

# Human Age Ranking Using Hodge Rank

Chan Lok Chun

Department of Biochemistry and Cell Biology, HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

## Introduction

Being able to extract global ranking from a set of pairwise-comparison data is an important task in many different aspects, spanning from large-scale election to priority decision by Internet search engines. In this study, we intend to rank 30 human face images based on their ages. The dataset also contains 14011 pairwise comparison done by 94 “voters”. Due to the incompleteness of the comparison (i.e. each voter did not compare each image with 29 other images), we employed Hodge Rank method to solve the rank aggregation problem. The predicted global rank was then compared with the ground-truth to access the model accuracy. In addition, we also compared the performance by four most commonly used models, which are Uniform, Bradley-Terry, Thurstone-Mosteller and Angular transform model.

## Background

HodgeRank method is primary built on Hodge decomposition of a graph formed by the pairwise comparison data which is given by:

$$\hat{Y} = \hat{Y}^g + \hat{Y}^c + \hat{Y}^h$$

The first term is the gradient flow, more importantly, the second and the third terms refer to curl flow and harmonic flow respectively. Both of them contribute to the inconsistency in the ranking: curl flow indicates the local cyclic ranking, which can be regarded as local inconsistency. Whereas harmonic flow indicates global cyclic ranking that triggers global inconsistency in the ranking.

## Method

To evaluate the model, we measure the rank inconsistency by considering the curl flow and harmonic flow components. The total inconsistency of the model is given by:

$$Inc.Total(\hat{Y}) = \frac{\|\hat{Y}^c\|_w^2 + \|\hat{Y}^h\|_w^2}{\|\hat{Y}\|_w^2}$$

Moreover, we are also interested in the global inconsistency of the predicted ranking (global cyclic ranking). We therefore suffice to calculate the harmonic inconsistency as well:

$$Inc.Harm(\hat{Y}) = \frac{\|\hat{Y}^h\|_w^2}{\|\hat{Y}\|_w^2}$$

## Experiment 1

We first perform the HodgeRank analysis using the original pairwise comparison dataset which contains the ratings of all voters.

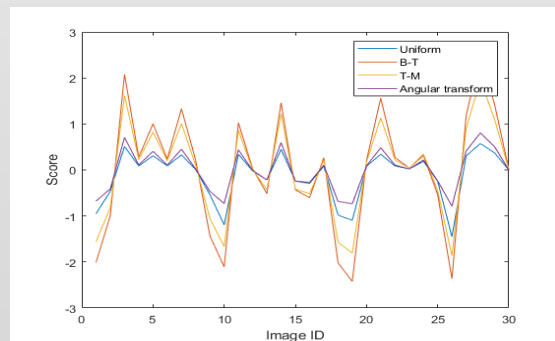


Fig 1. Rank score for all 30 images by four different models

From figure 1, we observe that all four models give similar distribution of rank score (it is a score assigned by the algorithm, with further treatment, it gives the global rank of the dataset).

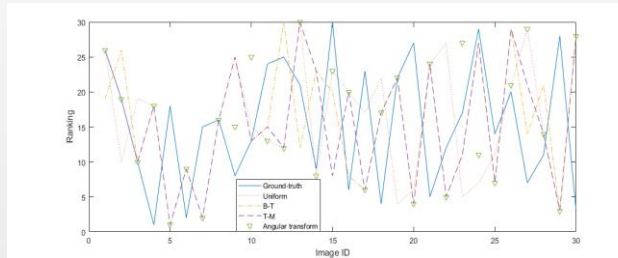


Fig 2. Distribution of global rank of the ground-truth and the four models

However, as shown in figure 2, all four models return rankings with pretty low correspondence to the ground-truth. We employed Spearman as the evaluation metric of the model accuracy, and found that all four models yield below 30% efficiency.

## Experiment 2

We wondered if the poor performance is due to the presence of some “poor” voters which disrupt the dataset, so we perform the HodgeRank analysis on each of the voters and finally ruled out those with >60% total inconsistency or >30% harmonic inconsistencies.

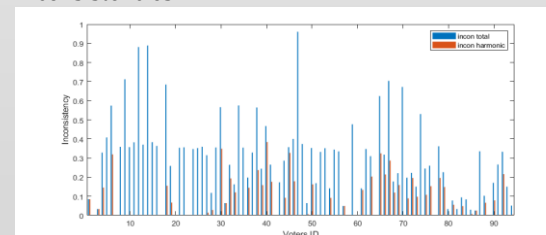


Fig 3. Bar plot of total and harmonic inconsistency of each voters

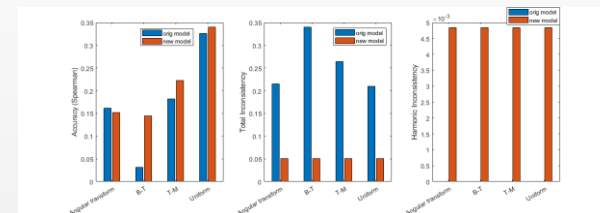


Fig 4. Comparison between new model and original model

After trimming down the data, we create a new ranking model, the new model was then compared with the original model (i.e. model without cutting down “poor” voters). It was found that except the Angular transform model, all others give higher accuracy. Despite having a small increase in harmonic inconsistency, all the new models give much improved total inconsistency.

## Experiment 3

In the final experiment, we are interested in the effect of random sampling of pairwise comparison data on the rank inconsistencies and prediction accuracy.

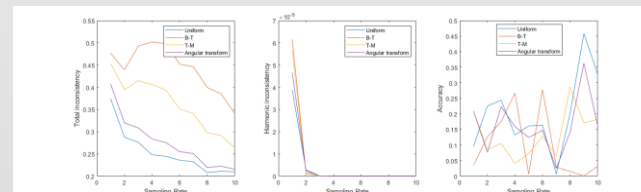


Fig 5. Effect of random sampling on rank inconsistency and accuracy

It is noticed that with increased size of samples being randomly selected, both total and harmonic inconsistencies decrease; however, we could not observe any notable increasing trend for the model accuracy. Instead, it fluctuates with increasing random sample size. The reason behind shall be further investigated. Moreover, we found that among four models, the Uniform model outperforms the others most of the time and has lowest total inconsistencies.