

A Study on Sports Data Matching for Virtual Competitions

Kai-Li Wang and Kai-Chun Wang

Abstract—Technological advances have given rise to increasingly diverse sports data and applications. Among them, integrated sensors, big data and cloud computing are the most intriguing ones. Their combination with social media enables users to share their sports data, expanding their network. In this study, we will go through a screening analysis model that employed advanced mathematics, and derive our solution through the approximation theory. Our goal is to find a more accurate matching model for users of a big sports database to join a virtual competition. Compared with traditional sports rating methods, our desired model can more flexibly match each piece of sports data in a practical manner.

Index Terms—Sports data matching, screening analysis, approximation theory, virtual competition.

I. INTRODUCTION

A. Research on Sports Data

Sports data can range very widely: from the results of games, to an athlete's physical stats or training programs. In the field of information technology, studies used to be centered around the analysis of sports games, with most of them focused on sports betting [1], [2]. There were also, of course, some research concerning non-betting prediction of game results [3]. While topics such as physical stats or training programs should fall under the categories of sport physiology and sport coaching science, the advance of sensing elements has urged more researchers to explore today's diversified wearable technology. Traditional studies confined themselves to the relevance between physiological data and sports performance. Two examples of this include the study on the association between a Formula 1 racer's heart rate and his racing score [4], and that between a boxer's heart rate and his muscle response time [5]. On the other hand, some researchers were dedicated to integrating new technology into training, for instance, advanced heart rate monitors that can improve training efficiency and prevent accidents from happening [6], [7]. Further examples are sensor-based assisted bicycle trainers (ABTs) system, and wireless sensor networks (WSNs). Designed to track cycling performances, these technologies allow trainees to more accurately adjust their speed and help each other [8], [9]. Many other studies explored the application of information technology to commercial products or services. For instance, a runner's real-time physical status can be obtained by

monitoring his heart rate. By transmitting this data to a treadmill that supports realistic sports casting, the image's speed can then be adjusted in line with the runner's physical status [10]. Another featured function is the playback of different music files. The music informs users about their exercising stages by monitoring biological data and collecting feedback. This way, they may be more motivated to do sport [11]. Rather than heartbeat sensors, some applications feature accelerometers, gyroscopes or magnetometers. Take, for instance, the recording of ball-hitting movements in tennis and badminton through a sensor [12], [13].

B. Effect of Wearable Technology

Now we know most of these applications of sports data have close relationships with the innovation of wearable technology. In fact, today's newest trend is how to integrate big data with social media. The case in point is Nike+, a platform where registered members can share their sports data. The data includes daily running paces, average speed, as well as the total running distance, scores and total calories burned over a specified period of time. As the most featured function, users may add friends through contact lists or Facebook communities, and compete in the following way. They can specify a certain distance, and see who in the circle is the fastest to achieve the goal and how fast the winner is. Nike+ also tracks where a runners goes by taking advantage of GPS, a function now available in most wearable devices and smartphones. In addition, the platform allows users to publish their routes onto social websites, urging more of their friends to compete against each other. Another example is Garmin Connect. Developed by the automobile navigation giant Garmin Corporation, the app is dedicated to users of its wearable devices, such as sports bracelets and smartwatches. In addition to the Nike+ functions mentioned above, Garmin Connect provides an overview of other users' paths on the map. By adding friends or joining communities, a user can request that his/her friends share the electronic data of the paths, and transmit the data into his/her smartwatch. This is like a virtual game where contestants compete by following the same route yet at different times.

C. Demand for Sports Data Matching

We can imagine that these virtual competitions can be really interesting if the contestants have comparable abilities athletically. Wearable devices offer real-time information about whether a wearer stays ahead or lags behind his/her contestant on the same path at a different time point. If joined by two users whose athletic abilities differ largely, however, the virtual competition can be boring. We may therefore find out what users would demand from the platform: knowing in advance whether a downloaded path was provided by a user at a comparable level. This especially holds true when a user

Manuscript received May 20, 2015; revised July 23, 2015.

K. L. Wang is with the National Taiwan Sport University, Taoyuan City, Taiwan 33301 ROC (e-mail: elverz@ms8.hinet.net).

K. C. Wang was with Fu Jen Catholic University, New Taipei City, Taiwan 24205 ROC. He is now with the Commerce Development Research Institute, Taipei City, Taiwan 10665 ROC (e-mail: kennynnek123@gmail.com).

is challenging an unknown route. Screening criteria for choosing paths available on the platform range from the average speed, previous user experience, to the height/weight, calories consumed and heart rate of both the owner and the downloader of the electronic data. This is also what is termed as “sports data matching” in this study. To our surprise, very little has been studied on the field of sports data matching. This is presumably because this matching technology has not been available until recently. Most researchers used to explore “sport rating,” a model of rating sports performances. Rating methods vary substantially because of today’s diversified sports and systems of evaluating sports performances [14]. While some of the studies were focused on the rating of score-based games [15], more of them dealt with that of paired-comparison game. The most famous example of this is Elo’s system, which is often used for chess games [16], [17]. Most of these rating models are based on one important criterion — game results. However, sports data matching considers more than this, as a contestant’s biological data or experience can be of equal importance. In this regard, some users would lay more emphasis on other facts (e.g. calories consumed) rather than the game results. In this situation, a rating-based matching model would make limited contributions. Given these facts, this study is aimed at an algorithm for sports data matching. Its results can apply to future devolvement of similar services for social sports platforms.

II. BASIC MATHEMATICAL MODEL

A. Definition and Question

This study is based on one assumption: there is a sports database that collects sports data from a number of wearable device users. One of them now wants to download the data of other users through a wearable device in order to compete with them in a virtual game in pursuit of more pleasure. Nowadays, there are many competitive sports. This study chose running as the example for easier understanding.

We $f_i(x)$ define = User (i)’s sports performance on Path (x), $x = \alpha, \beta, \gamma, \delta, \dots$; $i = 1, 2, 3, \dots, n \& A$, where User A is the virtual game contestant. Sports performance is either linear or non-linear combination of data such as time, speed, heart rate or calories consumed. Under this definition, this study will explore the following issues rather than the linear/non-linear combination, which will be discussed in the future.

Suppose that User A wants a virtual opponent (i.e. sports data) of comparable athletic abilities to compete with him (User A) through a wearable device on Path δ . Mathematically, we then need to find the most proper $f_i(\delta)$ and make it closest to $f_A(\delta)$.

B. Proof

To be more concrete, we now suppose there are four paths: α, β, γ , and δ . User A has run on Paths α, β and γ without using Path δ . To find the most proper $f_i(\delta)$ and make it closest to $f_A(\delta)$, we process as follows. We also assume that $f_i(x)$ can be standardized as numeric values 0-100.

Then, we have $f_A(\alpha), f_A(\beta), f_A(\gamma)$, and

$$f_i(\alpha) = \{f_1(\alpha), f_2(\alpha), f_3(\alpha), \dots, f_n(\alpha)\},$$

where $f_i(\alpha) \in [0, 100]$

$$f_i(\beta) = \{f_1(\beta), f_2(\beta), f_3(\beta), \dots, f_n(\beta)\},$$

where $f_i(\beta) \in [0, 100]$

$$f_i(\gamma) = \{f_1(\gamma), f_2(\gamma), f_3(\gamma), \dots, f_n(\gamma)\},$$

where $f_i(\gamma) \in [0, 100]$

Also, we have

$$f_i(\delta) = \{f_1(\delta), f_2(\delta), f_3(\delta), \dots, f_n(\delta)\},$$

where $f_i(\delta) \in (0, 100)$

Because we assume that User A did not run on Path δ , one intuitive way to derive the most proper $f_i(\delta)$ is to find at least anyone user’s sports data about running on Paths α, β or γ , and compare this data with how well User A has performed on these paths. If the two pieces of data are the closest, the owner of the piece of the data, i.e. $f(\delta)$, may probably be the $f_i(\delta)$ requested in this study. As such, we may process in following manner.

$$\|f_A(\alpha) - f_i(\alpha)\| = \{f^{1,\alpha}, f^{2,\alpha}, f^{3,\alpha}, \dots, f^{n,\alpha}\} = f^{i,\alpha}, i = 1, 2, 3, \dots, n$$

$$\|f_A(\beta) - f_i(\beta)\| = \{f^{1,\beta}, f^{2,\beta}, f^{3,\beta}, \dots, f^{n,\beta}\} = f^{i,\beta}, i = 1, 2, 3, \dots, n$$

$$\|f_A(\gamma) - f_i(\gamma)\| = \{f^{1,\gamma}, f^{2,\gamma}, f^{3,\gamma}, \dots, f^{n,\gamma}\} = f^{i,\gamma}, i = 1, 2, 3, \dots, n$$

If

$$f_i(\alpha) = 0, f_i(\beta) = 0, f_i(\gamma) = 0,$$

Then the calculation above is insignificant. Such that

$$\min\{f^{i,\alpha}, f^{i,\beta}, f^{i,\gamma}\}$$

Now we have found another runner X. Because his sports performance on any one of Paths α, β and γ is closer to User A’s, his sports data on Path δ may probably be proper. However, we need to consider one possibility: X’s sports data on Path β might be closer to User A’s, but his and User A’s performances on other paths might differ a lot, or the data might be not even available. With only one path, therefore, it could be improper to choose X’s performance on Path δ as the most proper data, hence the low reliability. To avoid this, we then employ the concept of approximation to find a more stable answer.

Suppose there is an user $X \in (1, 2, \dots, n)$, and

$$f^{X,\beta} = \min\{f^{i,\alpha}, f^{i,\beta}, f^{i,\gamma}\},$$

then we Give a $\varepsilon > 0$ and check

$$\varepsilon < \|f^{X,\beta} - f^{X,\alpha}\|, \varepsilon < \|f^{X,\beta} - f^{X,\gamma}\|$$

\Rightarrow Case1: If both

$$\varepsilon < \|f^{x,\beta} - f^{x,\alpha}\| \text{ and } \varepsilon < \|f^{x,\beta} - f^{x,\gamma}\|$$

is not held, then $f_x(\delta)$ is the proper answer.

⇒ Case2: If

$$\varepsilon < \|f^{x,\beta} - f^{x,\alpha}\| \text{ or } \varepsilon < \|f^{x,\beta} - f^{x,\gamma}\|$$

is held, then $f_x(\beta)$ is not qualified (because not reliable). So we need to find the second minimum one, and check the same thing. Continuing the same process until

$$\varepsilon < \|f^{\bar{x},\beta} - f^{\bar{x},\alpha}\| \text{ and } \varepsilon < \|f^{\bar{x},\beta} - f^{\bar{x},\gamma}\|$$

is not held, where $\bar{x} \in (1, 2, \dots, n)$

⇒ Case3: If $f^{x,\beta}$ is the minimum but $f^{x,\alpha}$ and $f^{x,\gamma}$ do not exist (because of insignificance), it means $f^{x,\beta}$ is unstable, then $f_x(\delta)$ will not be the answer and we need to find the second minimum one (back to check Case 1 and Case 2 above).

⇒ Case4: After we process Case 1- Case 3, if we can't find the proper answer (usually in low-sample situation), then we may directly pick the minimum one.

III. DISCUSSION

A. Other Choice of Approximation

According to the mathematical model above, the so-called most proper $f_i(\delta)$ is considered X's performance on Path δ , i.e. $f_x(\delta)$, under the condition that X's sports data on anyone of Paths α, β and γ is the closest to User A's, except where Case 2 or Case 3 takes place. This model can be explained as a method to pursue high validity, with low reliability as the exclusion criterion.

Another model is based on a reliability-prioritized rule. For instance, we have a ε (one that is different from the ε mentioned above). As long as $f^{i,\alpha}, f^{i,\beta}$ or $f^{i,\gamma} < \varepsilon$, the corresponding $f_i(\alpha), f_i(\beta)$ and $f_i(\gamma)$ will become potential matching options. If the person with the most paths selected as potential matching options (e.g. X, whose $f_x(\alpha), f_x(\beta)$, and $f_x(\gamma)$ are presumably the potential matching options), then $f_x(\delta)$ will be the most proper matching option for User A. If more than one person is selected, we may then consider narrowing down the value of ε , and performing the same steps to derive the best answer.

Generally speaking, if the database contains a large amount of sports data, the model described in this section can more efficiently find a proper answer. On the contrary, if there is not enough data, that in Section 2 may work better.

B. Flexible Solutions

As mentioned in Section I, the model suggested here is more flexible, as compared with the sport rating method, where rating precedes the search for the most proper match.

1) The model suggested here does not confine itself to game results. For instance, the sports

performance $f_i(\cdot)$ is commonly described as $f_i(\text{path, time, HR, Calorie, ...})$. In most cases, our goal is to find the path, with "time" as the screening criterion. But we may also refer to this model by choosing "calories burned" as the screening criterion in order to find the path.

- 2) The model can more flexibly include many indicators. Because sports performances are evaluated using many indicators, this study allows one indicator to be screened to a certain degree, and uses the second indicator for subsequent screening. More importantly, this procedure offers more flexibility in choosing a desired indicator, as compared with a mere statistical model.
- 3) This model can be applied to many sports. The only difference in sports is the function $f_i(\cdot)$ for sports performance, but we can still use the screening method and the approximation theory. Therefore, compared with some sports rating methods that can only be used for certain sports, the model proposed here is more practical.

IV. CONCLUSION

The application of health and fitness remains the most concerned topic in the world of app development since smartphone was introduced. As mentioned earlier, Nike+ and Garmin Connect are based on the monitoring of sports performance and sharing on social media with the aim of creating better user experience. Although today's products and services have the function of challenging between or among community members, they only compare numeric values. This is why their rivalry can be too weak to create in-depth user experience and loyalty.

Therefore, this study proposes a mathematical screening model to help a user find sports data from users at comparable athletic levels. The proposed model allows for sharing of sports data, making a virtual competition possible. Different from a statistical one that predicts scores, this model focuses on the search for proper data (i.e. matching) by referring to various kinds of sports data. Furthermore, the combination of approximation theory with algorithm facilitates the data matching process according to what users need from a big sports data sharing platform.

To sum up, we have employed a concrete model for runners to match their sports data. Theoretically, the model can be applied to all sports rather than just running. However, the key to the solution lies in what is important about a sport and what its contestants compete for. And can we come up with a technology that quickly collects a large amount of sports data at low prices? These issues require further discussion in the future.

REFERENCES

- [1] I. Hale, "Why Spain will win. maybe?" *Engineering & Technology*, vol. 5, no. 8, pp. 24-27, 2010.
- [2] R. P. Schumaker, O. K. Solieman, and H. Chen, *Sports Data Mining*, New York: Springer, 2010, ch. 12, pp. 109-114.
- [3] D. S. Tan, G. Smith, B. Lee, and G. G. Robertson, "AdaptiviTree: Adaptive tree visualization for tournament-style brackets,"

- Visualization and Computer Graphics, vol. 13, no. 6, pp. 1113-1120, 2007.
- [4] R. Bedini, A. Belardinelli, G. Palagi, D. Franchi, A. Ripoli, and R. Ceccarelli, "Dynamic ECG analysis in drivers during a F1 championship," *Computers in Cardiology*, pp. 645-648, 1996.
- [5] C. J. Mei, C. M. Thomas, and M. Eid, "A biofeedback interactive boxing system for optimal performance," in *Proc. Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2014, pp. 5-9.
- [6] N. H. Mahmood, N. Uyop, N. Zulkarnain, F. K. C. Harun, M. F. Kamarudin, and A. Linoby, "LED indicator for heart rate monitoring system in sport application," in *Proc. 7th International Colloquium on Signal Processing and its Applications*, 2011, pp. 64-66.
- [7] N. S. A. Zulkifli, F. K. C. Harun, and N. S. Azahar, "XBee wireless sensor networks for heart rate monitoring in sport training," in *Proc. International Conference on Biomedical Engineering*, 2012, pp. 441-444.
- [8] A. Le, T. Jaitner, and L. Litz, "Sensor-based training optimization of a cyclist group," in *Proc. the 7th International Conference on Hybrid Intelligent Systems*, 2007, pp. 265-270.
- [9] S. K. Gharghan, R. Nordin, and M. Ismail, "Design consideration of an energy efficient wireless sensor network for high performance track cycling," in *Proc. International Conference on Information Science and Applications*, 2014, pp. 1-5.
- [10] H. Noma, and T. Miyasato, "A new approach for regulating person's athletic ability by the pulse rate for realistic sports casting," in *Proc. 7th International Conference on Virtual Systems and Multimedia Proceedings*, 2001, pp. 479-488.
- [11] G. Wijnalda, S. Pauws, F. Vignoli, and H. Stuckenschmidt, "A personalized music system for motivation in sport performance," *Pervasive Computing*, vol. 4, no. 3, pp. 26-32, 2005.
- [12] D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, M. Walsh, and C. O'Mathuna, "Multi-sensor classification of tennis strokes," *Sensors IEEE*, pp. 1437-1440, 2011.
- [13] I. Pernek, G. Stiglic, and P. Kokol, "How hard am I training? Using smart phones to estimate sport activity intensity," in *Proc. the 32nd International Conference on Distributed Computing Systems Workshops*, 2012, pp. 65-68.
- [14] R. T. Stefani, "A taxonomy of sports rating systems," *IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans*, vol. 29, no. 1, pp. 116-120, 1999.
- [15] S. Samothrakakis, D. Perez, P. Rohlfshagen, and S. M. Lucas, "Predicting dominance rankings for score-based games," *IEEE Transactions on Computational Intelligence and AI in Games*, DOI 10.1109/TCIAIG.2014.2346242.
- [16] T. Cheng, D. Cui, Z. Fan, J. Zhou, and S. Lu, "A new model to forecast the results of matches based on hybrid neural networks in the soccer rating system," in *Proc. the 5th International Conference on Computational Intelligence and Multimedia Applications*, 2003, pp. 308-313.
- [17] T. Fenner, M. Levene, and G. Loizou, "A discrete evolutionary model for chess players' ratings," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 2, pp. 84-93, 2012.



K. L. Wang works as an assistant professor in Graduate Institute of International Sport Affairs, National Taiwan Sport University. His area of interest is the law issue and public policy in leisure, entertainment and sport field. Recently, he did many researches about the image right, patent, public policy and legislation regarding the sport industry based on the approach of economics of law.



K. C. Wang is a M.S. student of the Department of Mathematics at Fu Jen Catholic University. With his major in mathematics, he engaged in all fields of numerical analysis involving model constructing, programming, implementing, and simulation with various numerical methods applied. optimizing the numerical solution to problem of mathematical model in all kind of industry is his speciality. He is a researcher now, and serve in the Commerce Development Research Institute (CDRI), which is a think tank in ROC (Taiwan). He is engaged in investigating the trading and investment circumstances for Taiwan enterprises, including in developed market and emerging market, and still continuing to study the mathematical model in all fields in leisure time.