HodgeRank, PageRank on ranking human age and their application in Hidden Markov Model

Xu Siao(sxuao@connect.ust.hk) ID:20377629 Department of mathematics HKUST
Li Juncheng(jlicv@connect.ust.hk) ID:20377124 Department of mathematics HKUST
Qi Bu(qbu@connect.ust.hk) ID: 20090594 Department of physics HKUST

Abstract

We use HodgeRank method ranking the age of 30 people. We gain different results with using different statistical models. We employ Hodge decomposition theorem for analyzing inconsistency of our graph. Then, we compare the results from HodgeRank model with that from PageRank model and find something interesting. Finally, we extend HodgeRank model further with application in hidden Markov model.

1 Introduction

Ranking has been widely applied to resolve conflicts in our society. In large datasets, people are often reluctant to offer complete ranks because of long time. Alternatively, we might seek crowdsource to simplify such task, by distributing pairwise comparison tasks to different accessory and analyzing results with appropriate models. Potential models are HodgeRank on Random Graphs and PageRank. We would also explore how different sampling methods in the dataset may affect the ranking results using Erdos-Renyi (ER) and Barabasi-Albert (BA) random graphs. Moreover, we extend the outcome, by predicting annotators' decision with Hidden Markovian Model. Hidden Markov model (HMM) is typically used to predict the hidden regimes of observation data. Therefore, this model finds applications in many different areas, such as speech recognition systems, computational molecular biology and financial market predictions. In our project, we use the global photo score derived from our HodgeRank as medium of input, by which we derive state transitional matrix and observation probability matrix. Then, using hidden Markov model predicts decision, Y_{ij}, based on the validity of decision as input.

2 data source

Data is provided by prof. Yao from course home page. Data is the pairwise comparison data we collected. 30 images from human age dataset are annotated by a group of volunteer users on ChinaCrowds platform. The annotator is presented with two images and given his choice of which one is older (or difficult to judge). Totally, we obtain 14,011 pairwise comparisons from 94 annotators.

3 Methodology

3.1 HodgeRank Model

Let D = {1,⋯,m} be a set of participants

V = {1,⋯,n} the set of photos to be ranked

$$\mathbf{Y}_{\mathbf{ij}}^{\boldsymbol{\alpha}} = \begin{cases} 1 & if\ \alpha\ prefers\ i\ to\ j \\ -1 & otherwise \end{cases} \quad \text{and } \mathbf{Y}_{\mathbf{ij}}^{\boldsymbol{\alpha}} = -Y_{ji}^{\alpha}$$

$$\boldsymbol{\omega}_{\mathbf{ij}}^{\boldsymbol{\alpha}} = \begin{cases} 1\ if\ \alpha\ makes\ a\ comparison\ for\ \{i,j\} \\ 0\ other \end{cases}$$

$$\min \sum_{i,j,\alpha} \omega_{ij}^{\alpha} \left(s_i - s_j - Y_{ij}^{\alpha}\right)^2$$

which is equivalent to $\min \sum_{i,j} \omega_{ij} \left(s_i - s_j - Y_{ij}\right)^2$ where

$$\mathbf{Y}_{\mathbf{ij}} = (\textstyle\sum_{\alpha} \omega_{ij}^{\alpha} Y_{ij}^{\alpha})/(\sum_{\alpha} \omega_{ij}^{\alpha}) \quad \text{and } \boldsymbol{\omega}_{\mathbf{ij}} = \sum_{\alpha} \omega_{ij}^{\alpha}$$

3.2 PageRank Model

The PageRank of a page is defined by the sum of the PageRanks of pages that point to it. Given a directed network structure, we may find the corresponding stochastic matrix and the “Google matrix”, and the PageRank scores are obtained from eigenvalue decomposition. In the age data, if one assessor determines that node i is older than node j, we model it by adding a directed edge from node j to node i. Following similar arguments, we may conclude that the nodes pointed by old nodes are likely to be old.

3.3 Random Networks

In this report two kinds of random networks are taken into consideration: ER and BA graphs. For ER graphs, given the number of nodes N, each pair of nodes are connected with the same probability p. Therefore, the expectation value of the number of edges is N(N-1) p/2 if it is an undirected graph. For BA networks, also known as "scale-free" networks, the degree distribution follows the "power law" for large degrees. In terms of sampling, BA graphs are less balanced than ER graphs and we would like to see how they differ in Hodge inconsistency.

3.4 Hidden Markov Model

The basic elements of a hidden Marko model are:

- Length of observation data, T
- Number of states, N
- Number of symbols per state, M
- Observation sequence, O = $\{\mathbf{O_t}, \boldsymbol{t} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \boldsymbol{T}\}$
- Hidden state sequence, Q = $\{\mathbf{q_t}, \boldsymbol{t} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \boldsymbol{T}\}$
- Symbol values of each state, $\{\mathbf{S_i}, \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \boldsymbol{N}\}$
- Transitional matrix, A = $(\mathbf{a_{ij}}), \boldsymbol{where\ a_{ij} = P(q_t = S_j\ | q_{t-1} = S_i)}$
- Vector of initial probability of being in state 1, $\mathbf{p_i} = \boldsymbol{P(q_1 = S_i), i} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \boldsymbol{N}$
- Observation probability matrix, B = $(\mathbf{b_{ik}}), \boldsymbol{where\ b_{ik} = P(O_t = v_k\ | q_t = S_i), i} = \mathbf{1}, \mathbf{2}, .., \boldsymbol{N\ and\ k} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \boldsymbol{M}$

There are three sets of algorithm which solves out three problems:

- $\boldsymbol{the\ probability\ of\ obsevations, P(O|\lambda)}$
- $\boldsymbol{choose\ the\ best\ corresponding\ state\ sequence\ Q = \{q_1, q_2, q_3, \dots, q_T\}}$
- $\boldsymbol{calibrate\ the\ best\ HMM\ parameter\ \lambda(A, B, p)\ to\ maximize\ P(O|\lambda)}$

(1) Forward algorithm:

initialization:

for i = 1, 2,⋯,N

$$\boldsymbol{\alpha_{t=1}(i) = p_i b_i(O_1)}$$

recursion:

for t =2, 3,⋯,T and for j = 1,2,⋯,N

$$\boldsymbol{\alpha_t(j) = [\sum_{i=1}^{N} \alpha_{t-1}(i)\alpha_{ij}] b_j(O_t)}$$

output:

$$\boldsymbol{P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)}$$

(2) The Viterbi algorithm

initialization:

$$\boldsymbol{\sigma_1(j) = p_j b_j(O_1), \qquad j = 1, 2, \ldots, N}$$

$$\boldsymbol{\gamma_1(j) = 0}$$

recursion:

for 2 ≤ t ≤ T, 1 ≤ j ≤ N

$$\boldsymbol{\sigma_t(j) = \max_i [\sigma_{t-1}(i)\alpha_{ij}] b_j(O_{t+1})}$$

$$\boldsymbol{\gamma_t(j) = argmax_i [\sigma_{t1}(i)\alpha_{ij}]}$$

output:

$$\boldsymbol{q_t^* = \gamma_{t+1}(q_{t+1}^*), t = T - 1, \ldots, 1}$$

$$\boldsymbol{q_T^* = argmax_i [\gamma_T(i)]}$$

(3) <u>The Baum-Welch algorithm:</u>

initialization:

input parameters λ, the tolerance tol, and a real number del

recursion:

repeat until del < tol

Calculate P(O|λ) using forward algorithm

Calculate new parameters λ^*: $\boldsymbol{for\ 1 < i < N}$

$$p_i^* = \gamma_1(i)$$

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \ 1 \le j \le N$$

$$b_{ik}^* = \frac{\sum_{t=1, O_t = v_k}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}, \ 1 \le k \le M$$

Calculate del = |P(O,λ^*) – P(O,λ)|

Update $\lambda = \lambda^*$

output:

parameters λ

4 Curl Inconsistency for filtering careless voters

We would like to know more about how ranking methods interconnects and wonder if we can improve the ranking outcome by experimenting on methods.
With the method provided in section of methodology, we would like to find out careless voters by estimating each assessor's curl inconsistency, the measure of local cycle in percentage. Higher curl inconsistency may imply that assessors vote under a noisy setting. We may discard these observations and compare complete dataset's estimated global and local inconsistency with those of truncated dataset.

The results are displayed as follows:

Sample Pct:100%	harmIncon	TotalIncon
Uniform	6.50E-13	1.95E-01
Bradley-Terry	3.55E-13	3.21E-01
Thurstone-Mosteller	5.24E-13	2.45E-01
Arcsin	6.21E-13	1.98E-01
Sample Pct:85%	harmIncon	TotalIncons
Uniform	1.54E-10	0.202640506
Bradley-Terry	1.12E-10	0.401498169
Thurstone-Mosteller	1.21E-10	0.309979909
Arcsin	1.37E-10	0.215677286

Higher inconsistency does cast doubt on our analysis. Maybe we need some other criterion to filter participants.

5 implementation of HodgeRank and PageRank

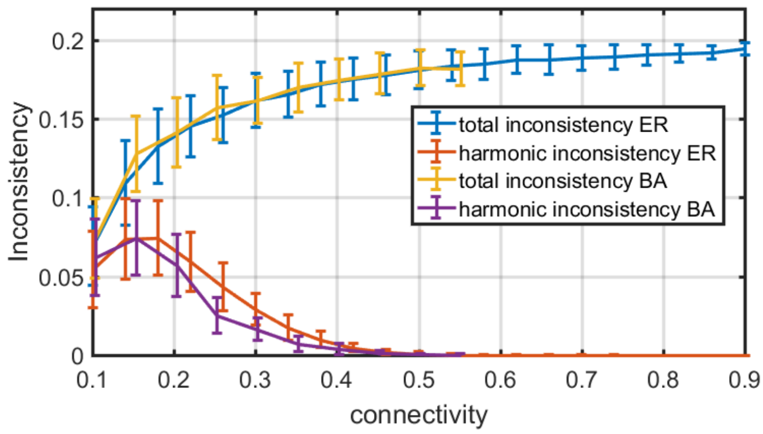
Results and comparison using different ranking scheme:

Ground Truth	HodgeRank				PageRank
	Uniform noise model	Bradley-Terry	Thurstone-Mosteller	Arcsin	α=0.85
1	2	2	2	2	2
2	1	1	1	1	1
3	5	18	5	5	5
4	7	5	18	7	7
5	4	4	4	4	3
6	18	7	7	18	10
7	3	12	3	3	4
8	10	3	12	10	18

9	12	10	10	12	12
10	22	13	13	22	16
11	13	9	22	13	22
12	9	20	9	9	13
13	15	15	15	15	15
14	11	11	11	11	9
15	6	22	20	6	20
16	20	6	6	20	6
17	14	16	16	16	11
18	16	14	14	14	8
19	8	8	8	8	14
20	21	25	21	21	21
21	19	19	19	19	19
22	25	21	25	23	23
23	23	23	23	25	24
24	24	24	24	24	17
25	17	17	17	17	25
26	26	26	26	26	28
27	29	28	28	27	27
28	27	27	27	29	26
29	28	30	29	28	30
30	30	29	30	30	29
Spearman	0.8501	0.8318	0.8372	0.8509	0.8461
total inconsistency	0.1948	0.3206	0.2449	0.1984	

From the results above, we may notice that according to the Spearman's index, the PageRank can also be used to rank the ages based on pairwise comparisons with similar accuracy to HodgeRank.

Random network sampling:



From the figure above, it seems that BA graphs may have lower harmonic inconsistency but slightly higher total inconsistency when the connectivity (sampling rate) is small. Therefore, we cannot determine which graph is better.

6 further extensions: global score with application in hidden Markov model

In the end of section 5, we gain a global score for thirty pictures, with which we can appeal to hidden Markov model for an interesting extension, that is, the prediction for next decision of annotator.

Firstly, we set annotators' decisions, Y, as hidden state which have two values, 1 and -1 respectively, and validity of specific decision as observation, which have two values as well, right and wrong respectively. Then, we define two probability matrix as follows:

State transitional probability matrix

	Y = 1	Y = -1
Y = 1	P(Y=1|Y=1)	P(Y=-1|Y=1)
Y = -1	P(Y=1|Y=-1)	P(Y=-1|Y=-1)

observation matrix for each state

	Y = 1	Y = -1
O = right	P(O=right|Y=1)	P(O=right|Y=-1)
O = wrong	P(O=wrong|Y=1)	P(O=wrong|Y=1)

For each state, there is a probability of observation conditional on corresponding state and there's also a transitional probability matrix connecting current state to next state. The medium how we establish matrix above is global score for each photo. For example, we have original data like that (1,3,6,1), which means first guy make decision that the people on photo 3 is older than that on photo 6. Then, we check out the global score on photo 3 and 6. If the global score on photo 3 is greater than photo 6, we note down that the decision is right and otherwise wrong if global score on photo 3 is smaller than that on photo 6. So on and so forth. In the end, we count up all decisions of 1, within which we find out "right" or "wrong" based on mean above each by each. Therefore, P(O=right|Y=1) is just the number of "right" under Y=1 divisible by total number of Y=1. Following the same logic, we can compute other three probabilities and all of four constitute observation matrix for each state.

For State transitional probability matrix, we count up numbers of Y=1, next to which is Y=1 and Y=-1 respectively and numbers of Y=-1 next to which Y=1 and Y=-1 respectively. Thereafter we generate P(Y=1|Y=1) as (number of Y=1, next to which is Y=1) divisible by (number of Y=1, next to which is Y=1 + number of Y=1, next to which is Y=-1). The other three are computed similarly. in the end, we establish full State transitional probability matrix.

State transitional probability matrix

(four all the same since the same state path data extracted from original data)

	Y = 1	Y = -1
Y = 1	0.6804522463	0.6452518579
Y = -1	0.3195477536	0.354748142

Uniform model

	Y = 1	Y = -1
O = right	0.825052067	0.817040951
O = wrong	0.1749479321	0.1829590488

Bradley-Terry model

	Y = 1	Y = -1
O = right	0.805712585	0.8208388375
O = wrong	0.194287414	0.1791611624

Thurstone-mosteller model

	Y = 1	Y = -1
O = right	0.8101755429	0.8173712021
O = wrong	0.189824457	0.1826287978

Angular transform model

	Y = 1	Y = -1
O = right	0.8231181196	0.8173712021
O = wrong	0.1768818803	0.1826287978

The above four matrices for observation probability and state transitional matrix evolve for each state by calling The Baum-Welch algorithm, which is a bit variation of EM algorithm, calibrating parameters for each state on which new input enter (the procedure is the same as that of first turn but include one more input for new turn).
Thereafter, we call on Viterbi algorithm which predicts the current state given new observation of validity of decision as input for each turn. And in the end, we have a very complete paths for predictive state which we compared with true decision paths given from original dataset. We gain the accuracy as follows:

(the accuracy for four models are all the same, why?)

	Uniform	Bradley-Terry
Accuracy	0.52606041	0.52606041

	Thurstone-mosteller	Angular transform
Accuracy	0.52606041	0.52606041

You may feel so strange why the accuracy for four are all the same? The reason is that the state transitional matrix is the same for four model and the global score for four are similar and the variation is so tiny that insulting in similar observations probability matrix which is even neglected while calling Baum-Welch algorithm for calibrating parameters. Finally, the prediction state paths for four model are almost equal to each other and therefore the final accuracy is the same when comparing with same real state path.
The biggest benefit of decision prediction is cost saving. For example, given global score and success rate of annotator decision (we assume 100% when the gap between two photo is so big and 50% when that of two photo is ambiguous), we can gain sufficient sampling without asking annotator to do questionnaire online. Furthermore, decision path prediction can be used to simulate people’s behaver, which we apply in such disciplines as psychology and economics.

7 conclusion and further work

- In cases where the dataset is complete, PageRank share similar results with different grading schemes of HodgeRank. We also try to use BA random network to sample the data and compare with ER network. However, we do not draw a clear conclusion on the comparison between ER and BA networks and further study and analysis is needed.

- We also tried the curl inconsistency based selection method proposed in the 1st Reference Paper. Total inconsistency increases as we decrease the threshold of inconsistency that screen out some participants; A systematic investigation is needed to identify how these assessors may contribute to the consistency of total ranking.

- The combination of HodgeRank and hidden Markov model above illustrate that the accuracy of which is 0.52606041. it means we can successfully predict almost half of decision that annotators make, which, I think, performs not so well. How to improve the accuracy of prediction? I think the global score we derive from HodgeRank model is the key. As we see above, when the global score for four models are similar with the same decision paths, the transitional state matrix and observation probability matrix are similar to each other as well, resulting in the same output in model. On the other hand, the way how to form the distribution of two probability matrix affects the model greatly. Obviously, improving the precision of global score in HodgeRank also improve the precision of prediction in hidden Markov model. And when the sample volume gets larger and larger, the precision of two probability matrix close to true situation by the law of large number.

8 contribution

Xu, Li and Qi evenly contribute to such everything as report writing, math modeling, data preprocessing and programming.

9 reference

[1] HodgeRank on Random Graphs for Subjective Video Quality Assessment, Qianqian Xu, Qingming Huang, senior Menber, IEEE, Tingting Jiang, Bowei Yan, Weisi Lin, Senior Member, IEEEE, and Yuan Yao.

[2] Google's pageRank and Beyond: The Science of Search Engine Rankings, Amy D. Langville and Carl D.Meyer.

[3] An Introduction to Hidden Markov Models, by L. R. Rabiner and B. H. Juang.

[4] Parsimonious Mixed-Effects HodgeRank for Crowdsourced Preference Aggregation, Qianqian Xu, Jiechao Xiong, Xiaochun Cao, Yuan Yao.