Heuristics Prediction of Olympic Medals using

Machine Learning

Heuristics Prediction of Olympic Medals using

Machine Learning

# Analysis and Prediction of Olympic Games

# 1.Abstract

# Objective

# According to the data of previous Olympic Games, including the sex, height, weight, nationality of competitors, competition items and medals, our task is to analyse and predict the number of gold medals of each country in the Tokyo Olympic Games in 2020 based on the given data.

# Motivation

# The Olympic Games are the world's sports events, and new records are broken every year. Researching the data of past years can show the changes of each Olympic Games and do the summary statistics of previous years. It is useful for countries and spectators who like to watch games to predict the future outcomes. More spectators are more willing to buy tickets to support competitions that their national teams can win.

# 2. Literature reviews

# There are number of research done related to Olympic Game and other Olympic movements. Those researches can be divided into two categories as macro level studies and micro level studies. Mainly the literature review was divided into two divisions as sport performance and Olympic performance. Under the category of sport performance, articles relating to individual sport performance were reviewed and under the category of Olympic performance articles related to Olympic successes were associated. Under the Olympic successes, there were mainly two sub categories as Olympic in 2000 and other Olympic Games. Related to the sport performance which is under the first category, 41 of previous studies done by scholars in different point of views, using different methodologies in different years to understand the factors affecting individual sport performance were also used. In the category of Olympic performance itself, 14 articles on 2000 Olympic Games were reviewed. Hence, five number of variables emerging from previous studies were used in W International Journal of Scientific and Research Publications, Those commonly emerged variables are country midyear population, GNP per capita in a particular country in the Olympic year, Social Development of the particular country (HDI), the political system of the particular country, and finally being hosting the Olympic Game. Accordingly, 05, number of studies done related to 2000 Olympic Game and Some of those important studies are as follows; The first study to analyze the factors influencing on the success at the Olympic Games appeared after the 1952 Olympic Games in Helsinki. According to Rathke and Woitek (2007) Jokl and co-authors (1956) in the study sport in the cultural patterns of the world: A study of the Olympic Games in 1952 at Helsinki were the first to use gross Domestic Product index (GDP) or GDP per capita as a potential predictor of Olympic success. Using GDP is more justifiable since it can indicate the country’s economic development and assumptions can be made about the resources for enabling athletes to be committed to sports preparation, building and maintaining training facilities, developing advanced educational system for coaches, supporting scientific research and consequently developing cutting edge training methods.

# I have studied the changes of the age of men and women and the number of participants in the Olympic Games, which may have more influence factors on the future Olympic Games. Using time series to predict the number of gold medals in the Tokyo Olympic Games in 2020 is conducive to promoting the development of national sports.

# 3. Detail of your approach

# I use the results before 2012 as variables to predict the known results in 2016, using the 2016 Olympic gold medal table as a prediction hypothesis, and finally predicting 2020 plus 2016 as variables.

# The Autoregressive Moving Average (ARMA) method models the next step in the sequence as a linear function of the observations and resiudal errors at prior time steps.It combines both Autoregression (AR) and Moving Average (MA) models.

# The notation for the model involves specifying the order for the AR(p) and MA(q) models as parameters to an ARMA function, e.g. ARMA (p, q). An ARIMA model can be used to develop AR or MA models.

# The method is suitable for univariate time series without trend and seasonal components.

# 图片包含 物体 描述已自动生成图片包含 屏幕截图 描述已自动生成图片包含 文字 描述已自动生成

# We use time series model to forecast and compare the results of 2016 Olympic Games

# 4. Dataset description

# We will explore a dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016.The data set contains 120 years of Olympic history athletes and results

# The file athlete\_events.csv contains 271116 rows and 15 columns; Each row corresponds to an individual athlete competing in an individual Olympic event (athlete-events). The columns are the following:

# ID - Unique number for each athlete;

# Name - Athlete's name;

# Sex - M or F;

# Age - Integer;

# Height - In centimeters;

# Weight - In kilograms;

# Team - Team name;

# NOC - National Olympic Committee 3-letter code;

# Games - Year and season;

# Year - Integer;

# Season - Summer or Winter;

# City - Host city;

# Sport - Sport;

# Event - Event;

# Medal - Gold, Silver, Bronze, or NA.

# 5. Experiment detail

# Distribution of the age of gold medalists.

# Women in Athletics.

# Medals per country.

# Disciplines with the greatest number of Gold Medals.

# What is the median height/weight of an Olympic medalist?

# Evolution of the Olympics over time.

# Based on the data of the top 20 countries in the Rio Olympic gold medals and using only the data of the Summer Olympic Games, we predict the number of medals in these 20 countries in the Tokyo Olympic Games in 2020. 图片包含 屏幕截图 描述已自动生成

# We can see that most of the gold medalists are in the 20-30 age range

# 图片包含 屏幕截图 描述已自动生成

# From the figure, we can see that more and more women athletes participate in the Olympic Games ,and increase year by year.

# 图片包含 文字, 照片 描述已自动生成

# The overall trend shows a growth trend, but there are three nodes declining. The first node is due to the Second World War, the second node is due to Contradictions between the Nationality of the People's Republic of China and the Republic of China, and the third is due to the Cold War boycott of the Moscow Olympic Games launched by the United States.

# 

# Women athletes participating in the Olympic Games have been showing a rapid growth trend.

# 

# The age trend of female athletes has increased a little in recent years.

# 

# It can be seen from the boxplot that the age of male athletes has not changed much.

# 6. Error analysis

# Using data from 2012 and previous years to predict the 2016 medal list is similar except for Britain, which values the influence of the Olympic Games after the London Olympic Games, encourages athletes to win gold medals, pays attention to athletes' training and adapts to the Brazilian climate. As for the prediction of 2020, the Japanese team, as the host, can adapt to the Japanese climate better, can play a better level, and the cancellation and reduction of strong sports in some countries can produce errors in the prediction results.

# 7.Conclusion

From the predict result we can see, USA will win the most gold medal. 

From result we can see if we only predict Total number of gold medals, the RMSE is 5.44, which is not bad.

REFERENCES

[1] Andreff, M., Andreff, W., (2010) Economic prediction of sport performance: from Beijing Olympics to 2010 FIFA World Cup in South Africa, International Association of Sport Economists, working paper series, paper no. 10-08, 12th conference of the International Association of sport economists 85th Western Economist Association International Conference, Poland, June 29th-July 3rd 2010.

[2] Andrew, B. B., Meghan, R. B., (2000) Who wins the Olympic Games: Economic Development and the medal total, National Bureau of Economic Development, NBER working paper series, No. 7998, JEL No. O10, L83.

[3] Anirudh, K., Eric, H., (2008) Why do some countries win more medals? Lesions for social mobility and poverty reduction, special article, Economic and political weekly, India, vol. 143.

[4] Bain. Xun., (2001) Predicting Olympic medal count: the effects of economic development on Olympic performance, the Park Place economist Volume xiii

[5] Bale, J., (2002) Geography and the Olympics: An evaluation of the work of Ernst JokI. Journal of Science and Medicine in Sport, 3 (3), 223-229.

[6] Bernard, A. B., (2008) Going for the gold: who will win the 2008 Olympic Game in Beijing, 100 Tuck Hall, Tuck School of Business, Hanover, NH 03755 [7] Bernard, A. B., (2008) Going for the Gold: Who will win the 2008 Olympic Games in Beijing, Tuck school of Business, Hanover, NH03755, Andrew. b. [Bernard@tuck.dartmouth.edu](mailto:Bernard@tuck.dartmouth.edu).

[8] Bernard, A., & Busse, M. R., (2004) Who win the Olympic Games: Economic Resources and medal total, Review of economics and statistics, 86, 413-417.

[9] Bernard, A., & Busse, M. R., (2000) Who win the Olympic Games: Economic Development and medal total (October 20, 2000) Yale school of management working paper No. ES-03; Amos Tuck school of Business working paper No. 00-02./on-line/.Retrieved January 11, 2008 from: http://ssrn.com/abstract=246937. [10] Christy, G., Daniel, G., Kristen, D., (2001) Factors Influencing Olympic performance: Interviews with Atlanta and Nagano US Olympians, Journal of Applied Sport Psychology, 13: 154-184, Association for Advancement of Applied Sport Psychology, University of North Carolina.

[11] Condom, E. M., Golden, B. L., & Wasil, E. A., (1999) Predicting the success of nations at the summer Olympics using neural network computers & operations research, 26, 1243-1265

# [12] Custonja, Z., Skoric, S., (2011) Winning medals at the Olympic Games Croatia have any chance? Kinesiology 43 (2011), 1: 107-114, UDL 796.032, 2:737.2 (497.5).

# [13] Tcha, Moonjoong, Pershin, Vitaly (2003) Reconsidering performance of the Summer Olympics and revealed Comparative advantage, Journal of Sports Economics, V.C. pp216-39

# [14] Wade, D. P., (2006) Predicting the medal wins by country at the 2006 winter Olympic Games: An econometrics Approach, National Graduate Institute for policy studies, Tokyo, Japan.

# [15] Xun, B., (2004) Predicting Olympic medal counts: the effect of economic development on Olympic performance, the park place economist, vo. XIII.

# [16] ZHANG Yunliang, CUI Degang, LIU Xueming (2008) Comparative Study on Predicting the Performance of Athletics of the 29th Olympic Games [J]; Journal of Tianjin University of Sport; 2008-05-020

# [17] http://www.olympic.org/Documents/Reference\_documents\_Factsheets/The \_Olympic\_Summer\_Games.pdf

# [18] http://www.olympic.org/content/errorpages/error404/?404;http://www.olym pic.org:80/uk/games/past/index\_uk.asp?olgt=1&olgy=1924http://www.oly mpic.org/content/error-

# [19] (PDF).http://www.aafla.org/OlympicInformationCenter/OlympicReview/19 92/ore301/ORE301p.pdf. Retrieved 2007-03-18