GraspSnooker: Automatic Chinese Commentary Generation for Snooker Videos

Zhaoyue Sun[1] , **Jiaze Chen**[2] , **Hao Zhou**[2] , **Deyu Zhou**[1*] , **Lei Li**[2] and **Mingmin Jiang**[1]

[1]School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, China
[2]ByteDance AI Lab, Beijing, China
{sunzhaoyue, d.zhou, mingm.jiang}@seu.edu.cn, {chenjiaze, zhouhao.nlp, lileilab}@bytedance.com

Abstract

We demonstrate a web-based software system, GraspSnooker, which is able to automatically generate Chinese text commentaries for snooker game videos. It consists of a video analyzer, a strategy predictor and a commentary generator. As far as we know, it is the first attempt on snooker commentary generation, which might be helpful for snooker learners to understand the game.

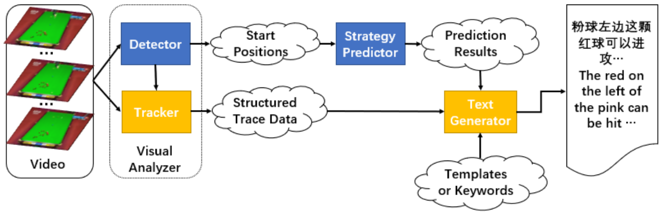

Figure 1: System Architecture of GraspSnooker.

1 Introduction

Snooker is an attractive sport which needs not only physical skills but also a lot of strategies. Novices of snooker may have difficulty in watching raw game videos without commentaries. Therefore automatically generating commentaries for snooker game videos can help them learn from professional players and also add fun to game watching.

Recently, several game commentary generation works have been developed. Yu et al. [2018] proposed a fine-grained NBA video narrative framework, and Jhamtani et al. [2018] presented a move-by-move commentary generation method for chess games. However, there is still no such work on snooker games. Moreover, deep analysis of sport games such as strategies was ignored in most previous approaches. Although some robotic billiards players have been introduced [Smith, 2006; Archibald and Shoham, 2009], their systems were designed under physical environment or simulation environment, not applicable for analyzing strategies from video and generating commentaries.

In this paper, we present a pipeline based system, GraspSnooker, to generate Chinese commentaries for snooker game videos. As far as we know, it is the first snooker video commentary system. For simplification, we generate commentaries for each single shot video which starts with a stationary table without moving balls and ends with a stationary table after one stroking. We collect 10.8k single shot videos and annotate 15.7k corresponding commentaries for training.

The system consists of three components: a) a video analyzer, b) a strategy predictor, and c) a text generator. The video analyzer processes videos, the strategy predictor predicts the target ball and pocket, and the text generator produces commentaries. As snooker commentaries are flexible, we classify them into three categories: score commentary, prediction commentary and track commentary. Specifically, score commentaries describe the score status of players, prediction commentaries show the narrators' guesses about the next shot, and track commentaries explain what happens during this shot. Some examples are shown in Table 1.

2 Demonstration Outline

Users can upload a single shot video to GraspSnooker, and the system will process the video and return the commentaries. On the upload page, users need to provide the information of whether it is required to hit a red ball or a coloured one in this shot. The result page shows the uploaded video with detected table and balls and its perspective mapping to a standard ratio table. The predicted target ball and pocket are marked by twinkling arrows. Figure 2 shows the screen shot of the process. Commentaries generated by two different methods are shown below the videos. English commentaries generated from templates are also provided to assist in understanding. Table 1 gives some commentary examples. GraspSnooker also provides a few preprocessed video clips which can be used to generate commentaries.

3 System Architecture

The architecture of GraspSnooker is presented in Figure 1. It mainly consists of three modules: 1) a video analyzer, which detects the table and balls from video screens, and tracks balls to extract important movement information; 2) a strategy predictor, which predicts the target ball and the target pocket; 3) a text generator, which generates commentaries based on the results generated from the video analyzer and the strategy

[*]Contact Author

	Template-based Generation	Keyword-based Generation
Score Commentary	双方是打到了第9局，目前比分是3比5，威廉姆斯领先。 The two players get into the 9th frame. The current score is 3 to 5. Williams leads.	双方是第9局，现在比分是3比5，威廉姆斯领先一局。 It is the 9th frame of the two players. The score is 3 to 5 now. Williams leads one frame.
Prediction Commentary	打右边库这颗红球。 Hit the red near the right cushion.	可以去打一下右边靠近库的这颗红球。 Maybe he can hit the red near the cushion on the right.
Track Commentary	白球有一些贴边库。 The cue ball is kind of near the cushion.	把白球贴到边库。 He hit the cue ball to the cushion.

Table 1: Comparison of texts generated by template-based and keyword-based methods. Blue marks differences. Red marks mistakes.

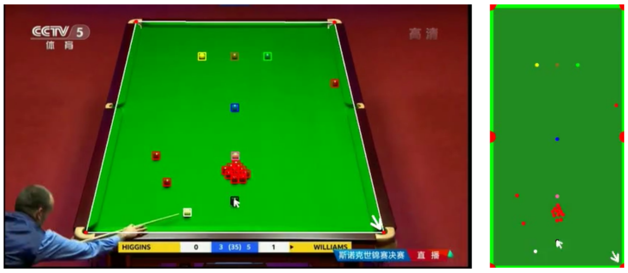

Figure 2: Illustration of video processing and prediction results.

predictor. The details of these three components are described in the following three subsections.

3.1 Video Analyzer

For better video understanding, it is crucial to detect the table and balls in the video (locating) and capture the traces of the moving balls (tracking).

For locating, front view frames showing the full table at a fixed angle are first extracted. Then the table is separated by green borders. And balls are detected by template matching. Once the table and balls are located, locations are mapped to a standard ratio table by perspective transformation. An example of detecting and mapping can be seen in Figure 2.

For tracking, moving balls are detected using background subtracting. And the positions of each ball are reduced to a simple polyline. For convenience, spins are out of consideration. For extracted traces, some crucial information such as which balls are bumped and which cushions are hit, are extracted. These key features will be used to generate commentaries described in Section 3.3. We implement the video analyzer using the python package of OpenCV[1].

3.2 Strategy Predictor

The role of the strategy predictor is to predict which ball to hit and which pocket to target. An adaptation of graph neural network [Li *et al.*, 2015] is employed to learn from data.

Specifically, each ball expect the cue ball on the start table is considered as a vertex in a graph. The initial feature of a vertex is a 28-dim vector, including the color, position and some geometric features inspired by [Smith, 2006]. And for a given ball, its nearest N balls are defined as its neighbors. The edge feature between a ball and one of its neighbors includes two values: the distance between the ball and a neighbor, and the cosine value of target ball-cue ball-object ball angle. Then a similar propagation layer as [Li *et al.*, 2015] can be used for balls' feature propagation. Next, for the output layer, ball and pocket predictions are learned simultaneously. And the output of the ball prediction is used as a soft attention that decides which balls are relevant to the pocket prediction.

The prediction model is trained on 2k training instances and 200 validation instances. Tested on 200 instances, the accuracy of target ball prediction is 62.25%, of target pocket prediction is 77.04%, and of both is 56.12%. The prediction model is implemented based on PyTorch Geometric[2] library .

3.3 Text Generator

In our demonstration, two solutions are provided to generate commentaries: template-based generation and keyword-based generation. The template-based method generates sentences from human-written templates, while the keyword-based method extends given keywords of templates to sentences by CGMH (*Constrained sentence Generation by Metropolis-Hastings sampling*) [Miao *et al.*, 2018], aiming to increase the language diversity. CGMH is an unsupervised method applying Metropolis-Hastings sampling [Metropolis *et al.*, 1953] to generate sentences under constraints. Readers may refer to [Miao *et al.*, 2018] for details of CGMH.

Examples of some commentaries generated by the two methods are illustrated in Table 1. Compared to template-based generated commentaries, keyword-based generated commentaries are more flexible. Blue texts show their differences. However, as CGMH only generates sentences according to language models, ignoring the context of the events, so that it usually loses some coherence. In addition, sometimes words that are contrary to the video information are introduced, which are marked as red in Table 1.

Two language models of forward and backward direction are trained on all 15.7k collected snooker commentaries, which get perplexities of 35.87 and 36.04 respectively.

4 Conclusion

In this work, we present GraspSnooker, the first application on snooker commentary generation. It combines video understanding and strategy prediction with commentary generation. Commentary generation for long snooker videos will be explored in the future.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (61772132) and the Natural Science Foundation of Jiangsu Province of China (BK20161430).

[1] https://pypi.org/project/opencv-python/

[2] https://rusty1s.github.io/pytorch_geometric/build/html/index.html#

References

[Archibald and Shoham, 2009] Christopher Archibald and Yoav Shoham. Modeling billiards games. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 193–199. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[Jhamtani *et al.*, 2018] Harsh Jhamtani, Varun Gangal, Eduard Hovy, Graham Neubig, and Taylor Berg-Kirkpatrick. Learning to generate move-by-move commentary for chess games from large-scale social forum data. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1661–1671, 2018.

[Li *et al.*, 2015] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493*, 2015.

[Metropolis *et al.*, 1953] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[Miao *et al.*, 2018] Ning Miao, Hao Zhou, Lili Mou, Rui Yan, and Lei Li. CGMH: Constrained sentence generation by metropolis-hastings sampling. *arXiv preprint arXiv:1811.10996*, 2018.

[Smith, 2006] Michael Smith. Running the table: an AI for computer billiards. In *Proceedings of the 21st national conference on Artificial intelligence-Volume 1*, pages 994–999. AAAI Press, 2006.

[Yu *et al.*, 2018] Huanyu Yu, Shuo Cheng, Bingbing Ni, Minsi Wang, Jian Zhang, and Xiaokang Yang. Fine-grained video captioning for sports narrative. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6006–6015, 2018.