MOTION PATTERN TRACKING CLASSIFICATION IN BOWLING MATCHES

By:

DARYL TAN HOCK ANN

(Matrix No.: 120366)

Supervisor:

Dr. Loh Wei Ping

June 2017

This dissertation is submitted to
Universiti Sains Malaysia
As partial fulfilment of the requirement to graduate with honors degree in
BACHELOR OF ENGINEERING (MECHANICAL ENGINEERING)

School of Mechanical Engineering
Engineering Campus
Universiti Sains Malaysia

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed……………………………………………… (DARYL TAN HOCK ANN)

Date…………………………………………………

Statement 1

This journal is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/ references are appended.

Signed……………………………………………… (DARYL TAN HOCK ANN)

Date…………………………………………………

Statement 2

I hereby give consent for my journal, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations.

Signed……………………………………………… (DARYL TAN HOCK ANN)

Date…………………………………………………

Acknowledgements

I would like to express my gratitude to School of Mechanical Engineering, Universiti Sains Malaysia for providing students with the Final Year Project course. Assuredly, this course would serve well as a solid foundation for our future engineering careers preparing students with great tool and platform to learn and be competent leaders towards a successful career. The Final Year Project provides students with the ability to understand the real world problem and being able to propose reasonable and attainable solution with proper planning.

I would like to convey a special acknowledgement to my project supervisor, Dr. Loh Wei Ping. Throughout the project, her guidance and lecture was highly relatable and useful. Without her help, the quality of the solution and steps proposed would not be as high as presently proposed. In addition, Dr. Loh Wei Ping was always eager to offer her motivation in times of stagnant progress towards the smooth completion of the project.

Lastly, I grab the opportunity to thank our fellow course mates undertaking this course as well. Although having different methods and ideas, our constant discussions on the project offer great motivations that make this accomplishment possible.

Table of Contents

List of Tables

List of Figures

Abstrak

Pengesanan corak pergerakan adalah mengesan gerakan objek dan memindahkan data maklumat untuk analisis. Kajian lepas pengesanan gerakan dalam bidang sukan memberi tumpuan kepada perubahan kedudukan pemain-pemain dan bukan pada interaksi segmen-segmen badan. Justeru, suatu jurang telah dikenal pasti bagi kajian selanjutnya. Oleh itu, kertas ini membentangkan pendekatan pengesanan corak pergerakan bagi pengelasan postur-postur dalam permainan boling. Objektifnya adalah untuk (i)membentuk rangka kerja bagi mengenal pasti dan mengkelaskan corak gerakan, (ii)meneroka urutan strategi gerakan, (iii)menganalisis gerakan bagi mengenalpasti pergerakan relatif pemain mengikut urutan postur dominan dan (iv)mengkaji ciri-ciri corak gerakan oleh analisis pengelasan dan kaedah-kaedah penaakulan bawah beberapa attribut iaitu bahu, bengkok badan, keseimbangan, sudut ayunan dan jarak kaki. Gerakan badan dikesan pada urutan imej daripada video rekod dan data berangka diperoleh melalui perisian Photoshop. Data yang telah dipraproses dikelaskan menggunakan perisian WEKA untuk maklumat analitikal dan pengelompokan gerakan badan kepada tiga kelas pra-takrif: BAIK, SEDERHANA, BURUK. Pengelas utama yang dipilih ialah “Random Tree”. Penemuan kajian menunjukkan empat syarat utama berkenaan gerakan badan untuk menghasilkan postur gerakan badan BAIK untuk permainan bowling bowling: Petua 1: sudut bahu akhir adalah <109.32º dan sudut bengkok badan akhir <50.13, Petua 2: sudut keseimbangan akhir <89.03º dan perubahan maksimum dalam sudut ayunan <82.41º dan 48.27º< sudut bengkok badan akhir <50.10º, Petua 3: 17.78cm < perubahan maksimum dalam jarak kaki <69.82cm) dan sudut bengkok badan akhir >39.17º, Petua 4: perubahan maksimum dalam jarak kaki >69.82cm dan sudut bengkok badan akhir <51.48º dan perubahan maksimum dalam sudut ayunan <56.19º.

Abstract

Motion pattern tracking is the tracing of object movements and transferring the informative data for analyses. Previous motion tracking studies in sports focused on the changing position of the players rather than actual body segment interactions. Thus, a gap is identified for further research works. Therefore, this paper presents a motion pattern tracking approach for bowling game posture classifications. The objectives are to (i)design a framework to recognize and classify motion patterns, (ii)explore sequences of motion strategies, (iii)analyze motion to recognize bowlers' relative movements by dominant posture sequences and (iv)examine motion pattern characteristics by classification analysis and rules-reasoning under several parameters namely shoulder, body bend, balance, swing angles and distance of feet. Motion is tracked on sequential image frames of video records and the numeric data retrieved using the Photoshop tool. Preprocessed data are classified using WEKA software for analytical information and grouping body motions into three predefined classes: GOOD, MODERATE, BAD. The main classifier chosen is the Random Tree. The findings show four main conditions concerning body motion to result in GOOD body motion postures for bowling mainly Rule 1: final shoulder angle is $<109.32^{o}$ and the final body bend angle$<50.13^{o}$, (Rule 2: final balance angle$<89.03^{o}$ and the maximum change in swing angle$<82.41^{o}$ and $48.27^{o}<$ final body bend angle$<50.10^{o}$, Rule 3: 17.78cm$<$maximum change in distance of feet$<$69.82cm and final body bend angle$>$39.17Rule 4: maximum change in distance of feet$>$ 69.82cm and final body bend angle$<51.48^{o}$ and maximum change in swing angle$<56.19^{o}$.

1. Introduction

Motion pattern tracking, sometimes being referred as the match moving, is the tracing of object movements and transferring of informative data for further analyses. Motion tracking, commonly known as the motion capture (mocap) includes the capturing motions of objects for matching with its stored motion library template. Its applications are mostly found in military, entertainment industry, medical applications, computer vision, robotics and even in the sports science field. Analyses reported in sports science studies attempted to understand relative movements of body segments and its relations to the successful performance in sports tournaments.

Previous studies had reported relative movements of athletes across the game, field or court or the movement of the game ball. Greater emphasis is placed on the changing position of the players or ball rather than the actual body segment interactions. For instance, Liu et al. (2009) had automatically tracked the movements of ice skaters on a large-scale complex and dynamic rink. The purpose of their study was to capture highly complex and dynamic scenes under fast moving camera. Ren et al. (2008) on the other hand, demonstrated the innovative techniques for estimating the trajectory of a soccer ball from multiple fixed cameras. The authors proposed an updated method for soccer ball detection and tracking from real video sequences. In it, a local matching process was proven to be effective in compensating the Kalman tracker in order to deal with merged balls. Thus, it can be seen that the current studies have mostly looked into positional changes of players and objects rather than the interactive movements of body segments. Previous works have also used data mining application in sports studies. For example, Cao (2012) focused on using machine learning algorithms to build a model for predicting the NBA game outcomes. The algorithms used in Cao (2012) involved Simple Logistics Classifier, Artificial Neural Networks, SVM and Naïve Bayes. The study considered automated data collection and cloud techniques to enable data management, a data mart containing NBA statistics data was built.

The previous works have observed some lacking from the aspect to relate different body segment motions during sports matches. This aspect motivates the current project to understand interactions of different body segments in games and how one body segment corresponds to another as well as work together to execute a specific motion. This project focuses on the bowling game as the case study; considering the distinctive motion patterns from several body segments to reflect the game performance during the ball throw. The case study was opted for two main reasons. Firstly, the bowling game requires integration of different body segments such as shoulder, back, arms and legs to efficiently deliver a consistent solid ball throw. Therefore, it is of great interest to understand how the body segments coordinate well to achieve the most efficient and successful throw. Secondly, there has been no significant study on bowling game reported so far and thus shall lead to a very broad area of new findings.

In standard bowling games, for the amateurs or beginners, there is lack of knowledge and insight on the proper form and posture needed to execute good shots. No significant academic research works has been carried out on the sport which may provide an experimental proof on which approach would be most efficient. Besides, data mining approach that provides information on useful parameters to be focused by the bowler is yet to be presented. With no analytical data on the sports concerning body motions, beginners are unaware on what body motion to adopt in order to bowl well.

The general purpose of the project is to track the bowler's motion patterns during matches and to classify the patterns on how performance of the players can be enhanced. The specific study objectives include to (i)design a framework that allows the recognition and classification of motion patterns in bowling matches, (ii)explore the sequences of motion strategies, (iii)analyze motion under several bowling shots to recognize bowlers' relative movements by dominant posture sequences and (iv)examine motion pattern characteristics by classification analysis and rules-reasoning under several identified parameters.

Efforts mainly involve designing a framework that allows the recognition and classification of motion patterns in bowling matches and classifying the sequences of motion strategies by type of game performances. Hence, while analysing motion under several shots to recognize the bowlers' relative movements, the emphasis are put on the dominant posture sequences and motion pattern characteristics by classification and rule-reasoning under considerable parameters. The first step involves experimental data collection with proper planning on the focus parameters;

which include the shoulder angle, balance angle, body bend, distance of feet and swing angle. Collected data undergo pre-processing stage in which the video data are transformed into numeric forms followed by data cleaning and filtering approach on potential outliers or any misleading data. Data classification analyses are performed on preprocessed data. Classification analyses by type of bowler postures characterized patterns according to the relevance of shoulder angle, balance angle, body bend angle, distance of feet and swing angle are carried out. The interpretation and evaluation is considered by if-else rules reasoning at knowledge discovery level. The distinctive constraint factors from the classified groups is analysed and converted into understandable relationships among dominant parameters. The movement of separate body segments to form good posture motion enables well executed shots. The knowledge learnt could be translated into informative discovery on how to maximize efficiencies in bowling game motion.

2. Literature Review

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. In motion pattern tracking, data mining has greatly improved the accuracy of a system to detect changes in motion as well as predicting sequential movements. In sports, data mining is used to either predict the results of games based on historical recorded events or predicting motions or movement of entities related to the sports.

In sports data mining, Cao (2012) focused on using machine learning algorithms to build a model for predicting the NBA game outcomes. Simple Logistics Classifier, Artificial Neural Networks, SVM and Naïve Bayes were applied. The study involved automated data collection and cloud techniques enabled data management, a data mart containing NBA statistics data being built. The limiting factor found was that due extracted feature relied on statistics from last 10 home/road games; models trained in the project would not have as high accuracy when predicting NBA games of first two months in a regular season as predicting the reset of the season.

Leung (2014) presented the sports data mining approach which helped to discover interesting knowledge and predict the outcomes of sports games such as college football. The approach returned predictions based on a combination of four different measures on the historical results of the games instead of the traditional approach of comparing statistics of two competing teams and predicting the outcome. The evaluation results showed the accuracy of the sports data mining approach in predicting the outcomes of football games in recent seasons.

Thomas et al. (2017) discussed various fundamental techniques are being applied in commercially available systems today that use computer vision for sports analysis such as camera calibration and tracking, player detection and tracking, as well as player modelling. Some topics that are currently being addressed in the research community are highlighted such as the inability to differentiate players among same team due to similar appearances and a lack of a comprehensive public dataset.

Liu et al. (2009) had automatically tracked the movements of ice skaters on a large-scale complex and dynamic rink. The purpose of their study was to capture highly complex and dynamic scenes under fast moving camera. Therefore, tracking amorphous skaters became a challenging task. However, the main problem in the proposed system is on how to improve the tracking performance when skaters are moving in groups during a long and continuously full occlusion.

Gwak, J. (2017), whereas, proposed a method to track multiple objects based on a single reference target object by finding discriminative relational feature differences. The method solves the problem of maintaining ID of an object partial occlusions and missed detections. The current drawback is on dealing with large class numbers in image sequences or videos. This is because such effort requires the same number of RFDs equivalent to the number of classes.

Ren et al. (2008) demonstrated the innovative techniques for estimating the trajectory of a soccer ball from multiple fixed cameras. An updated method for soccer ball detection and tracking from real video sequences was proposed. In it, a local matching process proved to be effective in compensating the Kalman tracker to deal with merged balls. The application of occlusion-reasoning and tracking-back results was significant in improving tracking accuracy and continuity of the ball trajectory.

Zhang et al. (2017) introduced the Martial Arts, Dancing and Sports dataset (MADS). The MADS dataset contains five categories of challenging actions; Tai-chi, Karate, jazz, hip-hop and a sports combo. Two martial art masters, two dancers and an athlete performed these actions while being recorded with either multiple cameras or a stereo depth camera. In the multi-view or single-view setting, three colour views for 2D image-based human pose estimation algorithms was provided whereas for depth-based human pose estimation, stereo-based depth images from a single view was provided. The results of the evaluation suggested that discriminative approaches perform better than generative approaches when there are enough representative training samples.

On the other hand, Vyas et al. (2015) presented object tracking in a multi-sport field. It was proposed to modify certain elements of the original Mean Shift algorithm so as to track entities in video streams with changing colour, shape and direction enabling calculation of the distance covered by an entity in the field. In the proposed method, the probability factor of tracking the right object is sufficiently high. A key feature was that the excess spatial information that was not desired for tracking of the player was ignored by separating the background from the player's pixels.

The key technical parameters that professional golf coaches associate with a top level golf swing was studied in Smitha et al. (2012), with the intention of using the results to enhance future golf biomechanics research and coaching technologies. A successful golf swing was defined through three elements, with "body motion" affecting "club motion" and resulting "ball flight". A golfer's "body motion" was highlighted by golf coaches and "posture" was identified as one of the five key technical parameters.

The overall state-of-the-art reviews gave indicators for this study, that is to analyze the primary data with main focus on how movements and relationship between body segments interact together smoothly in a particular sport. This is one of the key aspects of this study because a proper understanding of the relationship of selected dominant body segments could deliver the guide towards either BAD, MODERATE or GOOD bowling shots.

3. Methodology

The research implementation is generally divided into four stages: Data Collection, Preprocessing, Data Classification and Interpretation and Evaluation (Figure 1).

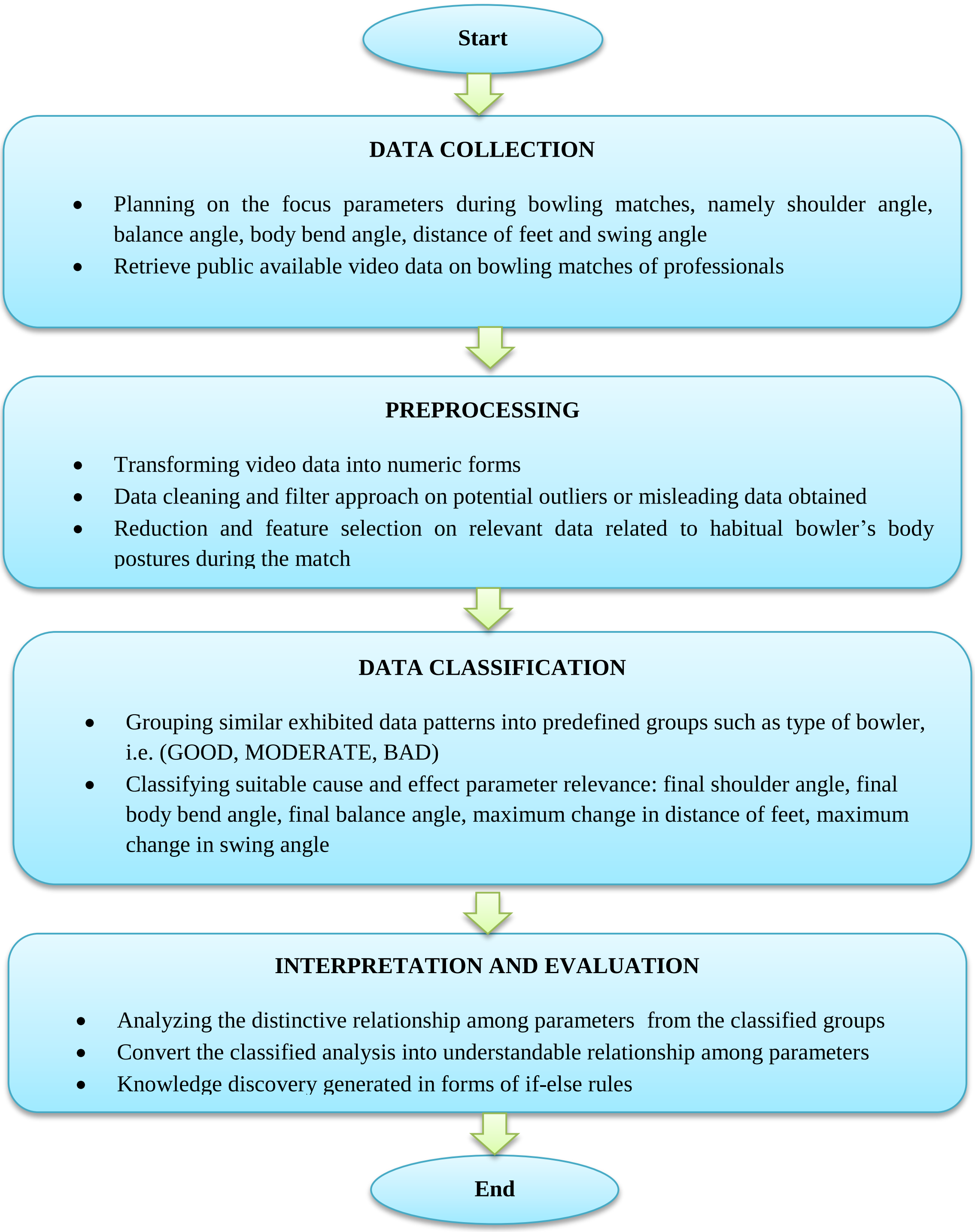

Figure 1: Flow chart of data mining approaches in understanding the bowling motion patterns

3.1 Data Collection

The study data are collected from two case studies: experimental study on bowling games and public available bowling video. The experimental dataset involved 10 amateur bowlers of various handedness, level of experience and gender chosen as the study subjects. Two cameras are stationed at two identified angle locations; back and side of the bowler (Figure 2). Each bowler was required to bowl five bowling shots on one lane and the video recording began from every shot thrown by the bowler. A total of 10 videos from separate shots for each bowler will be collected and saved by file names Bowler No *i* (*i* = bowl repetitions). The second case study involved eight public available bowling videos performed by two professional bowlers of different gender and bowling styles. Four videos for each professional bowler, among which two were captured from the back view and the other two from the side view similar to that of the amateur bowlers video records.

Camera 1: Record the back view of the bowler

Camera 2: Record the side view of the bowler

Figure 2: Cameras positioning during the recording of bowling game

3.2 Data Preprocessing

The raw video recordings for both case studies undergo pre-processing for editing for qualitative inspections. Outliers in bowlers' motions or irrelevant segments werediscarded to ensure no misleading data obtained. Among the five bowling shots taken by each amateur bowler, only two shots were chosen for analysis. The video sample files were segregated by the back and side view angle shots. Data transformation efforts took place to convert the original raw data from video into numeric forms for data recordings through the static image snapshots (video-image-numeric).

For each progression of step taken beginning with the 0^{th} step, the video is freeze into still shots and screenshot until a complete bowling shot is thrown. For example, if a bowler has a total of 5 steps including the 0^{th} step, 6 screenshots are taken, one of each step. Screenshots of such steps are considered observable to differentiate sequential postures. Since every bowler has their own ball throw style respective to different number of steps, individual video was separated into about 4 to 6 frames depending on how many steps taken by the bowler.

The screenshots are deployed to Photoshop, a photo editing software. The software extracts measurements of at each bowler's step progression; shoulder angle from the back view and balance angle, body bend angle, distance of feet and swing angle from the side view. Only significantly visible parameters from the bowler's postures: and position and knee bend observations corresponding to the game performance were extracted as the study parameters. In order to represent the data in a more extensive form, further calculations on the rate of change of average shoulder angle, balance angle, body bend angle, distance of feet and swing angle are also used.

Generally, based on the motion views, three bowling levels are predefined being data class; GOOD, MODERATE and BAD. A bowler who is labelled as a GOOD bowler is the bowler whose body posture does not change much throughout the approach, as well as having good balance and linearity of form. On the other hand, a BAD bowler is a bowler whose posture changes greatly throughout the approach; having large rate of changes thus losing balance as well as linearity of motin.

Table 1 generally shows the description of parameters measured. The recorded numeric dataset are stored in .csv format readable by the Waikato WEKA tool.

Table 1: Description of parameters measured

Parameter	Description	Scale type	Data range
Name	Aaron, Alimie, Angles, Chim , Daryl, Epol, Ipin, Lean, Syed, Teoh, Duke, Carolyn	Nominal	-

Time step	Time between consecutive steps	Numeric	{0,1,2,…,6} (s)
Shoulder angle	Angle between arm and body	Numeric	50 - 120 (°)
Balance angle	Angle between body plane and base	Numeric	70 - 100 (°)
Body bend angle	Angle between back and base	Numeric	35 - 100 (°)
Distance of feet	Distance between left and right food on consecutive steps	Numeric	0 - 13 cm
Swing Angle	Angle between forearm and upper arm	Numeric	30 – 180 (°)

The raw recorded data were segregated into three samples namely Sample A, B and C.

3.2.1 Sample A

Sample A concerns the direct raw data obtained through measurements using the Photoshop Software (Appendix E). There is no preprocessing works performed on the sample so the data is maintained in its original form with 40 missing values due to some bowlers having fewer amount of steps prior to bowl throw.

3.2.2 Sample B

The Sample B aims to rectify the missing values seen in Sample A (Appendix F). The preprocessing technique of substituting missing values with other values was adopted. In this case, all the missing data was filled up by which each bowler essentially had 6 steps meaning 6 data per throw. This is done by repeating the value of the 0th step of each bowler respectively to where the bowler ended on the 6th step. For example, as shown in Table 2, Aaron initially performed ball throw in 4 steps. Thus, with this preprocessing technique, the first step is repeated 3 times to where the final value ends on the 6th step. (Table 3)

Table 2: Example of Sample A data for a bowler

Name	0 step	1st step	2nd step	3rd step	4th step	5th step	6th step
Aaron	94	97.45	63.65	113.75	?	?	?

Table 3: Example of Sample B data for a bowler after preprocessing

Name	0 step	1st step	2nd step	3rd step	4th step	5th step	6th step
Aaron	94	94	94	94	97.45	63.65	113.75

3.2.3 Sample C

The setback of the Sample B is that it was unable to provide any useful information or visual illustrations for further analysis. Therefore, Sample C data was generated (Appendix G). In Sample C, instead of retrieving data on every single step of each bowler, only the selective data of each parameter were chosen by justifications as explained in Section 4.3. Table 4 shows the parameters that are chosen as well as the layout of the data collected.

Table 4: Layout data on selective parameters used for Sample C

Gender	**Handed-ness**	**No. of steps**	**Final Shoulder Angle**	**Final Balance Angle**	**Final Body Bend Angle**	**Max. Change in Distance of Feet**	**Max. Change in Swing Angle**	**Class**
M	R	3	113.75	85.3	51.45	103.407155	79.83247423	BAD
M	R	5	111.7	89.45	53.15	130.8943089	76.6057749	MODERATE
F	R	4	108.6	84.55	48.5	22.70955166	77.36418511	GOOD
:	:	:	:	:	:	:	:	:

3.3 Data Classification

Data classifiers namely ZeroR, InputMappedClassifier, NaïveBayes Multinomial, REP Tree, J48 Decision Tree, Hoeffding Tree, and Random Tree are used for classification analysis. These classifiers were selected due to their capability to provide consistent classification accuracies as well as visual representation of relationship among set parameters. The analysis groups the bowler's performance level by their posture variations. In bowling games, there are three notable levels: GOOD, MODERATE and BAD. The predefined classification by these levels is justified based on several reasons. Firstly, the level of GOOD is awarded to motions used by the professional bowlers (public available video-2nd case study) based on their notable success in the game. The MODERATE classification is defined for bowlers whose motions correspond to bowling performance at closer similarity to those motions adopted by professionals. This can be in terms of general body motion comparison as well as general accuracy of throw relating to bowler's set target on the lane. Meanwhile, the BAD classification is given to bowlers whose motions are inconsistent along with large difference postures from those shown by professional bowlers. The inconsistency can be referring to the variety of direction when the bowling ball travels towards a fixed target. Also, large difference refers to the general body motion when compared to the professional through visual checks besides a thorough comparison from the recorded data. The three bowling game levels are predefined as the data classes for further instance mappings.

The goal was set for classifying parameter data which would provide foundation for the analysis on how certain parameters must be for a GOOD bowling motion to be adopted by a bowler. The classification tasks are implemented with the aid of WEKA tool through classifier algorithms. The dataset was split by parameters to ensure a relationship between most parameters chosen can be determined. Thus, several runs on similar classification algorithms were performed on the different sets of conditions. All study data models were tested using full training set mode due to the small sample size of data available involving merely 12 bowlers. The full training mode enable more comprehensive results considering all the 12 bowlers' motions during game.

3.4 Interpretation and Evaluation

The classified groups are interpreted and evaluated mainly in terms of classification accuracy. Several performance metrics were considered including

i. correctly classified instances
ii. incorrectly classified instances
iii. Kappa statistic
iv. mean absolute error
v. root mean squared error
vi. relative absolute error
vii. root relative squared error
viii. total number of instances
ix. confusion matrix

Strong cause effect perspective relating the key body posture parameters with the classification performance is considered from multiple visual observations from tree diagrams constructed from J48 Decision Tree and Random Tree. The analysis is converted into if-else rules intellectual reasoning. Further analysis is performed within the respective groups to further simplify the reasoning process. The collective results were used to generate case-study based reasoning rules portraying the knowledge discovered through the data analysis to distinguish bowler's body posture pattern by level of bowling style adopted.

4. Results and Discussion

The classification results from the experimental case study are segregated by three samples: Sample A, Sample B and Sample C. Sample A is concerned with the raw data collected (Appendix E), Sample B is the preprocessed raw data collected by substituting missing values (Appendix F) and Sample C applies the selective parameters (Appendix G) as discussed in previous Section 3.2.3.

4.1 Sample A Details

In sample A, due to the different amount of steps taken by each bowler (as mentioned in Section 3.2), there are 8 missing values for each chosen parameter. Thus for all 5 parameters, there are a total of 40 missing values.

Hence, the amount of possible data classifiers to be used is greatly lacking since most classifiers such as J48 Decision Tree and Random Tree require complete data to provide any useful classification. Three algorithms used for Sample A, namely ZeroR, InputMapped Classifier and NaiveBayes Multinomial Text resulted in only 50% correctly classified instances (Table 5). Due to the low percentage of accuracy, the classification performed for Sample A can only provide partially useful knowledge.

Table 5: Summary of classification results from ZeroR, InputMappedClassifier, and NaïveBayes Multinomial Text for Sample A

Sample A Summary			
Classifier	**ZeroR**	**InputMapped Classifier**	**NaiveBayes Multinomial Text**
Correctly Classified Instances	6 50 %	6 50 %	6 50 %
Incorrectly Classified Instances	6 50 %	6 50 %	6 50 %
Kappa statistic	0	0	0
Mean absolute error	0.4148	0.4148	0.4148
Root mean squared error	0.4522	0.4522	0.4522

Relative absolute error	100 %	100 %	100 %
Root relative squared error	100 %	100 %	100 %
Total Number of Instances	12	12	12
Confusion Matrix a = BAD b = MODERATE c = GOOD	a b c | . 0 4 0 | a 0 6 0 | b 0 2 0 | c	a b c | . 0 4 0 | a 0 6 0 | b 0 2 0 | c	a b c | . 0 4 0 | a 0 6 0 | b 0 2 0 | c

4.2 Sample B Details

The weak performance observed in Sample A was rectified on Sample B is on the missing values treatment associated with the varying amount of steps taken by respective bowlers. Thus, a wider range of classifiers were found appropriate to classify Sample B. The algorithms selected to classify Sample B include NaiveBayes Multinomial Text, J48 Decision Tree and Rep Tree.

Using the NaiveBayes Multinomial Text and Rep Tree, the percentage of correctly classified instances is only 50 % (Table 6). Thus, this classifier is considered weak to produce reliable knowledge. Since the accuracy is low, no further steps or analysis can be carried out on its results.

Table 6: Summary of classification results from NaïveBayes Multinomial Text, REP Tree, J48 Decision Tree for Sample B

Sample B Summary			
Classifier	**NaiveBayes Multinomial Text**	**REP Tree**	**J48 Decision Tree**
Correctly Classified Instances	6 50 %	6 50 %	11 91.6667 %
Incorrectly Classified Instances	6 50 %	6 50 %	1 8.3333 %
Kappa statistic	0	0	0.8667

Mean absolute error	0.4148	0.4074	0.0889
Root mean squared error	0.4522	0.4513	0.2108
Relative absolute error	100 %	99.2143 %	21.4286 %
Root relative squared error	100 %	99.8187 %	100 %
Total Number of Instances	12	12	12
Confusion Matrix a = BAD b = MODERATE c = GOOD	a b c | . 0 4 0 | a 0 6 0 | b 0 2 0 | c	a b c | . 0 4 0 | a 0 6 0 | b 0 2 0 | c	a b c | . 0 4 0 | a 0 6 0 | b 0 2 0 | c

On the other hand, an efficient algorithm using the J48 decision tree was selected to present clearer and understandable visual tree diagram. J48 Decision Tree resulted in the percentage of correctly classified instances being 91.67%. Although the high accuracy was a positive step, however, unlike initially thought, the tree created through this classifier as shown in Figure 3 does not any provide any relatable results because the results neglect a large majority of the parameters involved. Hence, no further steps or analysis can be made when using J48 for Sample B.

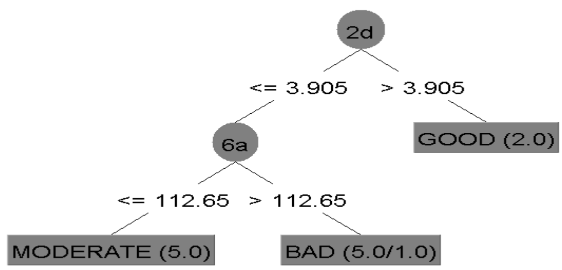

Figure 3: Tree diagram for Sample B classification using J48 Decision Tree

Findings from the first two sample data sets, Samples A and B indicate that the initial approach of having the coordinates at each step and presenting all these data to WEKA was not the most useful approach to provide any useful analysis. This indicates the inability data to explore fruitful information regarding good sequence of motions in the bowling game.

4.3 Sample C Details

Owing to the inability to present any useful informative classification or visual illustrations, a different approach is considered leading to Sample C. Comparing to Samples A and B, the Sample C does not consider the measurement for every step taken by the bowler for each and every parameter. However, a greater focus has been placed on key steps without neglecting the earlier set parameters. In this approach, the key parameters include the shoulder angle, balance angle, body bend angle, distance of feet and swing angle were maintained. However, instead of having the coordinates for every step, only selected data for each parameter were chosen based on the following criterion:

Shoulder angle, Balance angle, Body bend angle- the coordinate at the final step is chosen. For shoulder angle, based on the steps before the last step, there is no great disparity between all bowlers. The only notable difference was the coordinate of the final step. For the balance angle, the most important step is the final time step since it is the moment when the ball is released for throw. Thus, if there is a good balance in the final step, a more accurate and consistent shot can be made. Therefore, this parameter would serve to provide useful information for analysis to find out what final balance angle would be optimum to deliver a GOOD bowling shot. For body bend angle, the final step is chosen similar to balance angle, the final step is key in delivering a GOOD bowling shot because the body bend for the final step determines the amount of leverage a bowler has during his or her throw.

Distance of feet, swing angle - the step with the largest rate of change is chosen regardless of number of step. Since every bowler is of different body sizes, thus posing different leg lengths resulting in various lengths of step taken. Thus, a comparison between the distance of feet in various bowlers would not provide any useful information. Hence, the rate of change is chosen instead. The largest percentage difference would serve as a good indication and provide useful analysis. For swing angle, similar to distance of feet, the step with the largest rate of change is chosen. Since every bowler has their own respective swing, the rate of change would serve as a good parameter because although each person’s swing is different, the rate of change of swing angle is comparable. This rate of change would provide useful information for further analysis.

Sample C benefits for its more direct and simplified approach without compensating the value of parameters chosen.For Sample C, three classifiers, namely Hoeffding Tree, J48 Decision Tree and Random Tree were chosen. The results were compared and the most suitable and useful classifier would be chosen for its results explained in Section 4.2. For Sample C, three different sets are created. The 1st set relates final shoulder angle, final body bend angle and maximum change in swing angle, 2nd set relates to final body bend angle, final shoulder angle, final balance angle and maximum change in swing angle, 3rd set relates to maximum change in distance of feet, final body bend angle, and maximum change in swing angle.

4.4 Selection of Classifier Results for Analysis

Among the three classifiers (Hoefding Tree, J48 Decision Tree and Random Tree) used for the Sample C, further analysis is performed on the results and tree diagrams constructed from the Random Tree algorithm.

Random Tree is chosen here for several justifications. As shown in Table 7, the first classifier used, Hoeffding Tree has a lower accuracy (80.6 %) as compared to the other two classifiers showing perfect 100% accuracy. Besides, the Hoeffding Tree does not enable visual representation of the classification.

Table 7: Summary of classification results from Hoeffding Tree for 1st, 2nd and 3rd data set of Sample C

Sample C Summary			
Classifier: Hoeffding Tree	**1st set**	**2nd set**	**3rd set**
Correctly Classified Instances	9 75 %	10 83.3333 %	10 83.3333 %
Incorrectly Classified Instances	3 25 %	2 16.6667 %	2 16.6667 %
Kappa statistic	0.5184	0.7143	0.7143
Mean absolute error	0.1446	0.1381	0.1381

Root mean squared error	0.2899	0.2721	0.2721
Relative absolute error	34.8632 %	33.2945 %	33.2945 %
Root relative squared error	64.118 %	60.1859 %	60.1859 %
Total Number of Instances	12	12	12
Confusion Matrix a = BAD b = MODERATE c = GOOD	a b c | . 2 2 0 | a 1 5 0 | b 0 0 2 | c	a b c | . 2 2 0 | a 0 6 0 | b 0 0 2 | c	a b c | . 2 2 0 | a 0 6 0 | b 0 0 2 | c

Though both J48 Decision Tree and Random Tree show perfect 100% classification accuracy as indicated in Tables 8 and 9, however the visual trees constructed obtained differs in all data samples except for 1st set Tree Diagram (Figure 4).

Table 8: Summary of classification results from J48 Decision Tree for 1st, 2nd and 3rd data set of Sample C

Sample C Summary			
Classifier: J48 Decision Tree	**1st set**	**2nd set**	**3rd set**
Correctly Classified Instances	12 100 %	12 100 %	12 100 %
Incorrectly Classified Instances	0 0 %	0 0 %	0 0 %
Kappa statistic	1	1	1
Mean absolute error	0	0	0
Root mean squared error	0	0	0
Relative absolute error	0 %	0 %	0 %

Root relative squared error	0 %	0 %	0 %
Total Number of Instances	12	12	12
Confusion Matrix a = BAD b = MODERATE c = GOOD	a b c | . 4 0 0 | a 0 6 0 | b 0 0 2 | c	a b c | . 4 0 0 | a 0 6 0 | b 0 0 2 | c	a b c | . 4 0 0 | a 0 6 0 | b 0 0 2 | c

Table 9: Summary of classification results from Random Tree for 1^{st}, 2^{nd} and 3^{rd} data set of Sample C

Summary			
Classifier: Random Tree	**1^{st} set**	**2^{nd} set**	**3^{rd} set**
Correctly Classified Instances	12 100 %	12 100 %	12 100 %
Incorrectly Classified Instances	0 0 %	0 0 %	0 0 %
Kappa statistic	1	1	1
Mean absolute error	0	0	0
Root mean squared error	0	0	0
Relative absolute error	0 %	0 %	0 %
Root relative squared error	0 %	0 %	0 %
Total Number of Instances	12	12	12
Confusion Matrix a = BAD b = MODERATE c = GOOD	a b c | . 4 0 0 | a 0 6 0 | b 0 0 2 | c	a b c | . 4 0 0 | a 0 6 0 | b 0 0 2 | c	a b c | . 4 0 0 | a 0 6 0 | b 0 0 2 | c

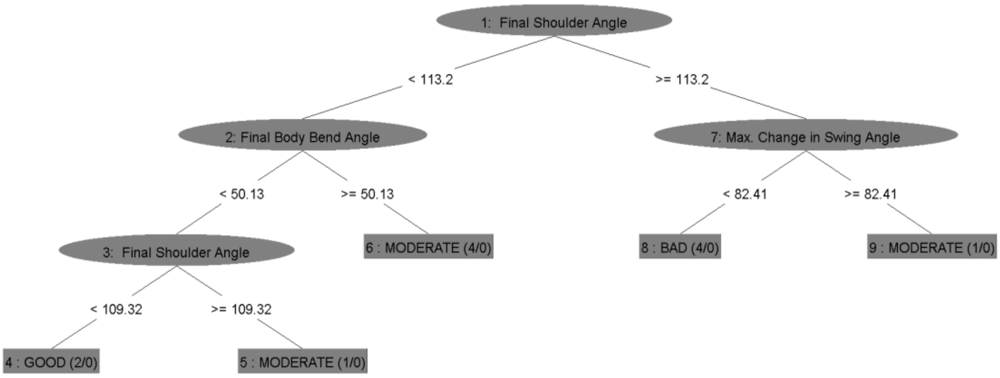

Figure 4: 1st set Tree diagram for J48 Decision Tree and Random Tree

In 1st set data tree diagram, the parameters used for algorithms classification include final shoulder angle, final body bend angle and maximum change in swing angle. As seen from the trees shown, the classification findings are similar.

However, for 2nd set Tree Diagram and 3rd set Tree Diagram, the classifications are based on different parameters. For 2nd set, the chosen parameters are final balance angle, maximum change in swing angle, final body bend angle and final shoulder angle. As seen from Figures 5 and 6, for classifier J48 Decision Tree, the classification does not include the final body bend angle whereas the classifier Random Tree includes it as well as having different stages of classification. This is not a particularly a mistake or error, however, for reasons of further analysis, it would be ideal to have to a greater interaction between as many parameters as possible.

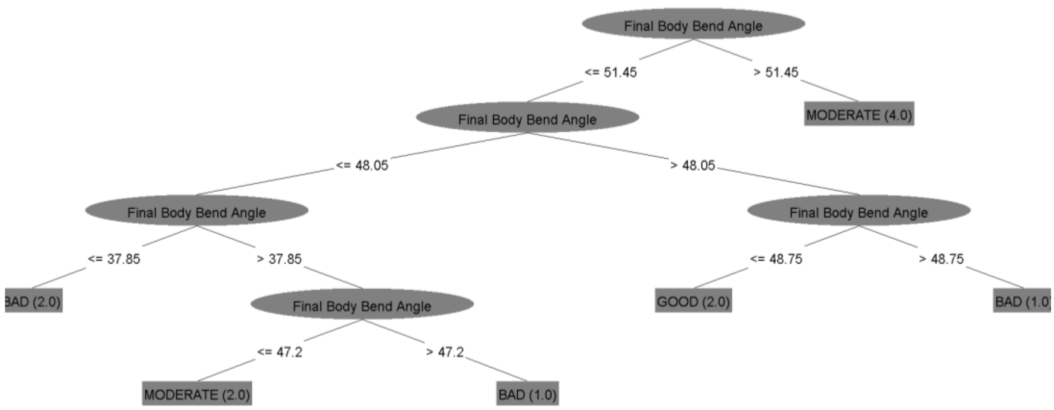

Figure 5: 2st set Tree diagram for J48 Decision Tree

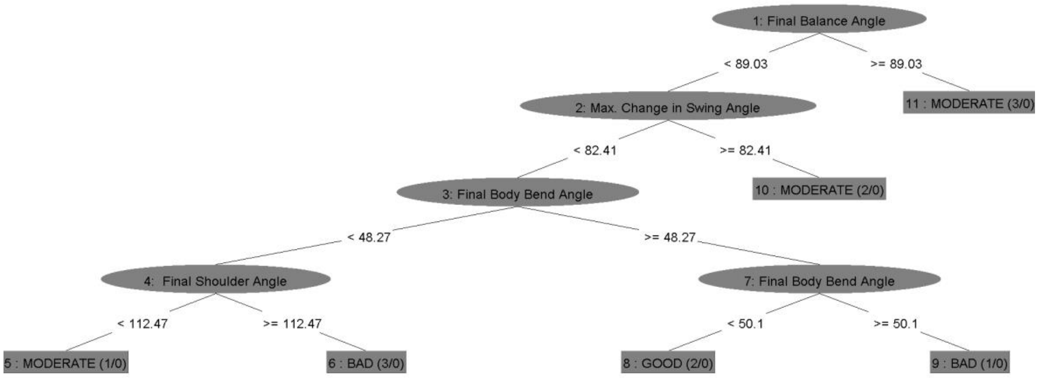

Figure 6: 2nd set Tree diagram for Random Tree

Next, for 3rd set, the parameters chosen for WEKA are maximum change in distance of feet, maximum change in swing angle, final body bend angle. As seen from Figures 7 and 8, for classifier Random Tree, there is a proper classification among there parameters where further analysis can be made. However, for classifier J48 Random Tree, the classification is only done for final body bend angle. Again, though it is not an error however it does not assist in further analysis. Thus, the classifier Random Tree is chosen.

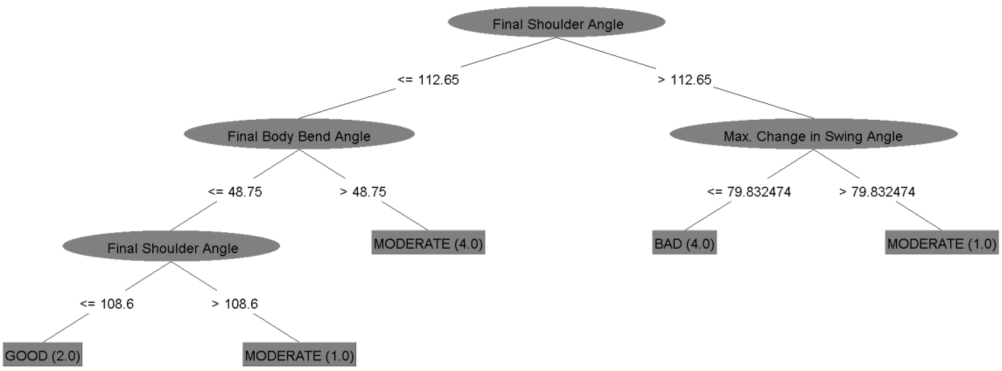

Figure 7: 3rd set Tree diagram for J48 Decision Tree

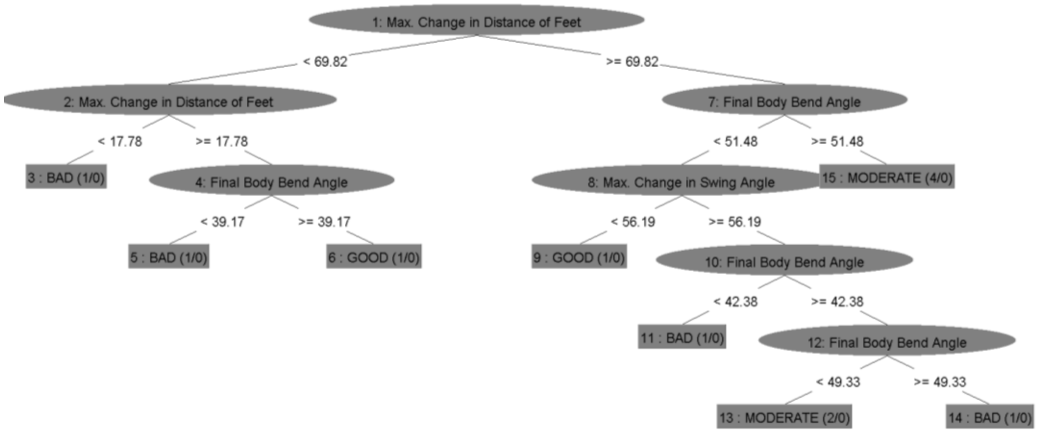

Figure 8: 3rd set Tree diagram for Random Tree

4.5 Knowledge Discovery

A closer inspection with reference to Figures 4, 7 and 8 of Random Tree classifier focusing merely on the ‘GOOD’ bowling posture, if-else reasoning statement rules are summarized as follows:

Based on 1st set**,**

a) If (final shoulder angle < 113.2^o) AND
(final body bend angle < 50.13^o) AND
(final shoulder angle < 109.32^o), THEN class=GOOD
else if
(final shoulder angle > 109.32^o), THEN class=MODERATE

b) If (final shoulder angle < 113.2^o) AND
(final body bend angle > 53.13^o) , THEN class=MODERATE

c) If (final shoulder angle > 113.2^o) AND
(maximum change in swing angle < 82.41^o), THEN class=BAD
else if
(maximum change in swing angle > 82.41^o), THEN class=MODERATE

Based on 2nd set,

d) If (final balance angle < 89.03^o) AND
(maximum change in swing angle < 82.41^o) AND

(final body bend < 48.27^o) AND

(final shoulder angle < 112.47^o), THEN class=MODERATE

else if

(final shoulder angle > 112.47^o) THEN class=BAD

e) If (final balance angle < 89.03^o) AND

(maximum change in swing angle < 82.41^o) AND

(48.27^o < final body bend angle < 50.1^o) THEN class=GOOD

else if

(final body bend angle > 50.1^o) THEN class=BAD

f) If (final balance angle < 89.03^o) AND

(maximum change in swing angle > 82.41^o) THEN class=MODERATE

g) If (final balance angle > 89.3^o) THEN class=MODERATE

Based on 3rd set,

h) If (maximum change in distance of feet < 17.78 cm) THEN class=BAD

i) If (17.78 cm < maximum change in distance of feet < 69.82 cm) AND

(final body bend angle < 39.17^o) THEN class=BAD

else if

(final body bend angle > 39.17^o) THEN class=GOOD

j) If (maximum change in distance of feet > 69.82 cm) AND

(final body bend angle < 51.48^o) AND

(maximum change in swing angle < 56.19^o) THEN class=GOOD

k) If (maximum change in distance of feet > 69.82 cm) AND

(final body bend angle < 51.48^o) AND

(maximum change in swing angle > 56.19^o) AND

(final body bend angle < 42.38^o) THEN class=BAD