Contents

Section 1 Introduction

As the sports develop, the coach,regarded as the soul of the team, has paid more attention to.Thus,how to evaluate them becomes a necessary problem sports fans concentrate on.

The magazine Sports Illustration is looking for the best all time college coach.The performance of coach is evaluated by the coaching years, number of winning Games, winning percentage, the number of champions in top contests, the number of honor and other factors which can not be quantized,such as historical contribution, whether bringing up the star, whether creating amazing records.In addition, if a coach has brilliant performance with an ordinary team, the coach should be ranked higher.

We are asked to construct a model to choose the best college coaches in different ages, gender and sports,it can be reduced to the ranking problem.According to the problem,specifically,we need to complete four tasks as follows:

- Verify whether it is different or not if we analyze with different time line horizon.
- Verify whether our model can be applied to the different gender.
- Check whether our model can be used for all the possible sports.
- We are required to represent our model's top 5 coaches in each of 3 different sports.

We can modify our model for better performance step by step in our modeling process.In the whole modeling process, we will give full consideration to validity, feasibility and cost-efficiency of our model.

Section 2 Assumptions

During the process of buiding mathematical models for the problem, the following assumptions are needed.

- Assume that the influence of other factors that are not among the five factors can be omitted.

- Assume that ratings of the coaches are linear-correlated to ranks of the coaches.
- Assume that performance of the team is highly related to its coach.
- Assume that all the competitions refered are fair and square.
- Assume that the ranks of coaches given by the sports institutions are authentic.

Section 3 Data Preprocessing

3.1 Data Source

The source data we use in this paper mainly come from the following websites.

- Basketball source data: http://www.sports-reference.com/cbb/coaches/
- Football source data: http://www.sports-reference.com/cfb/coaches/
- Baseball source data: http://en.wikipedia.org/wiki/List_of_college_baseball_coaches_with_1,000_wins
- Basketball ranking data: http://bleacherreport.com/articles/1155912-pat-summitt-and-the-50-best-college-basketball-coaches-ever-male-or-female
- Football ranking data: http://bleacherreport.com/articles/890705-college-football-the-top-50-coaches-of-all-time/page/52
- Baseball ranking data: http://www.mademan.com/mm/10-best-college-baseball-coaches-ever.html
- Champion data: http://www.ncaa.com

3.2 Data Formalization

1. Factor Extraction

There are lots of factors that may influence the judgement of a coach and many of them are reflected in the original data. In practice, we select 5 factors here to form the record of each coach. That is, coaching years, number of winning Games, winning percentage, the number of champions in top contests, the number of honor.

Table 3.1 is parts of source data and because of limitation of space, we won't show all the data in the papar. If necessary, please check the websites above for source data.

Rk	Coach	Year	School	Yrs	Gms	Pct	Chp	Hnr
2	Mike Krzyzewski	2014	Duke	39	975	0.764	69	0
7	Roy Williams	2014	North Carolina	26	715	0.793	53	4
8	Jim Boeheim	2014	Syracuse	38	942	0.75	51	2
15	Billy Donovan	2014	Florida	20	470	0.714	28	0
20	Tom Izzo	2014	Michigan State	19	459	0.717	33	2
21	Rick Pitino	2014	Louisville	28	681	0.74	47	0
35	Bob Huggins	2014	West Virginia	29	665	0.711	45	0
40	John Calipari	2014	Kentucky	22	585	0.774	44	0
48	Bill Self	2014	Kansas	21	524	0.756	40	2
6	Jim Calhoun	2012	Connecticut	40	877	0.697	58	1
37	Gary Williams	2011	Maryland	33	668	0.637	26	0
3	Bob Knight	2008	Texas Tech	42	899	0.706	47	5
14	Eddie Sutton	2008	San Francisco	37	806	0.71	46	3
13	Lute Olson	2007	Arizona	34	776	0.731	51	0
31	John Chaney	2006	Temple	24	516	0.671	31	3
38	Billy Tubbs	2006	Texas Christian	29	609	0.658	24	0
26	Lou Henson	2005	Illinois	41	775	0.649	25	0
45	Gene Keady	2005	Purdue	27	550	0.656	26	3
11	Lefty Driesell	2003	Georgia State	41	786	0.666	35	0
25	Jim Phelan	2003	Mount St. Mary's	15	179	0.425	5	0
41	Rollie Massimino	2003	Cleveland State	28	481	0.562	20	0

Table 3.1: Original data of coaches after the year 2003

2. Data Formalization

Given the source data, we first formalize the data as follows. A feature vector x_j is created from the sample data of each coach. x_j is composed of the 5 factors mentioned above. Instead of taking the source data directly, we normalize all the factors into the range (0,1) in order to be more accurate.For example, the feature vector of college basketball coach *Mike Krzyzewski* is (0.79545,1,0.86871,0.97183,0). List of features $x = (x_1, x_2, \ldots, x_n)$ is associated with list of ratings $y = (y_1, y_2, \ldots, y_n)$ The underlying assumption here is that the rating and the rank of a coach given by the sports institution is linear-correlated. Therefore, the values of y can be got from the rank information given by some sports institutions.Moreover, the ratings y are normalized into range (0,10) for convenience.

3.3 Symbols

Symbol	Meaning
x_j	Feature vector of the coach with index j
x	List of all fecture vectors. $x = (x_1, x_2, ..., x_n)$
n	The size of x
y_j	Given rating(label) of the coach with index j
y	List of all y_j, $y = (y_1, y_2, ..., y_n)$

Table 3.2: Symbols for data formalization

Section 4 Methodology and Metrics

In this problem, we are going to build a mathematical model to choose the best college coaches of different sports. At the same time, we will discuss three additional questions about the models. That is, *does it make a difference which time line horizon is used in the analysis? How can we apply the model across sports? How can we apply the model across genders?* In the following parts, we are going to give a detailed implementation plan for each question and articulate the metrics for accessment.

4.1 Metrics for Accessment

In this part, we are going to articulate the criterion for accessment. In this problem, we adopt one general criterion, *Average Euclidean Distance.*

$$D = \frac{\sum_{j=1}^{n} |y_j - z_j|}{n} \tag{4.1}$$

where y denotes the given rating of the sports institution and z denotes the ratings given by our models. n is the size of the data.

- Set a threshold for D, denoted as D_0 and if $D < D_0$, then we can say the model is applicable to this dataset.
- When comparing two models, if $D_1 < D_2$, then we can say that *Model 1* is better than *Model 2*.

4.2 Solution Methodology

4.2.1 Discussion for time-invariance

In this part, the way to find out whether or not the model is time-invariance is as follows.

1. Using a set of data of a certain time period as the sample data to train the model and get the parameters of the models.
2. Using another set of data with different time period as testing data and apply the model with parameters in step 1 directly to this dataset.
3. Using the criterion for accessment to determine if the model in step 1 is applicable to data in step 2.
4. If so, we can conclude that the model is time-invariant; otherwise, the model is not time-invariant.

A little more explanation:

The reason why this method can check if the model is time-invariant is that, if the dataset from the different time period fits the exactly same model and there is no need for adjusting parameters, then it implies that the model itself is not correlated to time line horizon and thus the different time line horizon makes no difference to the results.

4.2.2 How to apply the model across sports

In this part, the way to find out how to apply the model across sports is as follows.

1. Using a set of data of a certain sport as the sample data to train the model and get the parameters of the models.
2. Using another set of data of a different sport as the testing data and apply the model with parameters in step 1 directly to this dataset.

3. Using the criterion for accessment to determine if the model in step 1 is applicable to data in step 2.

4. If so, we can conclude that the model is applicable across sports; otherwise, go into step 5.

5. Using some samples of dataset in step 2 (randomly selected) as the sample data to train the model and get another set of parameters of the models.

6. Then run the model on the whole dataset in step 2 and use the criterion for accessment to determine if the model is correct for this dataset.

7. If so, we can say that the model itself is applicable across sports but it requirs additional adjustments to the parameters; Otherwise, the model is not applicable across sports and even incorrect for some sports.

A little more explanation:

In this problem, The data we get online have intrinsic inconsistency problem, to be specific, the data of different sports may differ in the range and weights. For example, the only top contest mentioned in the data of college football coaches is *Super Bowl*, which is much less than those in basketball. Therefore, the range of champions and honors in fooball's data is smaller than that in basketball's data and thus the weights of influnece on the judgement of the coach are slightly different.

Based on the reality above, it's acceptable that a little adjustment to the parameters is needed when applying the model across sports and the model can still be regarded as applicable across sports.

4.2.3 How to apply the model across genders

In this part, the way to find out how to apply the model across genders is as follows. The method is similar to the method of how to apply the model across sports.

1. Using a set of data of males as the sample data to train the model and get the parameters of the models.

2. Using a set of data of female as the testing data and apply the model with parameters in step 1 directly to this dataset.

3. Using the criterion for accessment to determine if the model in step 1 is applicable to data in step 2.

4. If so, we can conclude that the model is applicable across sports; otherwise, go into step 5.
5. Using some samples of dataset in step 2 (randomly picked up) as the sample data to train the model and get another set of parameters of the models.
6. Then run the model on the whole dataset in step 2 and using the criterion for accessment to determine if the model is correct for this dataset.
7. If so, we can say that the model itself is applicable across sports but it requirs additional adjustments to the parameters; Otherwise, the model is not applicable across sports and even incorrect for some sports.

A little explanation:

Data of males and females also suffer from the similar problem as above, so adjustments to the parameters may be needed.

4.2.4 Present the top coaches

In this part, we are going to give the way to output the top 5 coaches of a certain sport.

1. Using some samples of the whole dataset as the sample data to train the model and get the parameters.
2. Applying the model on the whole dataset. The outputs of the algorithm of all the models in this paper are ratings of the coaches. So, the descending order of the ratings will reflect the ranking of coaches. Final results are the 5 coaches with highest ratings.

Section 5 Regression: Pointwise Approach

When it comes to ranking problems, in its simple form, the so-called pointwise approach is to transform the ranking problems into the regression problem. Even though the result of pointwise approach is not as good as the results of the pariwise approach and listwise approach in the following sections, pointwise approach is still worth trying for its simplicity in form and it also serves as the base stone for the later on approaches. In the following of this section, we are going to introduce two regression models to solve the problem, that is, linear regression model and non-linear regression model based on BP neural network.

5.1 Model Clarification

5.1.1 Symbols

w	Vector of the coefficients of the regression equation
F	F denotes the regression funtion of the model, either linear or non-linear
y	ratings given by sports institution
z	ratings output by our model

Table 5.1: Symbols used in Regression Model

5.1.2 Model Foundation

In this parts, we are going to use regression model to solve the problem. That is to say, given the sample data, we first try to find the coefficients of the regression function, linear or non-linear. Then, apply the regression function to the whole dataset and get the ratings of all the coaches. The descending order of ratings reflects the ranking of the coaches.

To be more specific, firstly, we formalize the sample data following the instructions in *Section 3*. Here, x_j denotes the j^{th} feature vector and y_j denotes the corresponding rating given by sports institutions.

Secondly, define regression function. Here, we use F to denote the regression function of either linear model and non-linear model. For linear model, we just use the least square method to get the regression function. For non-linear model, since we do not know the explicit form of regession function, we consider using backpropagation neural network as an equivalent substitude. The structure of neural network is shown as Figure **??**. If we choose non-linear function like sigmoid function as weight and use at least one hidden layer, the neural network can be regard as equivalent to a non-linear function. In the following models, the objective is to minimize the loss functions.

Thirdly, formalize the whole dataset and apply regression fucntion F to them. For non-linear model by BP neural network, just take the whole dataset as instances. The outputs of both models are the ratings of the coaches. The top 5 coaches will be those with highest ratings.

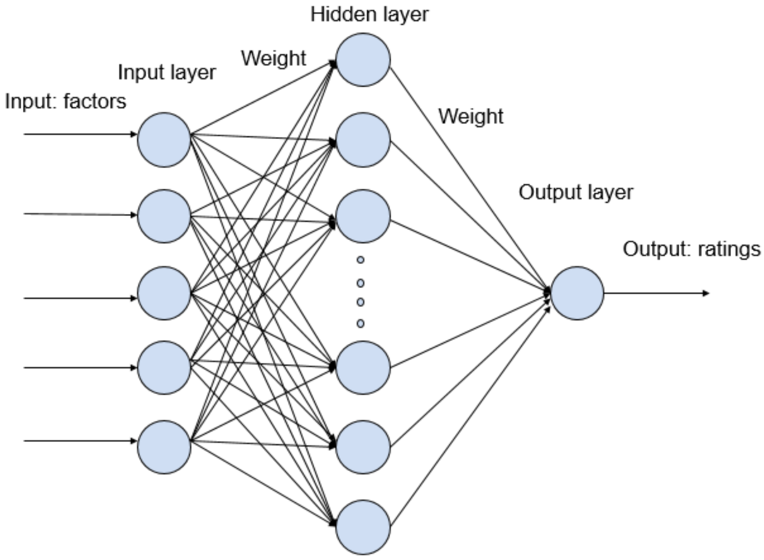


FIGURE 5.1: Curve Fitting of linear model in training process

The red one is expected curve and the blue one is the real curve.

5.2 Linear Regression Model

5.2.1 Model Implementation

We first implement the linear regression model on the data of college basketball coaches.

In the traing process, the coefficients of the regression function is shown in Table 5.2.

parameter	value
w_1	12.2872793196073
w_2	-6.61292839706653
w_3	-3.73319545690765
w_4	9.85778889758394
w_5	2.92944418563651

TABLE 5.2: coefficients of regression function

Also, we calculate the D-value of the model in training process, which is shown in Table 5.3

D	4.1000

TABLE 5.3: D-value of linear model for training process

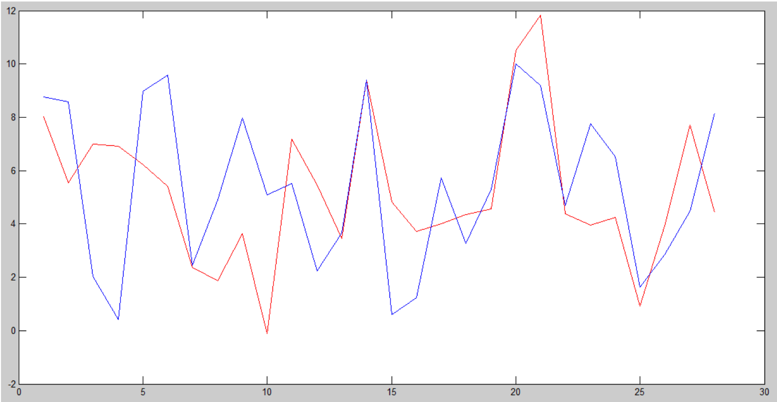


FIGURE 5.2: Curve Fitting of linear model in training process

The red one is expected curve and the blue one is the real curve.

From the Table 5.2, we can see there are two coefficients, w_2 and w_3 are negative, which make no sense in practice. Even though the curve fitting is not too bad, this linear model still needs improvements. So, in the following sections, we will introduce other improved models to solve the problem.

5.3 Non-linear Regression Model by BP Neural Network

5.3.1 Model Implementation

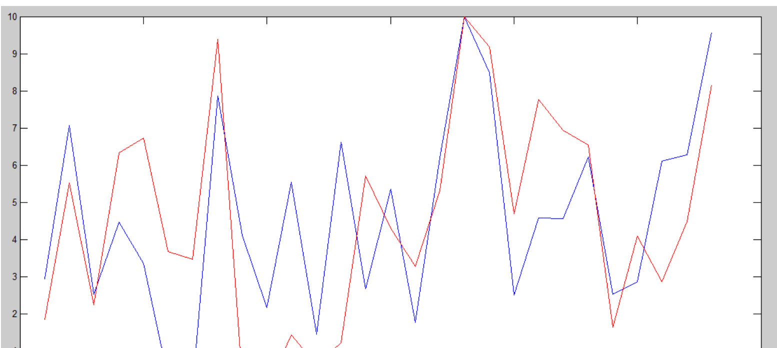
We first implement the linear regression model on the data of college basketball coaches.

In the traing process, we calculate the D-value of the model in training process, which is shown in Table 5.5

D	0.8000

TABLE 5.5: D-value of BPNN model for training process

Then, in the testing process, we compare the outputs of the model and the given ratings. Figure 5.3 shows the curve fitting. Table 5.6 shows the D-value of the model in testing process.



Analysis

From Figure 5.3, we can see that the two curves fit very well and at the same time, the D-values in the training process shown in Table 5.5 are very small. However, in the testing process, the D-values shown in Table 5.6 are relatively large. This phenomenon implies that the neural network overfits the expected curve and is not a good general model. Therefore, we are going to introduce other more effective models to solve the problem.

Section 6 Classification: Pairwise Approach

Pairwise Ranking is an approach of ranking using Machine Learning. It is similar to Pointwise Rank but need to pair the data so that to classify. On account of it, the method is efficient. We are going to describe a model based on Pairwise Ranking in this section.

6.1 Classification Model by Deep Learning

6.1.1 Introduction

The main idea of the Pairwise approach is to transform the ranking problem to the binary classification problem. In the model described later, data of coaches will be paired and classified to rank at first using the Deep Learning method, then the rank result can be conducted from it. In Pointwise method, regression is aimed at best fit in linear or nonlinear condition, that is, attaches importance on the whole trend, which may not lead to accuracy. Nevertheless, Pairwise approach is based on the details-ranking the coach list pair by pair, it is more efficient.

6.1.2 Model Clarification

6.1.2.1 Symbols

c_i	Coach i
x_i	a 5-dim factors of coach i
l_k	A Label of the kth coach pair.
$\mathscr{P}$	An implicit function represent the Deep Learning Network
R	$R =< r_1, r_2, ..., r_n >$, where r_i is the rating for coach i

Table 6.1: Symbols used in Classification model

6.1.2.2 Model Foundation

Our Classification Model by Deep Learning is a model using Deep Learning with the thought of Pairwise method. As we known, *Deep Learning* method has a good performance on classification problem, therefore, if we can convert this ranking problem into classification problem, we can use Deep Learning to solve it.

To conver this problem to classification problem, we put all the n coaches as well as their features into $n * (n-1)$ pairs $< c_i, c_j >$, and get the new paired feature $< x_i, x_j) >$, i.e.

$$X = \langle x_1, x_2, ..., x_m \rangle \Rightarrow X' = \langle (x_1, x_2), (x_1, x_3), ..., (x_{n-1}, x_n) \rangle \tag{6.1}$$

Then we use label $l_k = 1, k = 1..n * (n-1)$ to denote coach c_i is greater than coach c_j, and $l_k = 1$ otherwise. Next we train the Deep Learning Network with feature pairs $< x_i, x_j >$ as input and label l_k as expected outpt.

After the training, the Deep Learning Network can be regarded as an implicit function

$$\mathscr{P}(< x_i, x_j >) \in (0, 1), \tag{6.2}$$

which representing the probability that coach c_i is greater than coach c_j.

Then, for the new testing data, we firstly convert them to paired data similary, and put the paired feature as input into function $\mathscr{P}$, i.e. the Deep Learning Network, to get the probability. With the probability, we can construct the rating $R =< r_1, r_2, ..., r_n >$ for each coach by

$$r_i \leftarrow r_i + \mathscr{P}(< x_i, x_j >) \tag{6.3}$$

$$r_j \leftarrow r_j + (1 - \mathscr{P}(< x_i, x_j >)) \tag{6.4}$$

where $R = \mathbf{0}$ initially.

Finally, we can get the ranking sequence by sorting the rating R.

6.1.3 Solution and Result

6.1.3.1 Algorithm Implementation

The structure of Deep Learning Network is similar to the structure of BP Network, but the training process is different. The first step of Deep Learning is doing unsupervised learning without labeled outputs, which is by using Restricted Boltzmann Machines. Then it uses labeled outputs to do the supervised learning. Here we use the Matlab toolbox *DeepLearnToolbox* from Github and construct the Deep Learning Network as following.

```
dbn.sizes = [64 32];
opts.numepochs =   10;
opts.batchsize =   1;
opts.momentum  =   0;
opts.alpha     =   0.01;
dbn = dbnsetup(dbn, coFeature, opts);
dbn = dbntrain(dbn, coFeature, opts);
```

Because this is a classification problem, and we require the output to be limited in 0 to 1, we choose the sigmoid function as active function and output function to train the network.

```
nn = dbnunfoldtonn(dbn, 1);
nn.activation_function = 'sigm';
opts.numepochs =  100;
opts.batchsize = 1;
nn = nntrain(nn, coFeature, coLabel, opts);
```

6.1.3.2 Verification for time-invariance

In this section, we are going to verify that our model is time-invariance in this problem following the instrucions in *Section 4*

1. Training process

Table 3.1 shows the original data for the training process. They are data of college basketabll coaches after the year 2003.

Figure 6.1 shows the curve fitting in the training process. We can see from the figure that the two curves are nicely fitted and the trends are almost the same. To be more

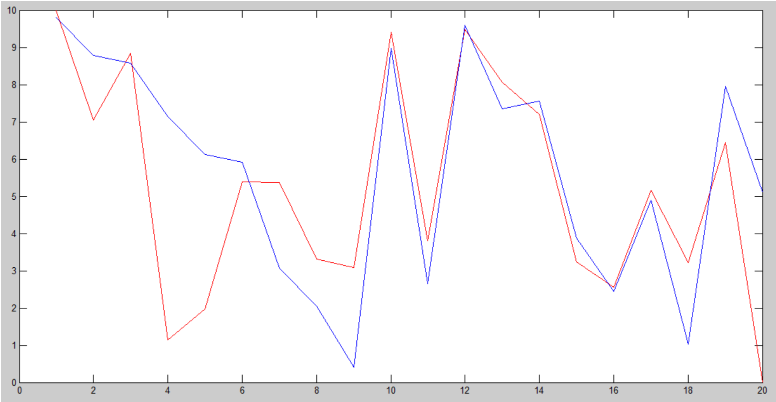


FIGURE 6.1: Curve Fitting of Classification model in training process

The red one is expected curve and the blue one is the real curve.

accurate, based on the criterion in *Section 4*, using Equation 4.1, we can calculate the D-value as shown in Table 6.2.

D	1.9000

TABLE 6.2: D-value of Classification model for training process

2. Testing process

In testing process, we use the data of coaches before 2003. Since the whole dataset is too large, we will only take part of the samples here.

Figure 6.2 shows the curve fitting in the testing process. We can see from the figure that the two curves are nicely fitted and the trends are almost the same.

To be more specfic, we calculate the D-value of the testing process as shown in Table 6.3.

D	4.3000

TABLE 6.3: D-value of Classification model for testing process

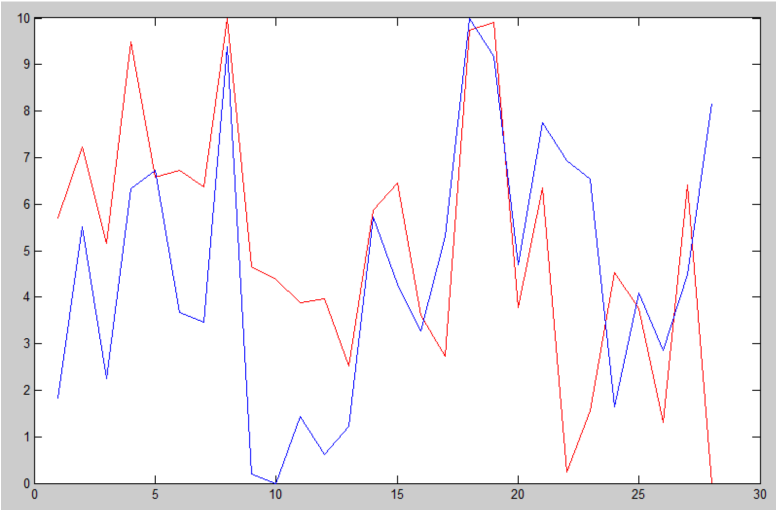


FIGURE 6.2: Curve Fitting of Classification model in testing process

The red one is expected curve and the blue one is the real curve.

As we can see, the D-value in testing process is small in an acceptable level and the curve fitting is good. Therefore, we can conclude that our model is time-invaraince. That is to say, the time line horizon used in the analysis makes little difference to the results.

6.1.3.3 Discussion for application across sports

In this section, we are going to discuss how our model can be applied across sports following the instructions in *Section 4*.

Firstly, we apply the model with parameters from basketball data to football data. As we can see in Figure 6.3, fitting of the two curves is not very well.

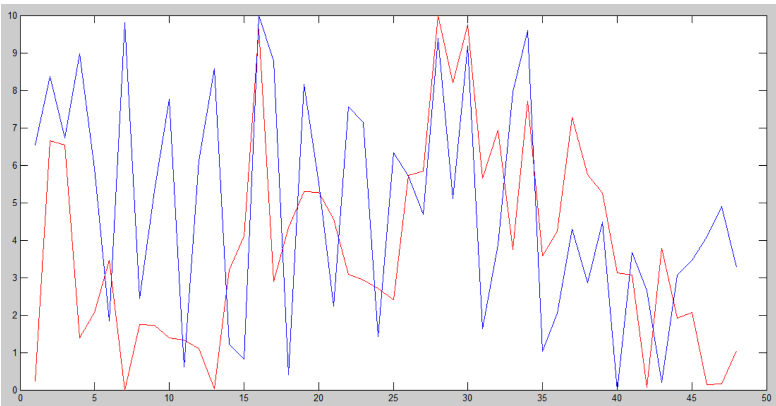


FIGURE 6.3: Curve Fitting directly of Classification model across sports

The red one is expected curve and the blue one is the real curve.

However, if we train some samples of college footabll coaches, the parameters of the model will slightly change.

Figure 6.4 shows the curves fit well in an acceptable level. To be more accurate, Table 6.4 shows the D-value of football data.

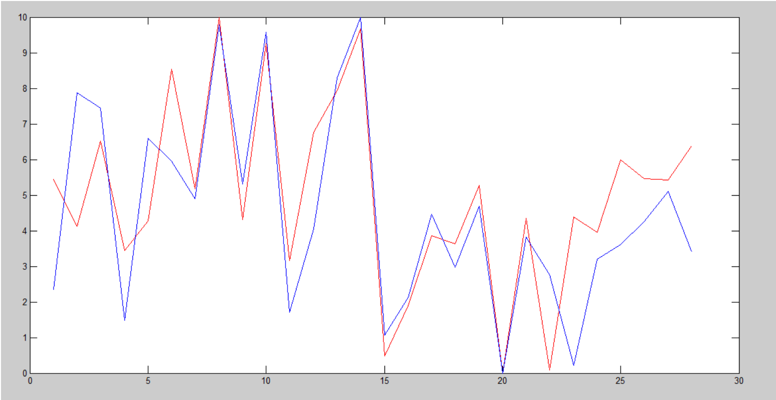


FIGURE 6.4: Curve Fitting adjusting parameters of Classification model

The red one is expected curve and the blue one is the real curve.

D	4.3571

TABLE 6.4: D-value of Classification model of football data

Analysis on the results is as follows.

As we can see clearly from the Figure 6.3 and Figure 6.4 that if we apply the model with parameters from data of college basketball coaches to the data of college football coaches, then the result is not good. However, if we adjust the parameters of the model, the result turns out to be very good. Therefore, as the instructions in *Section 4*, our model can be applied across sports with an extra step of adjusting parameters.

6.1.3.4 Discussion for application across genders

In this part, we are going to discuss how our model can be applied across genders.

Firstly, we apply the model with the parameters of male college basketball coaches to the data of female college basketball coaches. Figure 6.5 shows the curves fitting.

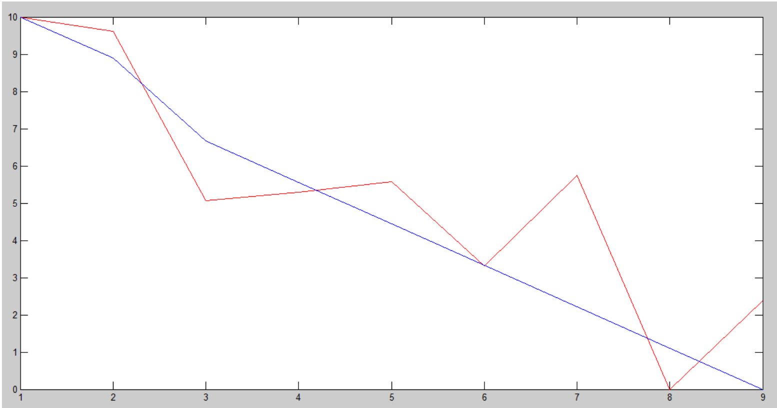


FIGURE 6.5: Curve Fitting of Classification model across genders

The red one is expected curve and the blue one is the real curve.

D	1.3333

TABLE 6.5: D-value of Classification model across gender

As we can see, the curves in Figure 6.5 fits well and the D-value is small, we can safely conclude that our model can be applied across genders directly without any changes.

6.1.3.5 Top 5 coaches for three sports

In this part, we are going to present our final results of top 5 coaches of 3 different sports following the instructions in *Section 4*.

1. Basketball
Firstly, we run sample data of basketball coaches to build up the Neural Network. After that, we formalize the whole dataset of basketball data and get ratings for the total 3000 records. Since the data is too large, we only show top 10 records in Table **??** here.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Mike Krzyzewski	39	975	0.764	69	0	10
2	John Wooden	29	664	0.804	54	12	9.902374
3	Dean Smith	36	879	0.776	70	1	9.889975
4	Jim Calhoun	40	877	0.697	58	1	9.801151
5	Bob Knight	42	899	0.706	47	5	9.762024
6	Adolph Rupp	41	876	0.822	71	1	9.732853
7	Jim Boeheim	38	942	0.75	51	2	8.844616
8	Eddie Sutton	37	806	0.71	46	3	8.66901
9	Denny Crum	30	675	0.696	57	0	8.478481
10	Lute Olson	34	776	0.731	51	0	7.924273

Mike Krzyzewski	John Wooden	Dean Smith	Jim Calhoun	Bob Knight

Table 6.7: Top 5 college basketball coaches of classification model

2. Football

Firstly, we run sample data of football coaches to build up the Neural Network. After that, we formalize the whole dataset of football data and get ratings for the total 2054 records. Since the data is too large, we only show top 10 records in Table 6.8.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Bear Bryant	38	323	0.78	15	3	10
2	Tom Osborne	25	255	0.836	12	1	8.927731
3	Pop Warner	42	311	0.733	1	0	8.580714
4	Fielding Yost	25	165	0.833	1	0	8.527589
5	Bo Schembechler	27	234	0.775	5	1	8.267956
6	Knute Rockne	13	105	0.881	1	1	8.240703
7	Amos Alonzo Stagg	42	275	0.681	0	1	8.178191
8	Bob Neyland	21	173	0.829	2	0	7.989987
9	Joe Paterno	46	409	0.749	24	1	7.957188
10	Frank Leahy	13	107	0.864	1	1	7.877672

Table 6.8: Results of top 10 college football coaches of classification model

Therefore, the top 5 college football coaches in this model is shown in Table 6.9

Bear Bryant	Tom Osborne	Pop Warner	Fielding Yost	Bo Schembechlerr

Table 6.9: Top 5 college football coaches of Classification model

3. Baseball

Firstly, we run sample data of baseball coaches to build up the Neural Network. After that, we formalize the whole dataset of baseball data and get ratings for the total 73 records. We only show top 10 records in Table 6.10 here.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Rod Dedeaux	44	1342	0.691	10	1	10
2	Gordie Gillespie	59	1893	0.665	5	1	9.81793
3	Don Schaly	40	1442	0.812	0	1	9.09571
4	Cliff Gustafson	29	1427	0.792	2	1	8.775067
5	Augie Garrido	45	1874	0.682	5	0	8.773391
6	Skip Bertman	18	870	0.724	5	1	8.539533
7	Ron Fraser	30	1271	0.742	2	1	8.339319
8	Ed Cheff	34	1705	0.799	0	1	8.28311
9	Frank Vieira	44	1127	0.776	0	0	7.95974
10	Bobby Winkles	13	524	0.752	3	1	7.469099

Table 6.10: Results of top 10 college baseball coaches of classification model.

Therefore, the top 5 college baseball coaches in this model is shown in Table 6.11

Rod Dedeaux	Gordie Gillespie	Don Schaly	Cliff Gustafson	Augie Garrido

Table 6.11: Top 5 college baseball coaches of Classification model

Section 7 Optimization: Listwise Approach

Listwise Ranking is also a methodology of learning to rank. The main difference between the pairwise approach and listwise approach is that input of listwise approach is a list of objects while input of pairwise approach is object pairs. Though pairwise approach offers advantages, it ignores the fact that ranking is a prediction task on list of objects. Therefore, in this section, we are going to introduce another model based on listwise approach.

7.1 ListNet Model

7.1.1 Introduction

In former sections, we talk about the pointwise approach and the pairwise approach and they do have obvious advantages when applied under current scenario. For pointwise approach, the complexity of the method is really low and it's easy to be carried out. For pairwise approach, firstly, existing classification methodologies can be directly applied. Secondly, the instances of the training set are easy to obtain. However, there are also problems with these approaches. The biggest problem of the pointwise approach is the accuracy. For pairwise approach, the intrinsic problem is that the objective of learning in pairwise approach is formalized as minimizing errors in classification of instance pairs rather than minimizing errors in ranking of instances. Moreover, the training process is computationally costly, since the number of paired instances is very large.

Since there are some unavoidable drawbacks in the former approaches, in this section, we are going to apply the listwise approach to the current scenario. That is to say, we take a list of objects as input and by optimizing the loss function; we can significantly improve the results. The detailed clarification is in the following sections.

7.1.2 Model Clarification

7.1.2.1 Symbols

Symbol	Meaning
$f(x)$	Ranking function given feature vector set
$f(x_j)$	Score of the coach with index j given by Ranking function $f(x)$
$f_w(x)$	Ranking function on parameter w
$f_w(x_j)$	Score of the coach with index j given by Ranking function $f_w(x)$
z	$z = (f(x_1), f(x_2), ..., f(x_n))$, list of scores
$P_y(\pi)$	Permutation probability
$L(y, f(x))$	Loss function
Γ	$\Gamma = \{(x_j, y_j)\}\|_1^n$, the training set

Table 7.1: Symbols used in ListNet model

7.1.2.2 Model Foundation

In this model, the main methodology comes from the area of learning to rank. Follow the tradition of machine learning, we train some data of the coaches whose rank can be found by some institutions online to get the parameters and then apply it to the whole dataset, both rank known and unknown.

To be more specific, firstly, we need to formalize the data. Follow the instructions in *Section 3*, we get the training set $\Gamma = \{(x_j, y_j)\}|_1^n$, where x_j is the j^{th} feature vector and y_j is the corresponding rating.

Secondly, we create a ranking function f with parameter w, where w is a 5-dimentional vector. In this problem, we use the linear function, that is,

$$f =< w \cdot x_j > \tag{7.1}$$

where $< \cdot >$ denotes the inner product of the two vectors. For each feature vector x_j, it outputs a score $f(x_j)$. The scores of each coach will be used for ranking the coaches. That is to say, the descending order of scores will be the ranking of coaches given by our model. For the list of feature vector x, we obtain a list of score, $z = (f(x_1), f(x_2), ..., f(x_n))$.

The objective of learning is formalized as minimization of the total losses with respect to the training data.

$$\sum_{j=1}^{n} L = (y, f(x_j)) \tag{7.2}$$

where L is the loss function which represents the distance of the two score lists. Theoretically, any metrics can be used for the representation and in this problem, we choose *Cross Entropy* as metric, and the loss function L becomes

$$L(y, z) = -\sum_{j=1}^{n} P_y(j) log(P_z(j)) \tag{7.3}$$

where $P_s(j)$, $s \in \{y, z\}$ is the top one probability of instance j, which is introduced for the purpose of representing the likelihood of a permutation(a possible order of a ranking list) given the ranking function. In this problem, we use top one probability for each instance and it is defined as

$$P_y(j) = \frac{exp(y_j)}{\sum_{k=1}^{n} exp(y_k)} \tag{7.4}$$

$$P_z(j) = \frac{exp(f(x_j))}{\sum_{k=1}^{n} exp(f(x_k))} \tag{7.5}$$

Thirdly, we are going to implement the gradient descent method to achieve the object of minimizing the total loss. Here, we need to calculate the gradient of $L(y, z(f_w))$ with respect to the parameter w as follows.

$$\begin{aligned}\Delta w &= \frac{\partial L(y, z(f_w))}{\partial w} \\ &= -\sum_{j=1}^{n} P_y(j)\frac{\partial f_w(x_j)}{\partial w} + \frac{1}{\sum_{k=1}^{n} exp(f_w(x_k))}\sum_{j=1}^{n} exp(f_w(x_j))\frac{\partial f_w(x_j)}{\partial w} \\ &= \sum_{j=1}^{n}\frac{\partial f_w(x_j)}{\partial w}[P_{z(f_w)}(j) - P_y(j))]\end{aligned} \tag{7.6}$$

After that we iterate the updating process for $1,000,000$ times with updating equation as

$$w = w - \eta \cdot \Delta w \tag{7.7}$$

Then we output the value of parameter w and f_w is the final ranking function.

At last, we formalize the whole dataset as the list of feature vector, and using the ranking function, we get the score of each coach in the database and the descending order of the score list will be the ranking of coaches.

7.1.2.3 Algorithm

The core part of this model is the calculation of *Loss Function* and the implementation of gradient decent. Here is our algorithm implemented by Matlab.

```
% T: iteration times
% alpha: learning rate
% err: target error
% x: factors
% w: coeficients
% loss: value of loss function
w = ones(dim,1);
for t = 1 : T
    z = x * w;
    ez = exp(z);
    ey = exp(y);
    sumez = sum(ez);
    sumey = sum(ey);
    c = - sumey .* log(sumez);
```

```
    loss(t) = sum(c);
    dw = ez / sumez - ey / sumey;
    dw = (dw' * x)' / samples;
    if norm(dw) < goal
        disp('target error achieved');
        break;
    end
    w = w - dw * alpha;
end
```

7.1.3 Solution and Result

7.1.3.1 Verification for time-invariance

In this section, we are going to verify that our model is time-invariance in this problem following the instrucions in *Section 4*

1. Training process

We adopt the gradient descent method here. Figure 7.1 shows the curves of loss fucntion $L(y, z)$ and we can see as the increase of iteration times, the total losses are approaching the minimum value. Figure 7.2 shows the curve of the gradient Δw and similarly, as the increase of iteration times, the gradient is approaching 0. These two figures will

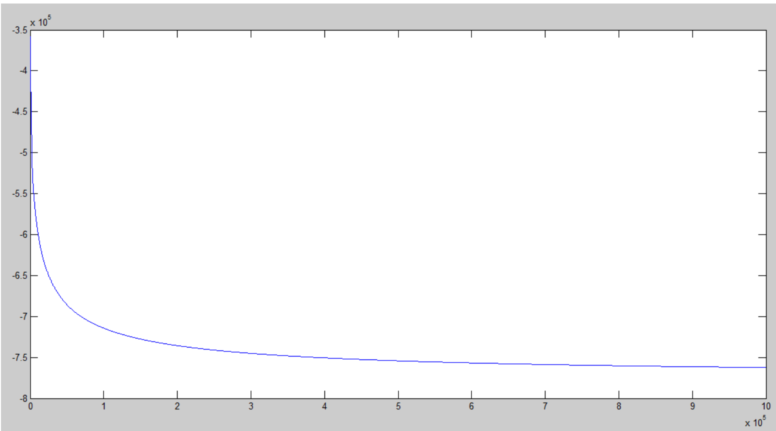


FIGURE 7.1: Curve of loss function

x axis is the iteration times.

show that our program indeed achieves the objective of minimization of total losses.

Table 3.1 shows the original data for the training process. They are data of college basketabll coaches after the year 2003. Table 7.2 shows the coefficients of the ranking

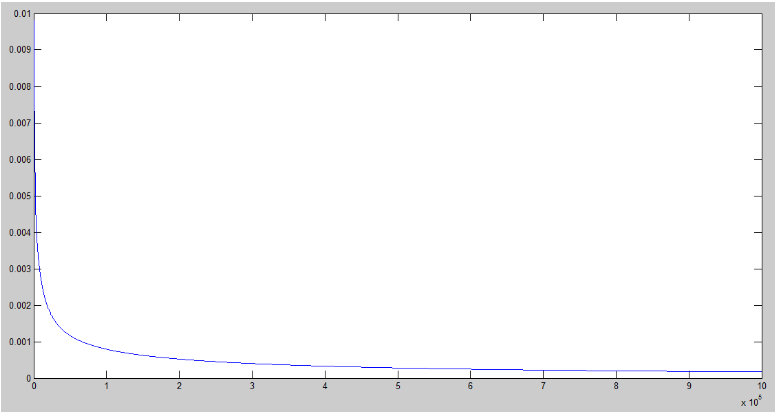


FIGURE 7.2: Curve of the gradient Δw

x axis is the iteration times.

function.w_1,w_2,w_3,w_4,w_5 are the coefficients of the 5 factors, *coaching years, number of winning Games, winning percentage, number of champions in top contests and the number of honors.*

parameter	value
w_1	2.55852534980252
w_2	2.23057544283441
w_3	2.57489617274287
w_4	4.97394937104007
w_5	3.47586804716193

TABLE 7.2: coefficients of the ranking function from data after 2003

Figure 7.3 shows the curve fitting in the training process. We can see from the figure that the two curves are nicely fitted and the trends are almost the same. To be more

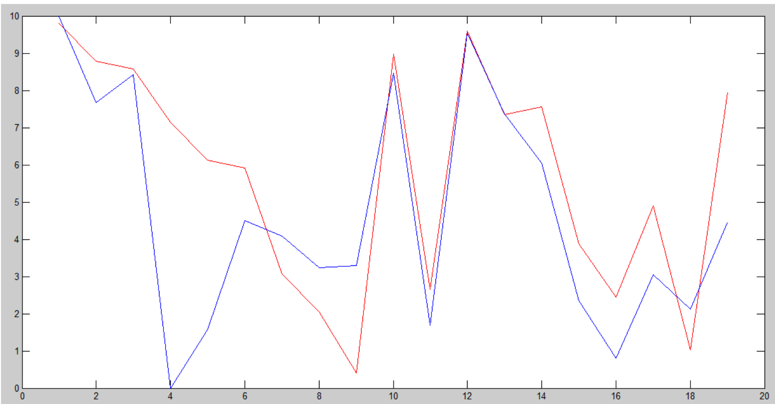


FIGURE 7.3: Curve Fitting of ListNet model in training process

The red one is expected curve and the blue one is the real curve.

accurate, based on the criterion in *Section 4*, using Equation 4.1, we can calculate the

2. Testing process

In testing process, we use the data of coaches before 2003. Since the whole dataset is too large, we will only take part of the samples here.

Figure 7.4 shows the curve fitting in the testing process. We can see from the figure that the two curves are nicely fitted and the trends are almost the same.

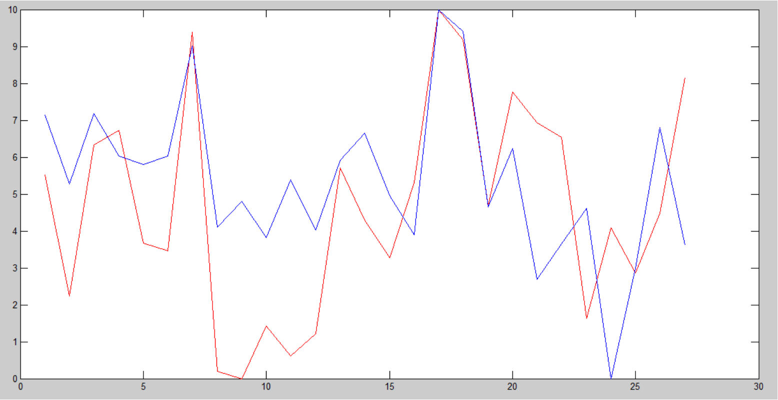


FIGURE 7.4: Curve Fitting of Listin testing process

The red one is expected curve and the blue one is the real curve.

To be more specfic, Table 7.4 shows the comparison of the given rank and the output rank of coaches before 2003. As we can see, each pair of rank is merely different in an acceptable level. Similarly, we calculate the D-value of the testing process as shown in Table 7.5.

Coach	Year	Given Rank	Output Rank	Ratings
John Wooden	1975	1	1	10
Frank McGuire	1980	34	30	6.1775045
Guy Lewis	1986	22	24	6.766940794
Jud Heathcote	1995	43	44	4.670083306
Dean Smith	1997	4	5	9.270363906
Don Haskins	1999	17	20	7.075891791

TABLE 7.4: Comparison of given rank and the output rank

D	4.2000

TABLE 7.5: D-value of ListNet model for testing process

Therefore, we can conclude that our model is time-invaraince. That is to say, the time line horizon used in the analysis makes little difference to the results.

7.1.3.2 Discussion for application across sports

In this section, we are going to discuss how our model can be applied across sports following the instructions in *Section 4*.

Firstly, we apply the model with parameters from basketball data to football data. As we can see in Figure 7.5, fitting of the two curves is not very well.

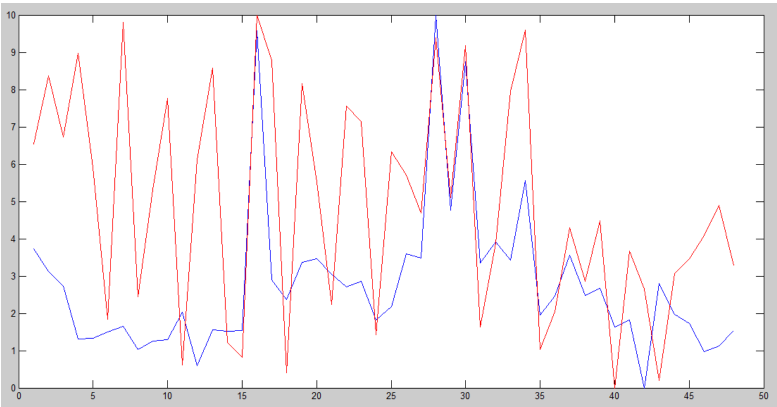


FIGURE 7.5: Curve Fitting directly across sports

The red one is expected curve and the blue one is the real curve.

However, if we train some samples of college footabll coaches, the coefficients will slightly change. Table 7.6 shows the adjusted coeffients.

parameter	value
w_1	1.91954141535799
w_2	1.58943349986935
w_3	5.17619727640685
w_4	0.353523905773156
w_5	0.675776883763109

TABLE 7.6: coefficients of ranking function from data of college football coaches

Figure 7.6 shows the curves fit well in an acceptable level. To be more accurate, Table 7.7 shows the D-value of football data.

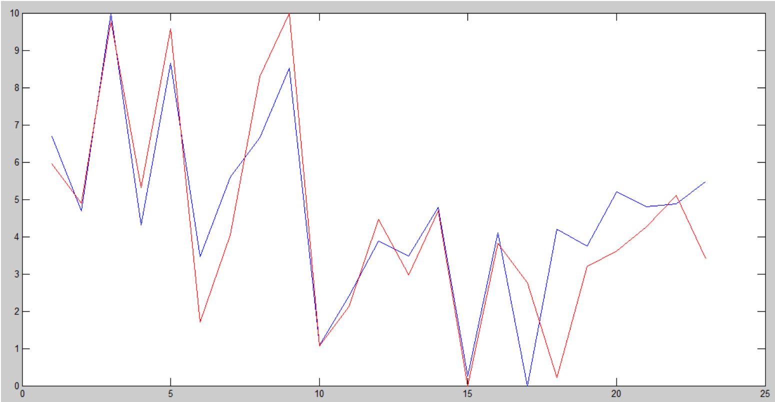


FIGURE 7.6: Curve Fitting adjusting coeffients

The red one is expected curve and the blue one is the real curve.

D	2.4000

TABLE 7.7: D-value of football data

Analysis on the results is as follows.

- Why the coeffients of different sports are different?
  We check the original data and find that on average the weights of *Champions* and *Honors* are significantly larger in basketball than those in football. This is because the only top contest recorded in the data of college football coaches is *Super Bowl* while in basketball records coming from several top contests are included. The intrinsic difference in the original data leads to the slightly different coefficients across sports.

- How can we apply our model in different sports?
  As Figure 7.4 and Figure 7.6 show, the model itself is applicable accross sports. However, since the data of different sports we get online have some intrinsic difference, the coefficients of different sports differ from each other slightly. Therefore, our model can be successfully apply to different sports with an extra step of adjusting coefficients. Moreover, if we can get more accurate data and avoid the problem above, our model may be able to be applied accross sports directly.

7.1.3.3 Discussion for application across genders

In this part, we are going to discuss how our model can be applied across genders.

Here, we apply the model with parameters of data of college male basketball coaches to the data of female basketball coaches. Figure 7.7 shows the curve fitting of the given rating and the output rating. As we can see, the trends of the two curves are almost the same, which implies that the model can be directly applied across genders without any changes. To be more specific, we calculate the D-value of the result, which is shown in

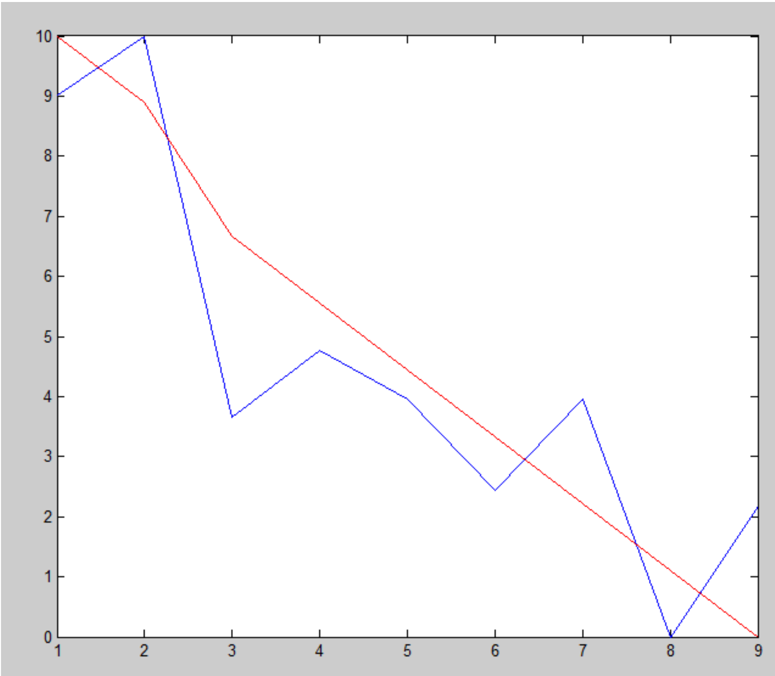


FIGURE 7.7: Curve Fitting across genders

The red one is expected curve and the blue one is the real curve.

Table 7.8

D_{10}	1.3333

TABLE 7.8: top10 D-value of female data

Therefore, based on the results above, we can say that our model can be directly applied across genders without any changes.

7.1.3.4 Top 5 coaches for three sports

In this part, we are going to present our final results of top 5 coaches of 3 different sports following the instructions in *Section 4*.

1. Basketball

The coefficients of the ranking function is shown in Table 7.9

parameter	value
w_1	2.55852534980252
w_2	2.23057544283441
w_3	2.57489617274287
w_4	4.97394937104007
w_5	3.47586804716193

TABLE 7.9: coefficients of the ranking function of basketball data

After that, we formalize the whole dataset of basketball data and get ratings for the total 3000 records. Since the data is too large, we only show top 10 records in Table 7.10 here and the top 60 records will be in the *Appendix*.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	John Wooden	29	826	0.804	54	12	10
2	Adolph Rupp	41	1066	0.822	71	1	9.358209
3	Bob Knight	42	1273	0.706	47	5	9.267523
4	Mike Krzyzewski	39	1277	0.764	69	0	9.266344
5	Dean Smith	36	1133	0.776	70	1	9.157346
6	Jim Calhoun	40	1259	0.697	58	1	8.787727
7	Jim Boeheim	38	1256	0.75	51	2	8.732497
8	Eddie Sutton	37	1135	0.71	46	3	8.322785
9	Roy Williams	26	902	0.793	53	4	8.212995
10	Lute Olson	34	1061	0.731	51	0	7.69237

TABLE 7.10: Results of top 10 college basketball coaches.

Therefore, the top 5 college basketball coaches in this model is shown in Table 9.2

John Wooden	Adolph Rupp	Bob Knight	Mike Krzyzewski	Dean Smith

Table 7.11: Top 5 college basketball coaches

2. Football

The coefficients of the ranking function is shown in Table 7.12

parameter	value
w_1	1.91954141535799
w_2	1.58943349986935
w_3	5.17619727640685
w_4	0.353523905773156
w_5	0.675776883763109

Table 7.12: coefficients of the ranking function of football data

After that, we formalize the whole dataset of football data and get ratings for the total 2054 records. Since the data is too large, we only show top 10 records in Table 7.13 here and the top 60 records will be in the *Appendix*.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Bear Bryant	38	425	0.78	15	3	10
2	Joe Paterno	46	548	0.749	24	1	9.764865
3	Bobby Bowden	40	485	0.74	22	1	9.142352
4	Amos Alonzo Stagg	42	425	0.681	0	1	8.808064
5	Pop Warner	42	446	0.733	1	0	8.773349
6	Tom Osborne	25	307	0.836	12	1	8.329462
7	Walter Camp	5	70	0.971	0	3	8.15314
8	Dan McGugin	30	271	0.762	0	1	8.11698
9	Bo Schembechler	27	307	0.775	5	1	8.088726
10	LaVell Edwards	29	361	0.716	7	1	7.976948

Table 7.13: Results of top 10 college football coaches.

Therefore, the top 5 college football coaches in this model is shown in Table 9.3

Bear Bryant	Joe Paterno	Bobby Bowden	Amos Alonzo Stagg	Pop Warner

Table 7.14: Top 5 college football coaches

3. Baseball

The coefficients of the ranking function is shown in Table 7.15

parameter	value
w_1	5.27074667521336
w_2	1.70611910382732
w_3	2.18164605915199
w_4	1.97022696549876
w_5	3.01940918525854

Table 7.15: coefficients of the ranking function of baseball data

After that, we formalize the whole dataset of baseball data and get ratings for the total 73 records. We only show top 10 records in Table 7.10 here and the top 60 records will be in the *Appendix*.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Gordie Gillespie	59	1893	0.665	5	1	10
2	Rod Dedeaux	44	1342	0.691	10	1	8.920852
3	Don Schaly	40	1442	0.812	0	1	7.440755
4	Ed Cheff	34	1705	0.799	0	1	6.923967
5	Cliff Gustafson	29	1427	0.792	2	1	6.420532
6	Ron Fraser	30	1271	0.742	2	1	6.009221
7	Bill Wilhelm	36	1161	0.683	0	1	5.748607
8	Ron Polk	35	1373	0.662	0	1	5.694265
9	Chuck Brayton	33	1158	0.687	0	1	5.431216
10	Augie Garrido	45	1874	0.682	5	0	5.416128

Table 7.16: Results of top 10 college baseball coaches.

Therefore, the top 5 college baseball coaches in this model is shown in Table 9.4

Gordie Gillespie	Rod Dedeaux	Don Schaly	Ed Cheff	Cliff Gustafson

Table 7.17: Top 5 college baseball coaches

Section 8 Strength and Weakness

8.1 Strength

- In the linear regression model, it is easy and quick to implement the algorithm. As the most common regression model, linear regression is always integrated into a function and can be easily used.
- In the non-linear regression model, by using the BackPropagation Neural Network, it can fit the training data better than many other approaches in a non-linear way.
- In the classification model, due to its *pairwise* algorithm - making each two coaches into a pair, it boost the quantity of training data, and therefore can perform well even if the training data is limited.
- In the optimization model, by using a more accurate loss function, we can obtain a more accurate metric for this ranking problem, and therefore get a better permutation result.

8.2 Weakness

- In the linear regression model, it cannot fit the data well, for the ranking is not linearly related to the factors.

- In the non-linear regression model, if the data are not enough, the neural network could easily over-fit the training data, lacking the generalization ability.
- In the classification model, it is hard to set the appropriate parameters for the Deep Brief Network, which has a markedly influence on the final results.
- In the optimization model, compared to other models, it will take a longer time for the loss function to convegent to its minimun value.

Section 9 Conclusion

According to the previous implementation results, we can list the permutation error (Average Euclid Distence) of each approach as Table 9.1.

Error	Linear Regression	BP Network	Deap Learning	ListNet
Training D	4.1	0.8	1.9	1.3
Testing D	7	6.2	4.3	4.2

Table 9.1: Permutaion error for each approach.

From the regression model, we can find that Linear Regression approach has the worst performance, in both training and testing situations. The BP Network has a very low training D but a high testing D. The most possible reason is that the neural network over-fits the training data, the weakness of BP Neural Network mentioned before, so that it lacks the generalization ability.

The performance of Deap Learning and ListNet are similar for the close permutation error. In addition, both of them are time and gender invarient without modifying coefficient, and can be applied across the sports with just changing the training data to get the adapted coefficient.

In conclusion, both of the Classification Model (Deep Learning) and the Optimization Model (ListNet) are qualified to this coach ranking problem. Finally, considering the smaller permutation error, we choose the Optimization Model as our final model. The top 5 coaches ranked by this model is shown as Table 9.2, Table 9.3 and Table 9.4.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	John Wooden	29	826	0.804	54	12	10
2	Adolph Rupp	41	1066	0.822	71	1	9.358209
3	Bob Knight	42	1273	0.706	47	5	9.267523
4	Mike Krzyzewski	39	1277	0.764	69	0	9.266344
5	Dean Smith	36	1133	0.776	70	1	9.157346

TABLE 9.2: Top 5 college basketball coaches.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Bear Bryant	38	425	0.78	15	3	10
2	Joe Paterno	46	548	0.749	24	1	9.764865
3	Bobby Bowden	40	485	0.74	22	1	9.142352
4	Amos Alonzo Stagg	42	425	0.681	0	1	8.808064
5	Pop Warner	42	446	0.733	1	0	8.773349

TABLE 9.3: Top 5 college football coaches.

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Gordie Gillespie	59	1893	0.665	5	1	10
2	Rod Dedeaux	44	1342	0.691	10	1	8.920852
3	Don Schaly	40	1442	0.812	0	1	7.440755
4	Ed Cheff	34	1705	0.799	0	1	6.923967
5	Cliff Gustafson	29	1427	0.792	2	1	6.420532

TABLE 9.4: Top 5 college baseball coaches.

Section 10 Article

Our magazine the Sports Illustrated is looking for the “best all time college coach” male or female for the previous century for sports fans and loyal readers.we have completed a rank for all time coaches of the basketball,football and baseball using the constructed model recently.Here are the results and corresponding analysis for reference.

We use the method of machine learning to train the sample data to get model coefficients.Then we use the model to deal with all data to implement ranking. In evaluation of the college coaches, many factors are considered such as coaching years, number of winning Games, winning percentage, the number of champions in top contests, the number of honor and other factors which can not be quantized-historical contribution, whether bringing up the star player, whether creating amazing records, our model concentrates on the former five factors.In addition, if a coach has brilliant performance with ordinary team, the coach should be ranked higher.

We prove our model can be applied across the time after several controlled trials whose independent variable is the time.Thus,we can apply our model to all data to attain the result of all time.

For the football ball coaches, we use the College Football Conference Statistics on sports reference to get the data of of all time coaches with five main factors referred above. We consider the team performance in the Bowl game,consider the honor such as Coach of year,Hall of Fame and the Bryant prize.Finally, we get the top 5 college coaches ever:Bear Bryant,Joe Paterno,Bobby Bowden,Amos Alonzo Stagg and Pop Warner.

For the basketball ball coaches, we use the Basketball Conference Statistics on sports reference to get the data of all time coach with five main factors. We consider the team performance (whether champion or not,whether top four or not) in the top contest such as regular season conference,conference tournament and NCAA tournament,consider the honor such as Coach of year,Hall of Fame and Wooden Prize.At last, we attain the top 5 all time basketball college coaches:John Wooden,Mike Krzyzewski,Bob Knight,Dean Smith and Adolph Rupp.

For the baseball ball coaches, we use the Baseball Conference Statistics on the wikipedia to get the data of all time coaches with five main factors. We consider the team performance in the NCAA tournament,consider the honor Coach of year,Hall of Fame.At last, we attain the top 5 all time baseball college coaches:Gordie Gillespie,Rod Dedeaux,Don Schaly,Ed Cheff and Cliff Gustafson.

To sum up, we rank the top 5 all time college coaches using our own model with parameters of main factors,the result is rational on common sense.

List of Figures

List of Tables

Section A Appendix A: Results of Basketball coaches of List-Net model

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	John Wooden	29	826	0.804	54	12	10
2	Adolph Rupp	41	1066	0.822	71	1	9.358209
3	Bob Knight	42	1273	0.706	47	5	9.267523
4	Mike Krzyzewski	39	1277	0.764	69	0	9.266344
5	Dean Smith	36	1133	0.776	70	1	9.157346
6	Jim Calhoun	40	1259	0.697	58	1	8.787727
7	Jim Boeheim	38	1256	0.75	51	2	8.732497
8	Eddie Sutton	37	1135	0.71	46	3	8.322785
9	Roy Williams	26	902	0.793	53	4	8.212995
10	Lute Olson	34	1061	0.731	51	0	7.69237
11	Denny Crum	30	970	0.696	57	0	7.557599
12	Jerry Tarkanian	30	963	0.79	49	0	7.400804
13	Ray Meyer	42	1078	0.672	20	4	7.260707
14	Lefty Driesell	41	1180	0.666	35	0	7.258451
15	Phog Allen	48	978	0.735	32	0	7.114043
16	Rick Pitino	28	920	0.74	47	0	7.016763
17	Bob Huggins	29	935	0.711	45	0	6.913867
18	Hank Iba	40	1085	0.693	29	0	6.806075
19	Lou Henson	41	1195	0.649	25	0	6.768746
20	Don Haskins	38	1072	0.671	29	0	6.652092
21	Ralph Miller	38	1044	0.646	17	3	6.617977
22	Guy Lewis	30	871	0.68	26	3	6.489658
23	Norm Stewart	32	967	0.656	30	1	6.450256
24	Bill Self	21	693	0.756	40	2	6.441149
25	John Calipari	22	756	0.774	44	0	6.404508
26	Tubby Smith	23	759	0.688	33	3	6.362676
27	E.A. Diddle	42	1061	0.715	17	0	6.30906
28	John Chaney	24	769	0.671	31	3	6.279938

29	John Thompson	27	835	0.714	37	0	6.258887
30	Gene Keady	27	839	0.656	26	3	6.255431
31	Cliff Ellis	36	1085	0.61	15	2	6.225424
32	Gary Williams	33	1048	0.637	26	0	6.19529
33	Mike Montgomery	32	982	0.682	24	0	6.040882
34	Lou Carnesecca	24	726	0.725	26	2	5.868104
35	Tom Izzo	19	639	0.717	33	2	5.823229
36	Billy Tubbs	29	926	0.658	24	0	5.756271
37	Nolan Richardson	22	716	0.711	35	0	5.729613
38	Norm Sloan	33	917	0.609	14	2	5.716663
39	Thomas Penders	33	1014	0.586	20	0	5.706974
40	Bobby Cremins	31	965	0.607	22	0	5.68281
41	Stew Morrill	28	873	0.684	24	0	5.676175
42	Tom Davis	32	953	0.626	16	1	5.675354
43	Hugh Durham	37	1064	0.596	13	0	5.644035
44	Rick Barnes	27	871	0.664	25	0	5.633417
45	Rick Majerus	25	733	0.705	30	0	5.616832
46	Frank McGuire	30	785	0.699	22	0	5.504691
47	Jim Larranaga	28	842	0.596	17	2	5.4931
48	Slats Gill	36	992	0.604	13	0	5.478072
49	Pete Carril	30	798	0.658	22	0	5.42725
50	Fran Dunphy	25	726	0.653	29	0	5.421529
51	Gale Catlett	30	890	0.635	19	0	5.418306
52	Don DeVoe	31	901	0.568	20	0	5.353275
53	Mark Few	15	491	0.804	36	0	5.295872
54	Jim Harrick	23	705	0.667	28	0	5.294933
55	Gene Bartow	24	744	0.66	25	0	5.250842
56	Lon Kruger	28	869	0.611	19	0	5.242419
57	Bob McKillop	25	754	0.615	26	0	5.239911
58	Jack Gardner	28	721	0.674	22	0	5.235978
59	Ben Howland	19	607	0.657	25	2	5.220568
60	Billy Donovan	20	658	0.714	28	0	5.210581

Section B Appendix B: Results of football coaches of ListNet model

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	John Wooden	29	826	0.804	54	12	10
2	Adolph Rupp	41	1066	0.822	71	1	9.358209
3	Bob Knight	42	1273	0.706	47	5	9.267523
4	Mike Krzyzewski	39	1277	0.764	69	0	9.266344
5	Dean Smith	36	1133	0.776	70	1	9.157346
6	Jim Calhoun	40	1259	0.697	58	1	8.787727
7	Jim Boeheim	38	1256	0.75	51	2	8.732497
8	Eddie Sutton	37	1135	0.71	46	3	8.322785
9	Roy Williams	26	902	0.793	53	4	8.212995
10	Lute Olson	34	1061	0.731	51	0	7.69237
11	Denny Crum	30	970	0.696	57	0	7.557599
12	Jerry Tarkanian	30	963	0.79	49	0	7.400804
13	Ray Meyer	42	1078	0.672	20	4	7.260707
14	Lefty Driesell	41	1180	0.666	35	0	7.258451
15	Phog Allen	48	978	0.735	32	0	7.114043
16	Rick Pitino	28	920	0.74	47	0	7.016763
17	Bob Huggins	29	935	0.711	45	0	6.913867
18	Hank Iba	40	1085	0.693	29	0	6.806075
19	Lou Henson	41	1195	0.649	25	0	6.768746
20	Don Haskins	38	1072	0.671	29	0	6.652092
21	Ralph Miller	38	1044	0.646	17	3	6.617977
22	Guy Lewis	30	871	0.68	26	3	6.489658
23	Norm Stewart	32	967	0.656	30	1	6.450256
24	Bill Self	21	693	0.756	40	2	6.441149
25	John Calipari	22	756	0.774	44	0	6.404508
26	Tubby Smith	23	759	0.688	33	3	6.362676
27	E.A. Diddle	42	1061	0.715	17	0	6.30906
28	John Chaney	24	769	0.671	31	3	6.279938

29	John Thompson	27	835	0.714	37	0	6.258887
30	Gene Keady	27	839	0.656	26	3	6.255431
31	Cliff Ellis	36	1085	0.61	15	2	6.225424
32	Gary Williams	33	1048	0.637	26	0	6.19529
33	Mike Montgomery	32	982	0.682	24	0	6.040882
34	Lou Carnesecca	24	726	0.725	26	2	5.868104
35	Tom Izzo	19	639	0.717	33	2	5.823229
36	Billy Tubbs	29	926	0.658	24	0	5.756271
37	Nolan Richardson	22	716	0.711	35	0	5.729613
38	Norm Sloan	33	917	0.609	14	2	5.716663
39	Thomas Penders	33	1014	0.586	20	0	5.706974
40	Bobby Cremins	31	965	0.607	22	0	5.68281
41	Stew Morrill	28	873	0.684	24	0	5.676175
42	Tom Davis	32	953	0.626	16	1	5.675354
43	Hugh Durham	37	1064	0.596	13	0	5.644035
44	Rick Barnes	27	871	0.664	25	0	5.633417
45	Rick Majerus	25	733	0.705	30	0	5.616832
46	Frank McGuire	30	785	0.699	22	0	5.504691
47	Jim Larranaga	28	842	0.596	17	2	5.4931
48	Slats Gill	36	992	0.604	13	0	5.478072
49	Pete Carril	30	798	0.658	22	0	5.42725
50	Fran Dunphy	25	726	0.653	29	0	5.421529
51	Gale Catlett	30	890	0.635	19	0	5.418306
52	Don DeVoe	31	901	0.568	20	0	5.353275
53	Mark Few	15	491	0.804	36	0	5.295872
54	Jim Harrick	23	705	0.667	28	0	5.294933
55	Gene Bartow	24	744	0.66	25	0	5.250842
56	Lon Kruger	28	869	0.611	19	0	5.242419
57	Bob McKillop	25	754	0.615	26	0	5.239911
58	Jack Gardner	28	721	0.674	22	0	5.235978
59	Ben Howland	19	607	0.657	25	2	5.220568
60	Billy Donovan	20	658	0.714	28	0	5.210581

Section C Appendix C: Results of baseball coaches of ListNet model

Rank	Coach	Years	Games	W-L	Champion	Honor	Rating
1	Gordie Gillespie	59	1893	0.665	5	1	10
2	Rod Dedeaux	44	1342	0.691	10	1	8.920852
3	Don Schaly	40	1442	0.812	0	1	7.440755
4	Ed Cheff	34	1705	0.799	0	1	6.923967
5	Cliff Gustafson	29	1427	0.792	2	1	6.420532
6	Ron Fraser	30	1271	0.742	2	1	6.009221
7	Bill Wilhelm	36	1161	0.683	0	1	5.748607
8	Ron Polk	35	1373	0.662	0	1	5.694265
9	Chuck Brayton	33	1158	0.687	0	1	5.431216
10	Augie Garrido	45	1874	0.682	5	0	5.416128
11	Bob Bennett	34	1300	0.631	0	1	5.277639
12	Jim Brock	23	1099	0.713	2	1	4.819458
13	Skip Bertman	18	870	0.724	5	1	4.699423
14	Bibb Falk	25	444	0.732	2	1	4.524913
15	Bob Todd	27	1025	0.647	0	1	4.314621
16	Gary Ward	21	1022	0.739	0	1	4.300907
17	Bill Holowaty	45	1404	0.727	0	0	4.264458
18	Frank Vieira	44	1127	0.776	0	0	4.229776
19	Jack Coffey	47	1160	0.705	0	0	4.084606
20	Gene Stephenson	36	1837	0.731	1	0	3.902506
21	Chuck Hartman	47	1444	0.638	0	0	3.879351
22	Bobby Winkles	13	524	0.752	3	1	3.579304
23	Mike Martin	34	1771	0.743	0	0	3.493421
24	John Barry	40	619	0.807	0	0	3.48487
25	Mark Marquess	37	1495	0.656	2	0	3.320165
26	Don Brandon	38	1110	0.713	0	0	3.061469
27	Larry Cochell	39	1331	0.621	1	0	2.924379
28	Larry Hays	38	1508	0.637	0	0	2.905481

29	George Valesente	39	1000	0.685	0	0	2.858622
30	Jim Morris	31	1391	0.694	2	0	2.806832
31	Al Ogletree	40	1208	0.63	0	0	2.779262
32	Tommy Thomas	40	1308	0.613	0	0	2.755486
33	Skip Wilson	46	1034	0.556	0	0	2.746107
34	Itch Jones	39	1242	0.623	0	0	2.647608
35	Bob Hannah	36	1053	0.694	0	0	2.634747
36	Jack Stallings	39	1255	0.61	0	0	2.56513
37	John Scolinos	45	1070	0.528	0	0	2.461972
38	Mike Fox	30	1174	0.745	0	0	2.44499
39	Dean Bowyer	36	1053	0.66	0	0	2.384593
40	Jim Mallon	34	1196	0.665	0	0	2.337053
41	Joe Roberts	32	1283	0.673	0	0	2.254858
42	Mike Metheny	31	1214	0.685	0	0	2.158655
43	Tom Austin	32	1003	0.696	0	0	2.140435
44	Tim Pettorini	30	1002	0.726	0	0	2.130959
45	Gary Grob	35	1020	0.643	0	0	2.111492
46	Jim Gilligan	35	1227	0.605	0	0	2.041602
47	Rudy Abbott	32	1003	0.682	0	0	2.037431
48	Norm DeBriyn	33	1161	0.641	0	0	2.010424
49	Jack Leggett	32	1224	0.638	0	0	1.937578
50	Ray Tanner	25	1133	0.699	2	0	1.894698
51	Jay Bergman	32	1210	0.631	0	0	1.871893
52	Dewey Kalmer	40	1032	0.53	0	0	1.865223
53	Andy Lopez	30	1090	0.621	2	0	1.850224
54	Paul Mainieri	29	1122	0.646	1	0	1.750656
55	Jack Smitheran	35	1097	0.581	0	0	1.73333
56	Pete Dunn	32	1202	0.612	0	0	1.723997
57	Gary Adams	35	1172	0.567	0	0	1.706302
58	Jim Dietz	31	1230	0.62	0	0	1.696627
59	Fred Hill	34	1061	0.596	0	0	1.69263
60	Les Murakami	30	1079	0.654	0	0	1.679222

Bibliography

[1] Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F., and Li, H. (2007). Learning to rank: from pairwise approach to listwise approach. In ICML, pages 129-136.

[2] Burges, C. J. C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., and Hullender, G. N. (2005).Learning to rank using gradient descent. In ICML, pages 89-96.

[3] Li, H. (2009). Learning to rank. In Tutorial Abstracts of ACL-IJCNLP 2009, page 5, Suntec, Singapore. Association for Computational Linguistics.

[4] Freund, Y., Iyer, R. D., Schapire, R. E., and Singer, Y. (2003). An ecient boosting algorithm for combining preferences. Journal of Machine Learning Research, 4:933-969.

[5] Gupta, P. (2011). Learning-to-Rank: Using Bayesian Networks. Master's thesis, DA-IICT, India.

[6] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang, and H.-W. Hon. Adapting ranking SVM to document retrieval. In SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval,2006

[7] Bertsekas, D.P.: Nonlinear Programming. Athena Scientific (1995)

[8] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2, 121–167 (1998)

[9] T. Zhang. Solving large scale linear prediction problems using stochastic gradien-t descent algorithms. In ICML '04: Proceedings of the twenty-first international conference on Machine learning, 2004.

[10] M. Ranzato, Y-L. Boureau, and Y. LeCun. Sparse feature learning for deep belief networks. In NIPS, 2008.