Incorporating Pre-Game Rankings to Colley, Massey, and Elo Methods

Jiayin (Joy) Liu
Loomis Chaffee School

Abstract

Colley, Massey, and Elo are three rating methods commonly used in the NFL and Major Division I NCAA sports where game results from the regular seasons can be used to seed major playoff brackets or to create predictive rankings. In this paper, we propose two possible seeding methods, the **weak dominance graph** and the **dominance graph,** that incorporate pre-game rankings to make up for the lack of gameplays in certain sports needed to create predictive Colley, Massey, and Elo ratings. The methods will be applied in two different sports to measure their effectiveness and where it can be best utilized. The first is seeding the Official World Golf Ranking to the Dell Technologies Match Play to the round-robin play results to create rankings predictive of the final single-elimination outcomes. The second includes incorporating the preseason power ranking into the weekly NFL games to create week-by-week rankings that are predictive of the outcomes in the following game week.

Dominance and Weak Dominance Graph

Suppose an arbitrary ranking of four players A, B, C, and D, the dominance graph shows that A beats B, C, and D, while B beats C and D, and finally, C beats D, whereas the weak dominance graph shows that A beats B, B beats C, and C beats D (refer to Figure 1 and Figure 2).

Rankings	Player
1	A
2	B
3	C
4	D

Table 1

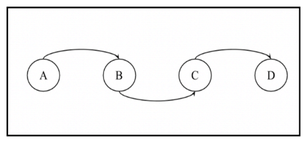

Figure 1: Weak dominance graph

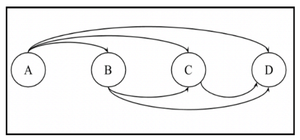

Figure 2: Dominance graph

Possible Advantages:

- generate games that are reflective of the ranking that can then be included in our dataset along with the actual games
- Ability to establish the strength of each competitor relative to others, especially when there is a lack of games played between the competitors in the field.

Rank Weighting

Standard Colley and Massey weight all games equally regardless of when the games are played, and, without modification both the pre-game ranking and actual games will be counted equally. For this experiment, we weigh the pre-game ranking portion of the dataset used for the rating algorithms as one-tenth of the actual games.

The Data – Dell Technologies Match Play

- Round Robin tournament results before the single elimination round and the Official World Ranking from the year of 2015 and 2016 are used to test the rank seeding methods.
- The results as put into a win-loss format to simulate ranks.
- Since the Massey Method takes point differences into account, we have incorporated the Official World Golf Ranking average points into the win-loss format.

The Data – NFL Week-To-Week Rankings

- The data from the regular seasons from 2010 to 2019 are used to test the proposed seeding methods.
- The rankings that are used for seeding are the corresponding years' ESPN preseason power rankings. The games are divided into 17 weeks according to the NFL calendar. The datasets are created in the format as illustrated in Figure 3.

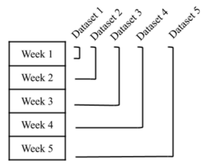

Figure 3

Results – Dell Technologies Match Play

Standard Method:

- For both 2015 and 2016, we could not produce accurate rankings using the standard Colley, Massey, and Elo methods since many competitors received the same ratings, leading to as many as 10 competitors tied in the same place.

Dominance Graph:

- For the year of 2015, the dominance graph seeding method to all three ranking methods produced rankings that are more like the Official World Ranking than the actual tournament results.
- The rank weight made little difference in the resulting ranking.

Weak Dominance Graph:

- For the year of 2015, the seeding method with Colley and Massey was able to predict the top 4s .
- For the year of 2016, the seeding method with all three methods was not able to perform better than the standard methods.

Results – NFL Week-To-Week Rankings

Predictability was used in determining whether the rating methods are successful in predicting games' out- comes in the following week. The ranking produced was used to compare with the actual outcome. If team A has a higher ranking than team B, and that team A had won the actual game in the following week, the methods have successfully predicted the outcome. The predictability was then calculated by the number of correct predictions divided by the total number of games.

Dominance Graph:

- The seeding method with Colley performed consistently better than the standard Colley for the first 2 to 3 weeks for 7 out of 10 years
- The seeding method for Elo performed better than the standard method in the first 2 to 3 weeks for most years except 2012 and 2018

Weak Dominance Graph:

- The seeding method for College made less visible difference in performance
- In 2011, 2015, and 2016, the seeding method for Colley underperformed in week 1 but performed better in week 2 and 3
- The seeding method with Elo created less predictive rankings than the standard methods in majority of the years, except 2013 and 2017.

Both the dominance graph and the weak dominance graph seeding method for Massey outperformed the standard method in the first 2 weeks for most years, since the standard method could not even produce rankings for these weeks due to the lack of gameplays and connections from teams to teams. In the year of 2016, 2018, and 2019, the dominance graph seeding methods created rankings more predictable than the standard Massey for the first 7 to 8 weeks

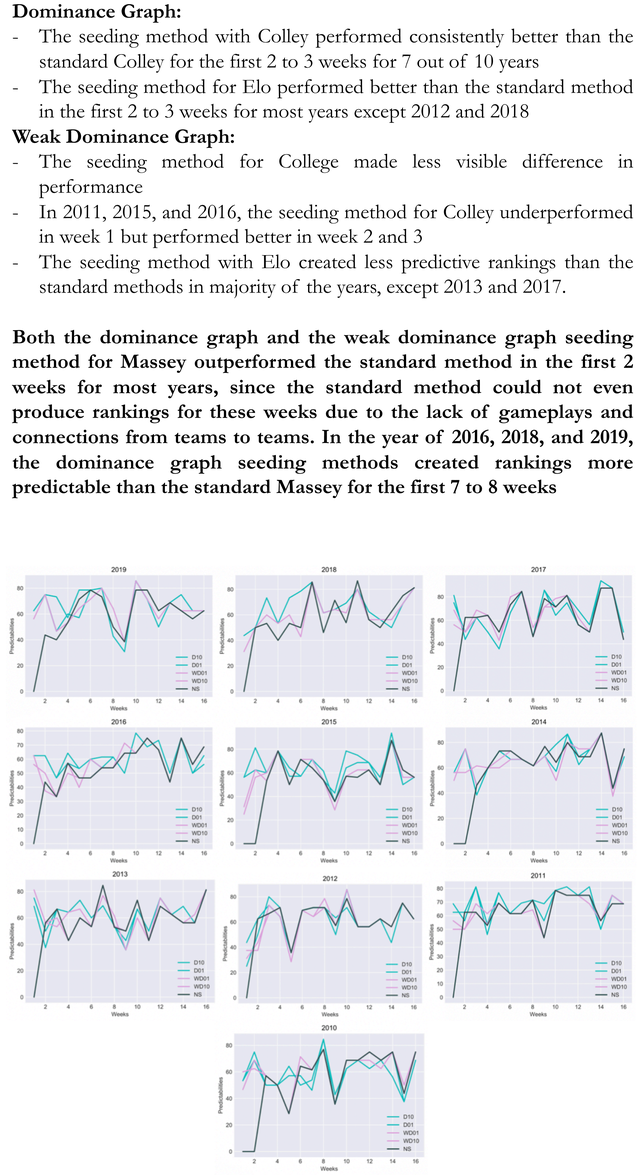

Figure 4. 2010 - 2019 NFL Week-To-Week Predictabilities Graphs - Massy

"D10"is the dominance graph seeding method with weighting as 1.0, "D01" is the dominance graph seeding method with weighting as 0.1, "WD01" is the weak dominance graph seeding method with weighting as 0.1, "WD10" is the weak dominance graph seeding method with weighting as 0.1, and "NS" is the standard method.

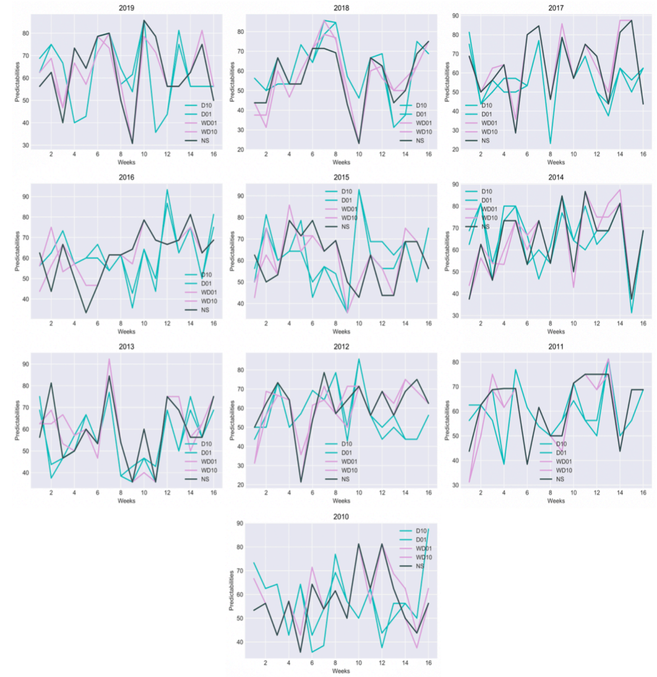

Figure 5. 2010 - 2019 NFL Week-To-Week Predictabilities Graphs - Colley

Conclusion and Future Work

- The rank seeding methods were an effective way to make up for the lack of gameplay available early into the NFL season but not for the Dell Technologies Match Play.
- The quality of the pre-game rankings can also drastically alter the resulting predictions
- The rank weights made little to no difference in the resulting ranking
- we can potentially use a mix of seeding methods and standard methods in the future.
  - We can remove the preseason ranking seedings around week 7 or 8 for the dominance graph seedings method for Massey, when the seeded methods are no longer performing better than the standard methods.
  - We can also only use the dominance graph seeding methods for the first 2 weeks and then removing them for the remaining weeks for Colley.

.

Acknowledgements

I would like to thank Dr. Tim Chartier of Davidson College for advising this work under the Pioneer Academic research seminar.