et al., 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 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 it: firstly, during a conversation, 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, an interviewer is always interviewing during the conversation; secondly, if one is the interviewer, and then the other must be the interviewee.

2 Related Work

Barzilay et al. (2000) exploited the lexical information (from ASR transcriptions) to identify 3 types 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 a 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 the one used in (Garg et al., 2008; Salamin et al., 2009; Salamin et al., 2010).

The AMI corpus is 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 participants play the following roles: the Project Man-ager (PM), the Marketing Expert

(ME), the User Interface Expert (UI), and the Industrial Designer (ID). Each participant 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 Table 1.

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

Table 1: Role distribution in the AMI corpus.

4 Methodology

4.1 Local Model

For each meeting M, let N be the total number of utterances.

$u = u_1, u_2, \ldots, u_N$ are the utterance sequence.

$s = s_1, s_2, \ldots, s_N$ are the speaker sequence.

$r = r_1, r_2, \ldots, r_N$ are the role sequence.

Where, speaker s_i says u_i, who has the role r_i.

The task is to assign the speakers to defined roles. Assume the role set is R and the speaker set is S.

$$r_i \in R, s_i \in S$$

For the local model, we can estimate the probability for a role r_i given the utterance u_i.

$$P(r_i|u_i)$$

A simple local model could be the lexical model used in (Garg et al., 2008).

4.2 Global Model

One of the global models could be Integer Linear Programming (ILP). The objective is to maximize the probability:

$$P(r|u)$$

If we assume utterances are independent with each other, then

$$P(r|u) = \prod_{i=1}^{N} P(r_i|u_i)$$

If we want to maximize the log of this probability, the objective function becomes to:

$$max \sum_{i=1}^{N} logP(r_i|u_i)$$

$P(r_i|u_i)$ is the local model.

Assume there are k different roles, then r_i is and only is one of the k roles. r_i can be formulized as a k-length vector,

$$r_i =< r_{i1}, r_{i2}, \ldots, r_{ik} >$$

where,

$$r_{ij} \in \{0, 1\}$$

$$\sum_{j=1}^{k} r_{ij} = 1$$

If $r_{ij} = 1$, then r_i is the j^{th} role.

The assumption that a speaker's role doesn't change can be formulized as

$$\forall i, j \; r_i = r_j \text{ if } s_i = s_j$$

The assumption that all the roles must be assigned can be formulized as

$$\forall j \sum_{i=1}^{N} r_{ij} \geq 1$$

In this experiment, we assume that the speakers are known. We can relax the assumption in the future.

5 Timeline

Sep 09 - Sep 22

- survey the related work regarding role recognition
- understanding the data, know how to extract and use the data

Sep 23 - Oct 20

- implement the method in (Garg et al., 2008) using the manual transcription, the lexical model will be used as the local model
- do Automatical Speech Recognition (ASR)

- run the local model on ASR results

Oct 21 - Nov 9

- implement ILP global model, using the manual speaker segmentation

Nov 10 - Dec 12

- propose a model without the manual speaker segmentation

- try other global model such as Bayes network, improved social network

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.