

# UWDS2-Wk8-ImprovedLrmcModel-jms206

*Jim Stearns, NetID=jms206*

*Due 10 Mar 2015*

## Improving the LRMC model for NCAA Basketball Ranking

(Assignment language is in *italics*)

### Overview

*This week we will combine the new feature you evaluated previously with the LRMC model to see if the combined approach can improve the base model. Please write a summary of your approach for combining the two models as well as a statement about comparing the combined model's performance to the basic model on the 2012 and 2013 data.*

```
# mds: Mobile Dealer Sales
#mds<- read.csv("MobileCarrierHolidayAnalysis2013.csv")
```

### Combining Models

*You may combine the new information with the base model in any way you like, however, here are two approaches that you may find useful...*

#### 1. Combine final scores

*Suppose that the new factor you chose allows you to produce a ranking model independent of the basic LRMC model. For example suppose you chose RPI score, which immediately allows you to create an ordered list of all teams. Let  $x$  be a team's LRMC score and  $y$  be its RPI score. The combined score could be a weighted average of these two, e.g.,  $ax + (1-a)y$  where  $a$  is a number between 0 and 1. For numerical stability, it will also be helpful in this case to make sure that the two scores have similar magnitude (i.e., both have similar minimum and maximum values). If they do not, then rescale one of them as necessary (consider the R function "scale").*

#### 2. Modify the probability matrix

*The LRMC model works by first constructing a probability matrix (the value in row  $i$  and column  $j$  is the probability that team  $i$  is better than team  $j$ ). If you have a method for updating this probability based on your feature, then the matrix can be modified and then the regular LRMC ranking process can be applied. Again, a weighted average of the form  $ax + (1-a)y$  might be useful.*

## Evaluation of New Model

*To judge the quality of each model we will use a relatively simple measurement: the number of head-to-head predictions it correctly makes for the games played during the 2014 NCAA tournament (you might also be interested in previous years, though this requires some tedious adjustments to the part of the LRMC code that collects the data). Only data available before the beginning of the tournament may be used to make predictions. If you have used an approach similar to the weighted average described above, you can now use this measurement to search for an optimal value of the weighting parameter.*