

# MCS: An In-battle Commentary System for MOBA Games

Xiaofeng Qi, Chao Li, Zhongping Liang, Jigang Liu, Cheng Zhang,  
Yuanxin Wei, Lin Yuan, Guang Yang, Lanxiao Huang, Min Li

Tencent Timi L1 Studio, Chengdu, China

sealqi, supeli, skylerliang, jigangliu, donyuzhang, winsterwei,  
tayloryuan, mikoyang, jackiehuang, konthigali@tencent.com

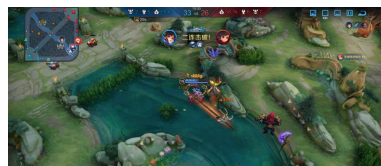
## Abstract

This paper introduces a generative system for in-battle real-time commentary in mobile MOBA games. Event commentary is important for battles in MOBA games, which is applicable to a wide range of scenarios like live streaming, e-sports commentary and combat information analysis. The system takes real-time match statistics and events as input, and an effective transform method is designed to convert match statistics and utterances into consistent encoding space. This paper presents the general framework and implementation details of the proposed system, and provides experimental results on large-scale real-world match data.

## 1 Introduction

In recent years, MOBA (Multiplayer Online Battle Arena) games have been popular all over the world. As a popular mobile MOBA game, Honor of Kings has a DAU of over 100 million, and has become an official sport in the 2022 Hangzhou Asian Games. In MOBA games, several players are divided into two teams and carry out a battle against each other. AI research in MOBA games has received many attention (Ye et al., 2020, 2022). Real-time commentary plays an important role in live streaming of e-sports competitions, assists the audience in the comprehension of game situation and events, expanding the presence of MOBA games.

In-battle commentary aims at generating coherent and informative natural language descriptions based on game situation, match statistics and player behaviour, as illustrated in Figure 1. Similar systems have already been widely applied in real-world applications such as automated news broadcast, television sports commentary, online video description, etc (Chen and Mooney, 2008; Wiseman et al., 2017; Nie et al., 2018; Puduppully et al., 2019; Kale and Rastogi, 2020). In this paper, we focus on the scenario of live e-sports commentary



Hero	Kills	Deaths	Assists	...	Gold	Gold (variation)
Shangxiang Sun	12	4	7	...	14028	16.23/second
Que Bian	3	2	20	...	13428	19.53/second
...	...	...	...	...	...	...
Ying Yun	8	6	8	...	15186	3.73/second
Yuanfang Li	9	7	8	...	12186	3.73/second
...	...	...	...	...	...	...

Killer hero: Shangxiang Sun

Be killed hero: Yuanfang Li

Event: double kill

Commentary:

孙尚香拿下双杀，厉害！

Shangxiang Sun accomplished

double kill, awesome!

Figure 1: The in-battle commentary generation task.

in mobile MOBA games and introduce MCS: an in-battle commentary generation system<sup>12</sup>.

Automatic commentary is grounded in in-battle match statistics and events, regarded as a data-to-text generation task. Recently, many large-scale language models have been proposed and promoted the advance in natural language generation (Radford et al., 2018; Devlin et al., 2019; Radford et al., 2019). These pre-trained language models are also beneficial for data-to-text tasks (Kale and Rastogi, 2020), in this work, we employ OpenAI’s GPT-2 model as the basic generator, and build a neural pipelined approach for commentary generation.

Although recent works have shown promising results regarding live comment generation (Chen and Mooney, 2008; Wiseman et al., 2017; Novikova et al., 2017; Dušek et al., 2018; Ma et al., 2019; Zhang et al., 2020), there are still many challenges for constructing a commentary generation model for MOBA games: (1) Previous studies incorporated static data like final scores, while live e-sports commentary uses changing match statistics as the game goes on. Current statistics as well as history statistics are both vital for text generation. For example, the narrowing of advantages may cause anxiety, while overturn from deficit can be

<sup>1</sup>This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <https://creativecommons.org/licenses/by/4.0/>.

<sup>2</sup>Demonstration video: <https://youtu.be/G0IKZKd7eco>.

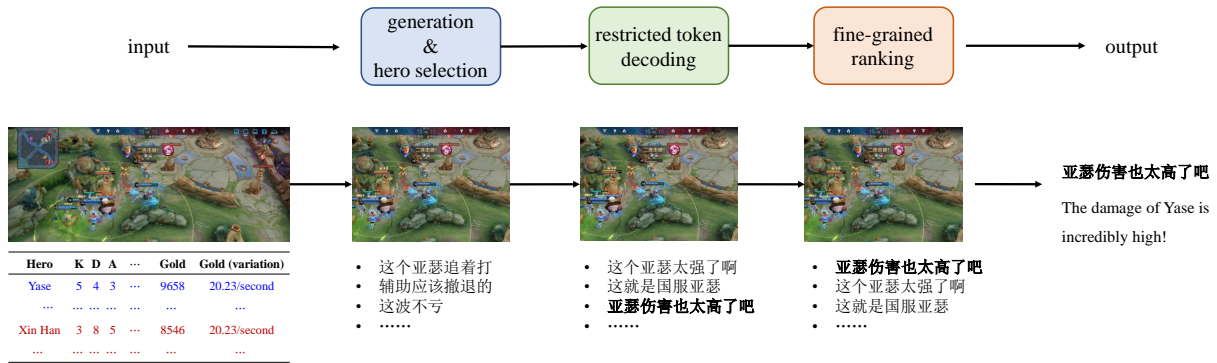


Figure 2: Pipeline of the proposed MCS, including three stages: initial commentary generation, restricted token decoding and fine-grained ranking.

a morale booster. Therefore, the problem of commentary generation employing dynamic structured data needs to be tackled. (2) Neural network based text generators are claimed to be unreliable in the consistency and accuracy of generated texts. However, commentary generation is highly correlated with in-battle data, which requires high conformity in the generated sentence and input data. In many approaches, retrieval methods are used instead of generative methods. How to solve the problem of generative methods and improve reliability remains challenging.

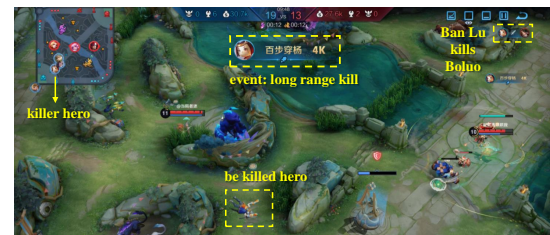
The proposed MCS has the following advantages: (1) We apply GPT-2 model into data-to-text task, and design a hero selection loss to focus on the target hero behaviour for commentary. (2) Previous studies on sports commentary generation incorporated only static data like final scores, while the proposed MCS uses dynamic match statistics during generation, as current and history statistics are both vital for e-sports commentary. (3) Commentary is highly correlated with in-battle data, which requires high conformity in the generated sentence and input data. A restricted token decoding method is employed to improve generation reliability.

## 2 System Framework

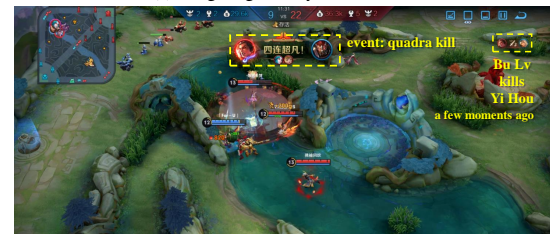
We first introduce the general framework of the proposed MCS system. The system consists of three modules in a pipeline, as presented in Figure 2. Taking in-battle statistics as input, the system produces a descriptive sentence based on current and history data.

### 2.1 Initial commentary generation

The first module focuses on generating initial texts: in this stage, several candidate sentences are gen-



(a) Long range kill by hero Ban Lu.



(b) Bu Lv quadra kill.

Figure 3: Examples of events. (a) A critical event of long range kill, active hero: Ban Lu, passive hero: Boluo; (b) A critical event of quadra kill, active hero: Bu Lv, passive hero: Wei Dian.

erated in an end-to-end manner, used as inputs for further polishing. All statistics of the ten heros as well as critical events are included in network input. The critical event is defined by a triplet including event type, active hero, and passive hero. For instance, single killing is a typical event that occurs frequently in MOBA games, in which the active hero is the hero kills someone, and the passive hero is the killed one. Figure 3 shows two examples of events. Beside all hero statistics, critical events are also included in network input as most commentaries focus on the heros involved in these events.

### 2.2 Restricted token decoding

The second module performs restricted decoding based on the initial text generator. Restricted generation can increase the reliability of neural lan-



(a) Data panel in game.

Data	Description
Hero ID	Unique identifier of the hero
Kills	Number of kills of the hero
Kills (variation)	The variation of kills in the last 30 seconds
Deaths	Number of deaths of the hero
Deaths (variation)	The variation of deaths in the last 30 seconds
Assists	Number of assists of the hero
Assists (variation)	The variation of assists in the last 30 seconds
Gold	In-game virtual currency income of the hero
Gold (variation)	The variation of Gold in the last 30 seconds

(b) Data and description.

Figure 4: Match statistics in game. Left: the kills/deaths/assists and other attributes are displayed on the data panel. Right: the data used as system input and description.



Figure 5: Serialization of structured data of a hero.

guage generation (Elder et al., 2020; Dathathri et al., 2020). Inspired by retrieval text generation, we construct event-specific corpus from the collected dataset, which is used in the token decoding process. With this approach employed, the generated sentences are expected to be controlled within a reasonable range around the corresponding event.

### 2.3 Fine-grained ranking

The final module conducts fine-grained ranking on sentences generated through the previous two modules. A pairwise scoring model is constructed following textual matching methods, to avoid topic shift and destruction of text fluency, the scoring model is trained against topic irrelevant commentaries and faulty sentences. After ranking, the commentary achieving the highest score serves as final output of the MCS system.

## 3 System Implementation

In this section, the implementation details of the proposed three modules are provided. In general, the network architecture are constructed following the pipeline of language models. As described in Figure 4, features are serialized and fed as the inputs, the network predicts a target hero ID and generates a comment as the outputs.

### 3.1 Initial commentary generation

For the initial commentary generator, the network input is composed of critical events and match statistics of the ten heros. A triplet including event

type and two hero IDs is used to represent the event, as described in Section 2.1. For match statistics, the data entries used in the proposed MCS includes Hero ID, Kills, Deaths, Assists, etc, as presented in Figure 4. Since the statistics are changing over time and the changes are important in indicating hero status and highlight events, we also include variations of the above data entries in network input. For example, a hero with high kills and gold usually means that it can cause high damage in battle. If the kills of a hero rise rapidly in a short time, it is likely that the hero has outstanding performance and human commentators will comment on it.

To adapt the structured data to a sequential input of the network, we design a method to transform hero attributes into decimal digits. For each hero, a unique identifier is assigned and placed at the beginning of its data sequence, followed by a token indicating whether the hero is alive or not. The other indicators are discretized by dividing into segments, Figure 5 illustrates an example of data serialization. Kills, Deaths, Assists and the variations are encoded by 600 digits, 100 for each, as they are usually less than 100 in a single match. In extreme cases where the indicators exceed 100, the real digits are replaced by 100 to avoid confusion. In a single match, the Gold of each hero usually varies from 0 to 40000, it is discretized by dividing into 20 segments, 2000 golds for one segment. The variation of Gold is encoded in a similar way, each segment represents 90 golds per minute, and the maximum variation is set to 4500 golds per minute.



At the end of the sequence, we use two extra tokens to represent the ranking of Gold and KDA of the corresponding hero among all 10 heroes. The data sequence of all 10 heroes are connected by a special token "[SEP]", and another special token "[EOD]" is added after the last data block to separate text from data sequence.

In general, the network architecture is illustrated in Figure 6. Previous studies have proven that pre-trained language models can improve data-to-text generation (Peng et al., 2020; Chen et al., 2020; Kale and Rastogi, 2020), we use Open-AI’s GPT-2 model to build the basic text generator. The GPT-2 model used in this work was pre-trained using Chinese copora LCCC (Large-scale Cleaned Chinese Conversation (Wang et al., 2020)). In addition to standard language model loss, we design a hero selection loss to focus on the match data of the target hero, which is helpful to increase generation consistency.

### 3.2 Restricted token decoding

In this stage, the collected dataset is divided into event-specific groups by a triplet retrieval key. The retrieval key is composed of event type, active hero ID, and passive hero ID. During generation, the triplet retrieval key can be extracted from in-battle statistics, and the corresponding event-specific corpus can be retrieved. In retrieval text generation approaches, a selection strategy is designed to choose a proper sentence from the corpus as the final result. In the proposed MCS system, we employ the event-specific corpus to restrict the sentences generated by neural networks.

The Plug and Play Language Model (PPLM) (Dathathri et al., 2020) module is employed to control the token decoding process, restricting the generated texts around the corresponding event. Basically, the restriction is accomplished by updating the last hidden layer features  $h_t$ :

$$h_t \leftarrow h_t - \nabla \sum_{i=1}^{|corpus|} \log(p(w_i|h_t)) \quad (1)$$

where  $|corpus|$  represents the size of an event-specific corpus and  $w_i$  denotes a word in the corpus. After a few iterations of updating, the hidden features are softly adjusted towards decoding the words in the event-specific corpus.

Table 1: Experimental results, best results are indicated in bold. The upper two rows are results of existing methods, the bottom rows are results of our method.

	ROUGE-1	ROUGE-2	ROUGE-L	Validity
Seq2Seq (Sutskever et al., 2014)	0.1306	0.0032	0.1296	0.69
GPT-2 (Radford et al., 2019)	0.1568	0.0102	0.1491	0.85
Initial commentary generator	0.1921	0.0111	0.1903	-
Restricted token decoding	0.2025	0.0140	0.2001	-
Fine-grained ranking	<b>0.2028</b>	<b>0.0144</b>	<b>0.2006</b>	<b>0.91</b>

### 3.3 Fine-grained ranking

The final stage performs a ranking on the generated candidate sentences. We build a pairwise ranking network on top of the GPT-2 generator. Taking two sentences as input, the GPT-2 outputs two encoding vectors respectively. These encodings are then fed into the ranking network, and a score is derived which indicates the relative order of the two sentences. After all candidate sentences are processed, the best sentence is regarded as the final commentary.

## 4 Materials and Results

The structured data and commentary history of 30 million matches in professional leagues and live broadcasts of Honor of Kings were collected and used to train and evaluate the proposed system. Perplexity is used to evaluate the fluency of generated sentences, ROUGE-1, ROUGE-2 and ROUGE-L are used to measure the distance between generated sentences and ground truth. We also calculate the generation validity by human evaluation.

The experimental results are presented in Table 1. Quantitatively, the system achieves an averaged perplexity of 1.27. Comparing the generated sentences with those written by human commentators among the collected data, the ROUGE-1, ROUGE-2 and ROUGE-L scores are 0.1921, 0.0111 and 0.1903 for the initial commentary generation, respectively. Combined with target hero prediction and restricted token decoding, the metrics improved to 0.2025, 0.0140 and 0.2001. The final ROUGE scores are 0.2028, 0.0144 and 0.2006 employing fine-grained ranking. Besides statistical indicators, we also carried out human evaluation of the system, and the generated commentaries achieve a validity of 0.91. This means that the system has the potential to be used in real-world applications. Our system is compared with two mainstream generation approaches: the RNN-based Seq2Seq (Sutskever et al., 2014) and the transformer-based standard GPT-2 (Radford et al., 2019), and outperforms both of them.

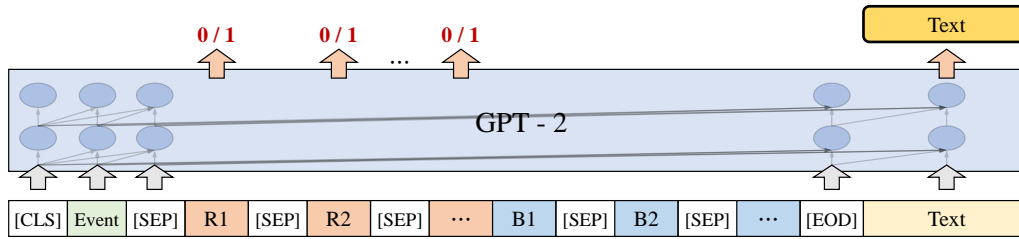
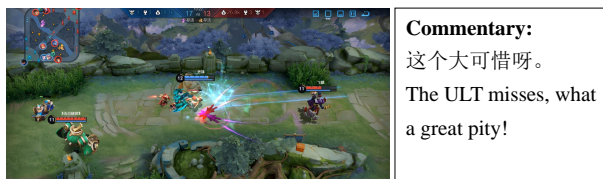
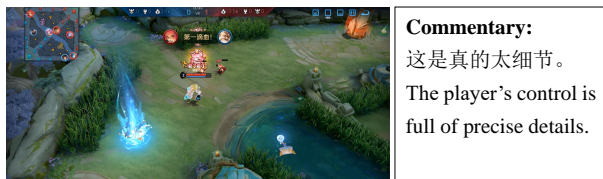


Figure 6: Architecture of the generation network. The network is based on GPT-2 model. The beginning of network outputs accomplish target hero selection, the remaining represents the generated tokens. "Event" denotes the triplet of critical event, "R1" - "R5" represents the data of the 5 heros on the red side, "B1" - "B5" represents the data of the 5 heros on the blue side.

Figure 7 shows two cases of the generated commentaries.



(a) The ULT of Ganjiang Moye misses, a few moments later, he is killed by Mozi.



(b) Anqila draws first blood by killing Master Fu.

Figure 7: The examples of commentaries generated by the proposed system.

## 5 Conclusion

We propose MCS, a model for in-battle real-time commentary generation in MOBA games. MOBA is a type of games popular all over the world, commentary plays an important role in game comprehension for audience. The proposed MCS is designed with the ability of analyzing real-time match statistics and generating coherent descriptive contents. It has a wide range of applications such as e-sports commentary, combat information analysis, etc, and has the potential to promote game development and improve match environment.

## References

- David L. Chen and Raymond J. Mooney. 2008. [Learning to sportscast: A test of grounded language acquisition](#). In *Proceedings of the 25th International Conference on Machine Learning*, page 128–135. Association for Computing Machinery.
- Zhiyu Chen, Harini Eavani, Wenhui Chen, Yinyin Liu, and William Yang Wang. 2020. [Few-shot NLG with pre-trained language model](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 183–190, Online. Association for Computational Linguistics.
- Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2020. [Plug and play language models: A simple approach to controlled text generation](#). In *International Conference on Learning Representations*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, pages 4171–4186.
- Ondřej Dušek, Jekaterina Novikova, and Verena Rieser. 2018. Findings of the e2e nlg challenge. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 322–328.
- Henry Elder, Alexander O'Connor, and Jennifer Foster. 2020. [How to make neural natural language generation as reliable as templates in task-oriented dialogue](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2877–2888, Online. Association for Computational Linguistics.
- Mihir Kale and Abhinav Rastogi. 2020. Text-to-text pre-training for data-to-text tasks. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 97–102. Association for Computational Linguistics.
- Shuming Ma, Lei Cui, Damai Dai, Furu Wei, and Xu Sun. 2019. [Livebot: Generating live video comments based on visual and textual contexts](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6810–6817.
- Feng Nie, Jinpeng Wang, Jin-Ge Yao, Rong Pan, and Chin-Yew Lin. 2018. [Operation-guided neural networks for high fidelity data-to-text generation](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages

- 3879–3889. Association for Computational Linguistics.
- Jekaterina Novikova, Ondřej Dušek, and Verena Rieser. 2017. The e2e dataset: New challenges for end-to-end generation. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 201–206.
- Baolin Peng, Chenguang Zhu, Chunyuan Li, Xiujun Li, Jinchao Li, Michael Zeng, and Jianfeng Gao. 2020. Few-shot natural language generation for task-oriented dialog. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 172–182, Online. Association for Computational Linguistics.
- Ratish Puduppully, Li Dong, and Mirella Lapata. 2019. Data-to-text generation with content selection and planning. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019*, pages 6908–6915. AAAI Press.
- Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training. *OpenAI Blog*.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*.
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.
- Yida Wang, Pei Ke, Yinhe Zheng, Kaili Huang, Yong Jiang, Xiaoyan Zhu, and Minlie Huang. 2020. A large-scale chinese short-text conversation dataset. In *NLPCC*.
- Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263. Association for Computational Linguistics.
- Deheng Ye, Guibin Chen, Wen Zhang, Sheng Chen, Bo Yuan, Bo Liu, Jia Chen, Zhao Liu, Fuhao Qiu, Hongsheng Yu, Yinyuting Yin, Bei Shi, Liang Wang, Tengfei Shi, Qiang Fu, Wei Yang, Lanxiao Huang, and Wei Liu. 2020. Towards playing full moba games with deep reinforcement learning. In *NeurIPS*.
- Deheng Ye, Guibin Chen, Peilin Zhao, Fuhao Qiu, Bo Yuan, Wen Zhang, Sheng Chen, Mingfei Sun, Xiaojian Li, Siqin Li, Jing Liang, Zhenjie Lian, Bei Shi, Liang Wang, Tengfei Shi, Qiang Fu, Wei Yang, and Lanxiao Huang. 2022. Supervised learning achieves human-level performance in moba games: A case study of honor of kings. *IEEE Transactions on Neural Networks and Learning Systems*, 33(3):908–918.
- Shengyu Zhang, Ziqi Tan, Zhou Zhao, Jin Yu, Kun Kuang, Tan Jiang, Jingren Zhou, Hongxia Yang, and Fei Wu. 2020. Comprehensive information integration modeling framework for video titling. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 2744–2754, New York, NY, USA. Association for Computing Machinery.