2014 World Cup Recap

Weichen Ning, Wen Bo

Department of ECE, Duke University

Objectives

We want to apply our machine learning knowledge to real world application of world cup.

- Be able to formulate machine learning problems corresponding to specific application.
- Understand the basic theory underlying machine learning.
- Understand a range of machine learning algorithms.
- Be able to apply machine learning algorithms to solve problems.

Introduction

We develop a Bayesian approach to predict the 2014 FIFA World Cup (especially focus on elimination games), by applying probit regression and Bayesian linear regression model with data augmentation. For our model, we develop the Gibbs sampler to perform the posterior inference. We apply our algorithms on the different datasets and the results show that we could obtain high accurate predictions.

Figure 1: Graphical Model

Materials

The following materials were required to complete the research:

- Bayesian linear regression.
- Bayesian motivation for proceduralist approach.
- Probit linear regression.
- Markov chain Monte Carlo.
- Gibbs Sampling.

The data were prepared according to the organizations outlined below:

- FIFA, OPTA, UEFA

Posterior Inference

The fitted parameters $\hat{\beta}$ should reflect the corresponding weights for each match statistic. We could employ the fitted $\hat{\beta}$ to accomplish prediction using new dataset $\tilde{X}$ as

$$\tilde{z} = \tilde{X}\hat{\beta} \tag{3}$$

A Gibbs sampler for Bayesian Linear Regression:

Sampling β from

$$(\beta|-) \sim \mathcal{N}(\beta; \hat{\beta}, \hat{\Sigma}) \tag{4}$$

where

$$\hat{\beta} = \hat{\Sigma}_0 (\Sigma^{-1}\beta_0 + X^T z)$$
$$\hat{\Sigma} = (X^T X + \Sigma_0^{-1})^{-1}$$

Sampling z_i for $i = 1, 2, \ldots, n$ from

$$(z_i|y_i = 0) \sim \mathcal{N}(z_i; X_i^T\beta, 1)\{z_i \leq 0\} \tag{5}$$

$$(z_i|y_i = 1) \sim \mathcal{N}(z_i; X_i^T\beta, 1)\{z_i > 0\} \tag{6}$$

Model

We aim to establish the relationship between match results $y = [y_1, y_2, \ldots, y_n]^T \in \mathbb{Z}^n$ and team statistics $X = [x_1, x_2, \ldots, x_p] \in \mathbb{R}^{n\times p}$ whose columns correspond to specific statistics. Since y is the binary data, we propose to use the following data augmentation approach to model the data: For $i = 1, 2, \ldots, n$,

$$z_i \sim \mathcal{N}(z_i; X_i^T\beta, 1) \tag{1}$$

where z_i is the latent variable; $y_i = 1\ if\ z_i > 0\ and\ y_i = 0\ otherwise$. Moreover, we assign the following prior on β:

$$\pi(\beta) \sim \mathcal{N}(\beta; \beta_0, \Sigma_0) \tag{2}$$

Important Result

Among the total 16 elimination matches of 2014 Brazil World Cup, we successfully predict 11 of them, obtaining the accuracy rate of about 70%.

Results

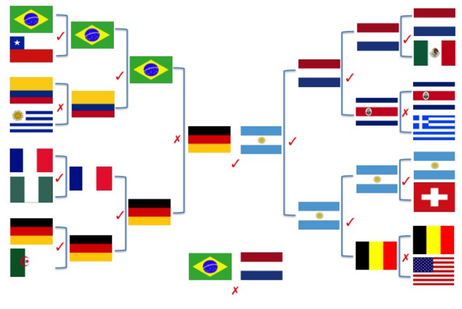

Figure 2: Prediction Result

In the game of Columbia vs Uruguay, Belgium vs USA, France vs Germany, Brazil vs Germany and Brazil vs Netherlands, we did wrong predictions according to the real data. Interestingly, some teams even win the game with the disadvantage in all fields.

Correct	Wrong	Accuracy
11	5	68.75%

Table 1: Prediction Outcome

Conclusion

This model introduces latent parameter for making prediction. In our model, we establish the relationship between match statistics in each soccer team and the result of a match. For our method, we implement the Gibbs sampling to perform the posterior inference. The experimental results tell us that the models can fit the data well and provide the reasonable prediction.

Additional Information

Further, we plan to apply some non-linear algorithms to make the prediction, like

- Non-linear SVM
- Gaussian process regression

References

[1] Alexander Spermann. *The Probit Model, University of Freiburg*. Sose, 2009.

[2] Peter D Hoff. *A first course in Bayesian statistical methods*. Springer, 2009.

Acknowledgements

We would like to express our special thanks of gratitude to our instructor (Sayan Mukherjee) who gave us the opportunity to do this interesting project of the World Cup topic, which also helped us doing a lot of research and we came to know about so much more.

Contact Information

- Web: http://www.ece.duke.edu
- Email: wen.bo@duke.edu
- Phone: +1 (919) 808 7710