In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA: Penn State.

Modeling Human Intention in a Live-Feeling Platform

Martin Lukac(martin.lukac@nu.edu.kz)
Department of Computer Science, Nazarbayev University,
Astana, 010000, Kazakhstan

Gaziza Oteniyaz (goteniyaz@nu.edu.kz)
Department of Computer Science, Nazarbayev University,
Astana, 010000, Kazakhstan

Michitaka Kameyama (michikameyama@isenshu-u.ac.jp)
Graduate School of Information Sciences,
Tohoku University, Sendai, Japan

Abstract

This paper presents a Bayesian Network (BN) model of human intention in an entertainment environment. We present a framework for live-feeling communication that allows an intelligent system to reconfigure its action based on user preferences and intention. The user intentions are learned from a training set of model situations and a hierarchical model of soccer game is integrated with the user intention into a single BN. The trained model is tested on a data set gathered from users. The BN model shows an accuracy of up to 85%.

Keywords: user intention estimation, live-feeling communication, live entertainment.

Introduction

Intention estimation is a crucial part of intelligent systems intended to be used in daily life or solving problems along with humans (Omori et al. 2007, Yokoyama et al. 2008, Kuan et al. 2010). The reason is that in order to make human-machine cooperation seamless and not requiring a constant direct input from human, the machine must be able to predict user's intention within the context of the task.

Watching a live game on site at a stadium allows the spectators to fully enjoy the entertainment feature since spectators can always view and follow what is happening by turning the head to the right direction. Moreover spectators can decide to observe not only the game but particular elements of the game, other spectators or the environment.

In order to allow spectator to fully enjoy a live entertainment event we propose a live-feeling platform that based on user preferences and the real time content of the event will provide the most desirable view to the user. The platform takes as input real time situation, user preferences and outputs set of commands to one of the available camera. The result of the camera commands are evaluated and compared to expected content that the spectator desires to see.

2. Live-Feeling Communication Platform

The base for our research is shown in the conceptual depiction of the live-feeling communication platform shown in Figure 1. The idea is as follows. The user, being at a remote location is observed by the computer and a real time recognition of emotion, as well as user preferences and game statistics are recorded in real time. The Figure 1 illustrates the camera and microphones to collect data of user intention. On the other end a set of mobile camera controlled by a computer are showing a real-tie event such a soccer or baseball.

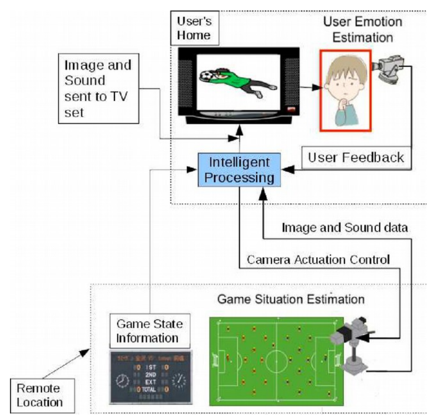

Figure 1: The live-feeling communication platform.

The target of the proposed setup is to show to the user the desired content based on his/her preferences in order to maximize his/her happiness and improve the experience of the event.

3. Data Collection

We extracted the following situations after considering them as the most important moments in a football game: offside, penalty kick, corner kick, goal kick, throw-in, free kick, moment of fouls and misconducts. In order to collect data that is equivalent to sensor processed results, we interviewed people, who are interested in football, for each of these situations. 10 people participated in interviews.

Output discretization

In order to allow predicting the camera control, we divided the football field into the grids and numerated them.

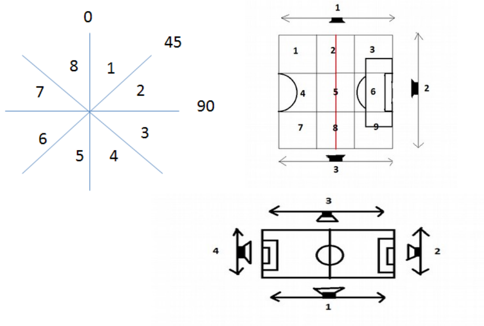

Fig.2 Camera's angles, positions and one half of the football pitch divided into grids

From figure 2 it can seen that in order to show position of some user preference, model specify number of cameras, grids where each should be located and angle. Zoom is specified by 3 levels: short, normal and wide. For example if network gives an output for offside situation "camera #3, grid #8, angle #1, normal zoom" then it means that during offside user may want to see position of goalkeeper since network showing the grid #6.

There are overall 5 cameras. We assume that camera #1, #2, #3, #4 can move, but camera #5 does not, because it is on the top.

4. Model of Human Intention

The Bayesian model of human intention consist of a set of input and output nodes. Inputs of the BN are user preferences and game event situations; outputs are number of camera, camera position, angle and zoom. Inputs node represent categories that were extracted from data analysis.

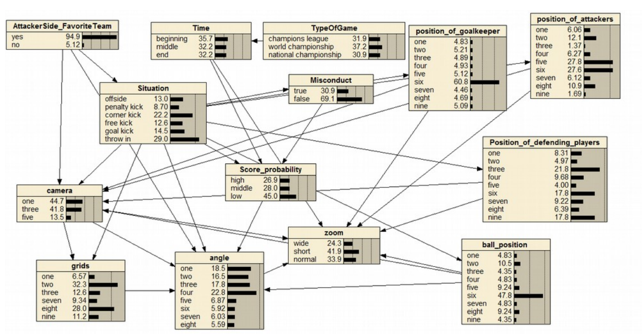

Fig.3 Network for Intention of User Estimation

5. Experiments and Results Discussion

Our goal is to have a human intention model that accurately recognizes human intention and respectively give an output as camera position, its angle and zoom that appropriately capture user preferences.

We conducted interviews to test model accuracy with user cognition. 14 people participated in interviews used for the testing.

In order to calculate accuracy of the network, we compared possible position of user preferences, which we obtained from testing, to position that is indicated by network. Network shows certain field position that we obtained by camera position, its angle and zoom. These places we specified as numerated grids.

Table 2: Network accuracy

Situation	User preferences	Expected locations	Network accuracy
Offside	**goalkeeper:20%**	**Grid #6**	20%
	defending players:20%	**Grid #6**	25%
	referee decision:20%	**Grid #7,8,9**	10%
	offside position in the pitch:40%	**Grid # 6, 5**	25%
			Overall:85%

Table 2 illustrates results of network accuracy in the offside situation. After testing network we obtained what user wants to see during offside situation. User preferences of people are goalkeeper, defending players, referee decision and position where offside happened. In the Table 2 it can be seen how many people prefer to see each of this user preferences out of 14 people in percentage. There are also expected locations of each user preferences. The last column illustrates percentage of accuracy out of 25% for each preference. What network shows during offside position is shown in the figure 4. As it can be seen the red square is the position shown by network with wide zoom. After comparison of expected positions with position that is indicated by red square on the figure 4, we identified that network can show goalkeeper, defending players and offside position in the pitch with high accuracy. However network cannot identify position of referee and its results, since referee can be located outside of the pitch near 7, 8, 9 grids. Despite this fact, there is 10% of accuracy for showing referee, since camera #1 can view 9th grid but with little accuracy since zoom is wide and it can clearly show only 6th grid.

6. Conclusion

In this paper we demonstrated a study on intention estimation during a live event in a live-feeling communication platform. The proposed model works well for the statistically gathered information however as of yet lacks the real-time ability to adapt for individual user's emotional expressions.

In the future a more complete model estimating only user intention (what is the desired content instead of camera control) is to be developed. Moreover building a model for individual user intention estimation is also intended as future work.

References

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 510-516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill.

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287-317.

Kuan, J. Y., Huang, T. H., & Huang, H. P. (2010, August). Human intention estimation method for a new compliant rehabilitation and assistive robot. In *SICE Annual Conference 2010, Proceedings of* (pp. 2348-2353). IEEE.

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

Oliver, N. M., Rosario, B., & Pentland, A. P. (2000). *A bayesian computer vision system for modeling human interactions. Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 22(8), 831-843.

Omori, T., Yokoyama, A., Okada, H., Ishikawa, S., & Nagata, Y. (2007, November). Computational modeling of human-robot interaction based on active intention estimation. In *Neural Information Processing* (pp. 185-192). Springer Berlin Heidelberg.

Lewis, C. (1978). *Production system models of practice effects*. Doctoral dissertation, Department of Psychology, University of Michigan, Ann Arbor.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Sun, X., Jin, G., Huang, M., & Xu, G. (2003, September). Bayesian-network-based soccer video event detection and retrieval. In Third International Symposium on Multispectral Image Processing and Pattern Recognition (pp. 71-76). International Society for Optics and Photonics.

Shrager, J., & Langley, P. (Eds.) (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

Yokoyama, A., Omori, T., Ishikawa, S., & Okada, H. (2008). Modeling of action decision process based on intention estimation. In *SCIS & ISIS* (Vol. 2008, No. 0, pp. 328-333). 日本知能情報ファジィ学会.