Journal of Knowledge Information Technology and Systems

ISSN 1975-7700

http://www.kkits.or.kr

A Study on Hadoop-based Self-Organizing Map for Golf Swing Model in Big Data Environment

Wansik An*

Department of Physical Education, SungKyul University

ABSTRACT

The healthcare and physical strength are critical factors to be considered in a highly competitive environment in human life. Many people prefer sport is designed to obtain the key of human life because it is a good characteristic both health and physical strength. One of the Golf Swing Model (GSM) in sport is defined as a designed, computer treated of complexity of motions automatically. Especially, GSM is apparently concerned with speed generation its adaptability such as golfer segment angular kinematics, kinetic energy and angular momentum. For this reason, the design of GSM is need to expertise on knowledge of motion patterns, improve by altering the sequence of rotations in the conventional golf swing. In our research paper, it is to study and evaluate the GSM by simulating modeling for experiment and analysis. The methodology used in our research is simulated by Self-Organizing Maps (SOM). SOM provide the design system as well as offer environment to which experiment of the system can be performing. Eventually, our GSM by using SOM is presented some researchable scenario. In addition, we extend the algorithm to handle attribute datasets containing both numeric and categorical attributes in Big Data Environment.

KEYWORDS: GSM(Golf Swing Model), Self-organizing maps, G-data, Big data, Motion patterns

ARTICLE INFO: Received 15 September 2018, Revised 9 October 2018, Accepted 12 October 2018.

*Corresponding author is with the Department of Physical Education, Sungkyul University, Sungkyul University-ro 53, Manan-gu Aanyang-si, Gyeonggi-do, 14097, KOREA. *E-mail address*: aws119@sungkyul.ac.kr

1. Introduction

There are a lot of application generating sport data daily. Growing numbers of golf player are dealing with more and more study of golf motion during the golf swing. Their research method has increased over the past years, but the aim of the research has remained the same. In a point of biomechanics view in golf, it has been to improve performance and reduce injury risk of golf player through the application of Newton's Laws of Motion to the golf swing. Nevertheless, it is hard to find out about the scientific method. Unlike the application of Newton's Laws of Motion to the golf swing, classification of this motion in a general context is an ordering procedure that respects certain rules. In the context of biology, for example, classification is used to organize a vast number of known large distances towards a target, number of strokes, or ball-club contacts each holes. In this method, classification is seen as a process of exquisiteness rules. Specifically, in this field of artificial intelligence, the task of the classifier is to perform a classification process. One of the most representatives is the neural networks classifier. However, this based on other theories are also used, such as Bayesian, k-nearest neighbor, and Fuzzy c-means. Every classifier has its specific classification methodology [1]. Although some these perform better than others on certain problems, there is no perfect classifier or one that performs best on any given problem. After all, our research paper is to define the performance of unsupervised and supervised Self Organizing Maps (SOM) classifiers [2]. Unsupervised and supervised stands for non-heuristic and heuristic for golf swing. Generally, SOM is an unsupervised (non-heuristic) learning algorithm that maps an input vector to an output array of nodes. This algorithm is used in simulations of classifiers for a specific task [3]. In addition, the aim of our research is to provide a good solution for G-data acquisition with refer to different types of G-data source in different sense. The following section of this paper provides as followings: Section 2 is Related works, Section 3 is Framework architecture on Hadoop, Section 4 is Experiments and Section 5 is Conclusions.

2. Related Works

2.1 Big Data

Intelligent-based and cloud-based system have been wildly used to analyze huge amount of data efficiently, and there are many applications made up it. It uses MapReduce paradigm and is easy to scale. However, for many applications built on it, designing, configuring, and deploying are still done in real world manually. Big data was defined in this situation to provide solutions to their systems. This is an intelligent-based and cloud-based framework that make an

effort to build on automatic system for supporting to data analysis [4-6].

2.2 Self-Organizing Map

Self-Organizing Map (SOM) is one of the most popular network models, and belongs to the category of competitive learning networks [7-9]. The SOM is based on unsupervised learning, which means that no human intervention is needed during the learning and that little needs to be known about the characteristics of the input data [10-12]. And also, this is used to clustering data without knowing the class memberships of the input data, and could be used to detect features inherent to problem with Self-Organizing Feature Map (SOFM). The SOM has been proven useful in many applications of complex machine and process states, otherwise very difficult to perceive and interpret [13]. The SOM can also utilized for the development of new pattern classification and target recognition system, whereby categorization of the input signal states is performed by it [14-15].

3. Framework Architecture on Hadoop

In our proposed system, we will provide two solutions of data acquisition of GSM (G-data). First, the traditional way of data acquisition will be given out. After that, a new way based our system, so called system name of AWS, as shown in <Fig. 1>. According to our analysis, AWS based solutions proves to be better, and G-data acquisition form different data sources will be introduced. Hadoop cluster is used for high performance computing, Hadoop runs MapReduce jobs on the cluster, and stores the results on HDFS after the jobs finish. G-data is fed to MapReduce jobs, and it must exist on HDFS before MapReduce jobs run. There are several steps to follow to run a MapReduce job in Hadoop. The first step is to prepare the required G-data set for the job, and copy it to HDFS. Secondly, submit the job to Hadoop, either by executing a java program invoking Hadoop API, or parsing a pig script by Pig. After the job finishes, the result will be stored in a directory specified by user on HDFS. The last step is to get out the result on HDFS. When Hadoop jobs work on these G-data files, it takes a lot of extra time.

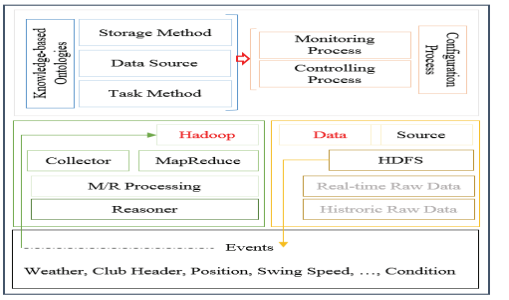

Figure 1. AWS Framework Architecture

HDFS does not support appending file content at the moment, and the only way is to acquire G-data from G-data sources and accumulate the G-data file to a proper size, and then copy it to HDFS. The performance

of G-data acquisition in this way is influenced by copying G-data to HDFS. When data file is relatively small, it takes almost no time to copy. But when there is a large file which is of several gigabytes, the copying time cannot be ignored. Especially in this paper, where the replicate number of Hadoop is set to two and three, the file size is doubled and three times as much as the original size. AWS is an open source G-data collection system designed for monitoring large distributed system. Been built on top of Hadoop, this inherits the scalability and robustness of it. For example in our scenario of AWS, in order to utilize the collected G-data in a better way, this provides a flexible and powerful toolkit to display, monitor, and analyze results. This has three main parts: Players, Collectors, and MapReduce jobs. The workflow of this system is shown in <Fig. 2>.

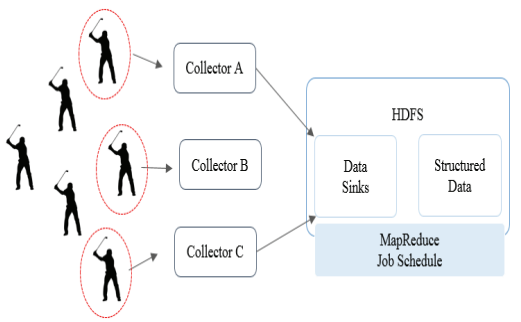

Figure 2. Workflow of AWS System Scenario

Players are responsible for collecting G-data through their adaptors. Adaptors interact with G-data source, and run inside of players that are collecting G-data. As adaptors are quite flexible within player, it is possible to have several adaptors for a player to collect G-data from different source at the same time. Players run on every host of Hadoop cluster, and G-data from different hosts may generate different G-data. But there is no need to use AWS in this situation as all players will collect the same G-data. Collectors gather G-data from different players through sensory or stream G-data, and then write all the data into a Hadoop sequence file called sink file. Sink file consists of records collected by players. After the sink file reaches a certain size or when a certain time is out, the files will be renamed and made available for MapReduce processing. In this scenario, collectors play an important role in reducing the number of HDFS files. Collectors also provide a uniform interface for adaptors. MapReduce job is made up of archiving jobs, and it makes the collected G-data organized and more convenient for analysis. AWSRecords are set of key-value pairs, and they are made available for Hadoop jobs to process by the collected G-data.

4. Experiments

In this paper, for the experiments of G-data, we supposed a very small G-data set with only two parameters, speed and height, which are already structured. Structure is usually performed to control the variance of vector components by SOM.

Table 1. Input G-data Records

His height is shorten from 160 to 169, so called *Small,* and is prospered from 170 to 179, so called Normal. Golf Swing speed is defined 'Fast' from 0.7 to 0.9. (Assuming Max Speed is 1)

	height	*speed*	
1	h_{11} = 160	S_{12} = 0.8	Small and Fast GS
2	h_{21} = 165	S_{22} = 0.7	Small and Fast GS
3	h_{31} = 170	S_{32} = 0.9	Normal and Fast GS
4	h_{41} = 175	S_{42} = 0.7	Normal and Fast GS
...	...	...	...

It is used to prevent the dominance in the map organization of vector components that have a variance which is significantly higher than the variance of other components.

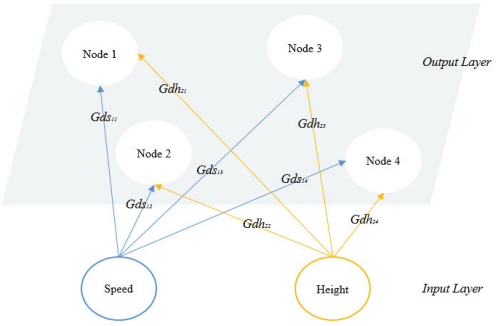

Figure 3. Simple SOM Network for G-data in AWS

As we mentioned in our scenario, we will use a 2 x 2 SOM Network to discover clusters in the G-data set. In the case of a small network the neighborhood radius is set to 0 so that only the 'good' node adjusts its height. The input set of some records is presented in Table 1. For this example the learning rate ή = 0.5. The heights h_{ij} are randomly initialized and presented in <Fig. 3>.

Accordingly, by the simple SOM Networks and Table 1, AWS approach defines the map size depending on the number of map units with nodes chosen by the operator. The number of map units changes from 1 to 18 with GS. An additional experiment is performed by Hadoop for the creation of a map. Then the experiment with different node lengths is repeated with the same values sometimes as shown in <Fig. 4>.

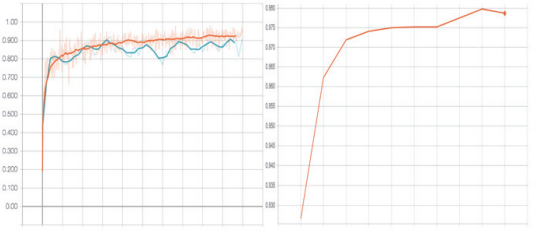

Figure 4. According to AWS-based GSM :
Left side is shown real-time raw data including events,
Right side is shown historical raw data without events

5. Conclusions

In this paper, we focused to on solutions for G-data acquisition with a good performance. There were two different G-data sources used in this research. Historical G-data came from without events such as indoor screen golf with gathering acquisition, but real G-data is difficult to gaining from real world. Two different solutions for G-data acquisition were presented by Hadoop and SOM on the AWS. However, we try to make a better to performances comparison by using example scenario, we don't have to achieve

in the quantitative and the qualitative research by considering of various events. There is a significant amount of development still needed in this research.

References

[1] J. Vesanto, *SOM-based data visualization methods*, Intelligent. Data Analysis, Vol. 3, No. 2, pp. 111-126, 1999.

[2] M. C. Su, and H. T. Chang, *A new model of self-organizing neural networks and its applications*, IEEE Transaction Neural Network, Vol. 12, No. 1, pp. 153-158, Jan. 2001.

[3] J. Vesanto, and E. Ahoniemi, *Clustering of the self-organizing map*, IEEE Transaction Neural Network, Vol. 11, No. 3, pp. 586-600, May 2000.

[4] X. Wu, X. Zhu, G. Q. Wu, *Data mining with Big Data*, IEEE Trans. on Knowledge and Data Engineering, Vol. 26, No. 1, pp. 97-107, Jan. 2014.

[5] Z. Zheng, J. Zhu, and M. R. Lyu. *Service -generated big data and big data-as-a-service: An overview*, in Proceeding IEEE BigData, pp. 403-410, Oct. 2013.

[6] W. Zeng, M. S. Shang, and Q. M. Zhang, *Can dissimiliar users contribute to accuracy and diversity of personalized recommendation?*, International Journal of Mordern Physics, Vol. 21, No. 10, pp. 1217-1227, Jun 2010.

[7] A. Bellogin, I. Cantador, and F. Diez, *An empirical comparison of social, collaborative filtering, and hybrid recommenders*, ACM Transaction on Intelligent Systems and Technology, Vol. 4, No. 10, pp. 1-37, Jan. 2013.

[8] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, *Fuzzy c-means algorithms for very large data*, IEEE Transaction on Fuzzy Systems, Vol. 20, No. 6, pp. 1130-1146, Dec. 2012.

[9] X. Liu, G. Huang, and H. Mei, *Discovering homogeneous web service community in the user-centric web environment*, IEEE Transaction on Services Computing, Vol. 2, No. 2, pp. 167-181, Apr.-Jun. 2009.

[10] F. Chang, J. Dean, and S. mawat, *Bigtable: A distributed storage system for structured data*, ACM Transaction on Computer Systems, Vol. 26, No. 2, pp. 1-39, Jun. 2008.

[11] Z. Zheng, H. Ma, and M. R. Lyu, *Qos-aware web service recommendation by collaborative filtering*, IEEE Transaction on Services Computing, Vol. 4, No. 2, pp. 140-152, Feb. 2011.

[12] S. Marsland, J. Shapiro, and U. Nehmzow, *A self-organising network that grows when required*, Neural Network, Vol. 15, No. 8-9, pp. 1041-1058, 2002.

[13] A. Ultsch, *Self-organizing neural networks for visualization and classification*, in Information and Classification-Concepts, Methods and Applications, O.B. Lausen and R. Klar, Eds. Springer-Verlag, 1993.

[14] G. Thilagavathi, D. Srivaishnavi, and N. Aparna, *A survey on efficient hierarchical algorithm used in clustering*, International Journal of Engineering, Vol. 2, No. 9. Sep. 2013.

[15] T. Niknam, E. Taherian Fard, and N.

Pourjafarian, *An efficient algorithm based on modified imperialist competitive algorithm and K-means for data clustering*, Engineering Applications of Artificial Intelligence, Vol. 24, No. 2, pp. 306-317, Mar. 2011.

빅 데이터 환경에서의 골프스윙모델을 위한 하둡 기반 자가조직맵에 관한 연구

안완식

성결대학교 사범대학 체육학과 교수

요 약

헬스케어와 체력은 치열한 경쟁 환경에서 살아가는 생활 속에 고려해야 할 중요한 요소이다. 대부분의 사람들은 스포츠가 건강과 체력에 모두 좋은 특성을 가지고 있기 때문에 인간의 삶의 주된 요인으로 가장 선호하고 있다. 이 중 골프는 스포츠 모델에서 사람들이 선호하는 가장 좋은 운동중의 하나이며, 컴퓨터를 활용한 복잡한 동작을 설계함으로서 골프스윙모델을 개발하였다. 특히, 골프스윙모델은 골프 선수들의 운동학, 운동 에너지 및 운동량과 유사한 데이터를 기반으로 운동 적응력 및 속도 생성에 관련된 가장 필수적인 요수중의 하나이다. 이러한 이유로, 골프스윙모델은 과거의 골프 스윙에서의 회전 순서를 응용하여, 모션 패턴에 대한 전문 지식을 습득할 수 있다. 따라서, 본 논문에서는 골프스윙모델을 기반으로 분석과 모델링을 통한 시뮬레이션을 제안하고자 한다. 또한, 빅데이터 환경 기반 자가조직맵을 활용함으로서 골프스윙모델에서 요구하는 데이터를 분석 및 처리를 위한 실험을 하고자 한다.

Wan Sik An received the bachelor's degree and the M.S. degree in the Department of Physical Education from the Hanyang University in 1987 and 1990. He received the Ph.D. degree in the Department of Physical Education from Dan-kuk University in 2001, respectively. He has been a professor in the Department of Physical Education at Sungkyul University since 2005. His current research interests include Gymnastics, Sport Biomechanics. He is life member of the KAHPERD, KSSB, KSGS, KSSLS.

E-mail address: aws119@sungkyul.ac.kr