Chatting Activity Recognition in Social Occasions Using Factorial Conditional Random Fields with Iterative Classification

Chia-chun Lian and **Jane Yung-jen Hsu**
Department of Computer Science and Information Engineering
National Taiwan University
yjhsu@csie.ntu.edu.tw

Motivation

Recognizing activities in social occasions plays an important role of building human social networks. For example, the recognition of social interactions could be of great help to determine whether any two attendees have the same interests in an academic conference or a cocktail party. Among the various types of social interactions, chatting with others is a significant indicator. Furthermore, the duration of a chatting activity may imply the strength of the interaction in reality. It is therefore important to recognize the patterns of chatting activities in social occasions. During a real-world conversation, a person often begins talking following the other person's utterance is completed. Linguistic experts have observed that chatting interaction is usually performed as an interlaced dialogic process. As a result, it is intuitive to apply dynamic probabilistic models to learning and detecting chatting activities.

Challenge

The main challenge of chatting activity recognition in social occasions is the existence of multiple people involved in multiple activities. That is, several conversations may take place concurrently, such that different combinations of multi-activity states will impact the final observations, causing a lot of confusion for the recognition of multiple chatting activities. To the best of our knowledge, most existing Bayesian Network models are not powerful enough to accommodate complex relationships among multiple people and activities.

On the other hand, complex probabilistic models, such as Conditional Random Fields (CRFs) (Lafferty & McCallum 2001) and various extensions, do support modeling of concurrent activities. Nevertheless, such models suffer from the problem that exact inference can be intractable when the graph structure has loops. To address this problem, Loopy Belief Propagation (LBP) has been proposed as an alternative algorithm to perform *approximate inference*. Unfortunately, LBP cannot guarantee absolute convergence and the performance is noticeably slow, especially when link density increases in the graph structure.

Related Work

Existing work has shown that sound can reveal many useful dynamic patterns to be learned for building daily-life activity models. Previous researchers used MFCC sound features for bathroom activity classification (Chen *et al.* 2005). Researchers of Human Dynamic Group at the MIT media lab have developed the *smart badges*, a wearable digital sensor that collects behavioral data to measure user interest and affiliation for conference attendees (Gips & Pentland 2006). However, they only used personal audio and accelerometer measurement as primary observations, rather than taking social interaction into account.

In the past, many researchers took efforts to construct individual conversation models by combing subject's style of interactions with their conversational partners (Choudhury & Basu 2004; Wyatt *et al.* 2007). However, as mentioned above, since multiple social groups usually exist concurrently in a real-world public occasions, different conversations is very likely to interfere with each other, adding complexity to the recognition of multiple chatting behaviors. As a result, modeling the interactions between multiple conversations of different social groups is inevitable.

Another work has addressed the above problem of detecting interaction group configurations based on the assumption of turn-taking behaviors, which are usually synchronized inside each conversational group (Brdiczka, Maisonnasse, & Reignier 2005). In their research, they used Hidden Markov Models (HMMs) to describe the transition possibility of dynamic changes among group interactions. However, they treated each possible combination of group configuration as a hidden state in HMMs, causing the model complexity of this approach grows exponentially with the number of group configurations to be recognized.

Proposed Solution

Some researchers have presented another Factorial CRFs (FCRFs), a dynamic-form generalization of basic CRFs in which each time slice contains a set of hidden variables and edges to represent the states of multiple labels and the cotemporal relationships among them (Sutton, McCallum, & Rohanimanesh 2007). So I advocate using FCRF model to conduct inference and learning from patterns of multiple concurrent chatting activities.

To maintain the advantage of co-temporal relationships in FCRF model and meanwhile avoid the use of inefficient LBP algorithm, I propose using Iterative Classification Algorithm (ICA) (Neville & Jensen 2000) as my inference method. To utilize ICA, our basic idea is to separate the original FCRFs model into several Linear-chain CRFs (LCRFs) to learn their corresponding chatting activities given other activities as observations. Because LCRFs provide a well-defined dynamic programming method to help accelerate learning and inferring process, we can combine this characteristic with ICA to do multi-label classification without discarding co-temporal relationships.

Model Design

In my FCRFs model design, each conversation between arbitrary two attendees will be modeled as a single LCRFs model and the hidden variables are represented as binary-state nodes to decide whether a chatting behavior is occurring or not. As shown in Figure 1, Y_i^t denotes the binary states of one possible chatting activity i at time t and X stands for the observation sequence of audio streams. To take co-temporal relationships into account, I let all of the variables Y_i^t at same time slice t be fully connected, representing the possibility of interactions among conversations.

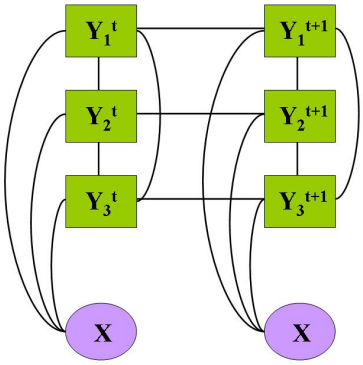

Figure 1: FCRFs example of 3 concurrent chatting activities.

Experiments and Conclusion

In my current experiment, I've collected 3-hours audio streams recorded by 4 participants who would chat with others at any time in a laboratory. I used this data set to train two FCRFs models by using LBP and ICA inference methods respectively. In addition, another LCRFs model is trained for comparison, which discards the connections between hidden variables. Meanwhile, I want to test if different chatting activities can share the same conversation style to be learned. To evaluate the performance, I used 10-fold cross-validation to test the experimental data and the results are summarized in Table 1, which compares the recognition accuracy, learning time, and inferring time.

As we can see, both FCRFs models outperform the LCRFs model, which provides us the conclusion that it is helpful to utilize the co-temporal relationship for chatting activity recognition. In addition, it is allowable for different conversationalists to share the same chatting activity model, which means that this characteristic lets learned model has the chance to be applied to arbitrary environments. Finally, FCRFs model using ICA inference approach takes much less time to do learning process than LBP method, while the inferring time is insignificantly higher.

Comparisons	LCRFs	FCRFs(L)	FCRFs(I)
Separate Model(%)	83.2	85.9	86.0
Sharing Model(%)	79.5	86.2	85.7
Learning Time(sec)	791.5	9332.3	1643.1
Inferring Time(sec)	0.2	1.5	5.4

Table 1: Performance comparison of LCRFs and FCRFs, where (L) and (I) denote LBP and ICA respectively

Acknowledgments

This research was supported by the National Science Council of Taiwan (#96-2218-E-002-008) and the NTU Excellent Research Projects (#95R0062-AE00-05).

References

Brdiczka, O.; Maisonnasse, J.; and Reignier, P. 2005. Automatic detection of interaction groups. In *Proceedings of the 7th International Conference on Multimodal Interfaces*.

Chen, J.; Kam, A. H.; Zhang, J.; Liu, N.; and Shue, L. 2005. Bathroom activity monitoring based on sound. In *Proceedings of the Third International Conference on Pervasive Computing*.

Choudhury, T., and Basu, S. 2004. Modeling conversational dynamics as a mixed-memory markov process. In *Proceedings of the Advances in Neural Information Processing Systems 17*.

Gips, J., and Pentland, A. 2006. Mapping human networks. In *Proceedings of the 4th Annual IEEE International Conference on Pervasive Computing and Communications*.

Lafferty, J., and McCallum, A. Pereira, F. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*.

Neville, J., and Jensen, D. 2000. Iterative classification in relational data. In *Proceedings of the AAAI 2000 Workshop Learning Statistical Models from Relational Data*.

Sutton, C.; McCallum, A.; and Rohanimanesh, K. 2007. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research* 8.

Wyatt, D.; Choudhury, T.; Bilmes, J.; and Kautz, H. 2007. A privacy-sensitive approach to modeling multi-person conversations. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*.