Biovistas

International Journal of Biological Research

ISSN: 2321-0524
Volume: 9 Issue: 1

Article Type: Original Research

Stochastic Model with Bayesian Noncompliance for a Ranking based Interaction and Collaboration Application

Sharang Thimmaiah[1], Maria Thomas[2]

Affiliation

1. Deparmtent of Mathematics, St. Joseph's College (Autonomous), Bengaluru, Karnataka 560027.

2. Department of Statistics, Christ University, Bengaluru, Karnataka 560029.

Abstract

Based on the needs of the Society of Jesus (S.J.), we have explored a mathematical model to implement a Jesuit Outreach application to undertake the goal of peace, justice and reconciliation among various ethnic and religious groups. This was achieved by trying to understand reasons behind various ethnic and socio-religious biases. This was done in order to present a means to implement this model in an android application and use the data collected through this to verify qualitatively our assumptionson migration with a special emphasis on student migrants under the classification of internal migrants. Our work has culminated in the development of an algorithm inspired by Brin and Page's PageRank Algorithm and a questionnaire to study ethnic biases. This algorithm is a stochastic model that may be updated and accounts for noncompliant behaviour using a Bayesian model. This will serve as a starting point for further study of ethnic biases on a global reach and on smaller scales as well. We believe that the model is supported by a detailed questionnaire with questions for certain foreseeable situations, which obtain enough information without making the concerned person uncomfortable.

Keywords: Jesuit Worldwide Learning, Displaced people, stochastic process, random interaction, Noncompliant Bayesian Inference, Platform abuse.

Introduction

In 2010, Fr. Adolfo Nicolas S.J., former Superior General of the Society of Jesus, called for drastic measures for reorganising the outreach and educational arms of the Society to follow St. Ignatius of Loyola's call of inclusivity, this leading to the inception of Jesuit Worldwide Learning [1, 2]. Tasked primarily with fulfilling the educational needs of forcibly displaced persons with an emphasis on women's education, the organisation required a formal network for allowing easy communication for members across the world, via local settings. An immediate project was to try and implement means to undertake the 16th goal under the United Nations sustainable development goals, that of peace, justice and strong institutions, and reconciliation among various groups [5, 6, 7].

Our project was to find a means of studying migration and its causes, in the backdrop of the migrant crises affecting the west and the middle-east currently. However, the general perception of migrants was too general to use in our analysis. We have taken up the special case of internal migration within the domain of the Republic of India and the subset of migrants who migrate in order to study in better institutions than what was available to them from their place of origin or under a scholarship/ free ship [1]. Keeping in mind the fact that many students not native to the region that they go to study in, we have attempted to make a statistical model for the implementationof an android application to allow such individuals to socialise, speak and receive help from volunteering locals. Taking into account the fact that many of the target demographic may have had troubled pasts and may be comfortable with certain types of individuals, we have attempted to not only map people to such suitable individuals, but to also over time expose these people to people outside of their "comfort zone", assuming this comfort zone to be ethnicity based [1]. We have also devised a way to take up data for abuse of the platform, and also information on volunteers who are not helpful, and then incorporate this into our mapping model to make this mapping more useful. Using a stochastic model, we will try to create a dynamic model, which may be updated based on a review system on board the application.

Received: 11.03.2021; **Revised:** 01.04.2021; **Accepted:** 10.05.2021

Corresponding Author
Ms. Maria Thomas
Research Scholar
Department of Statistics,
Christ University, Bengaluru, Karnataka.
Email: mdthomas215@gmail.com

Materials And Methods

Tools for modelling the problem

A questionnaire was designed to collect data from users on the platform to be used as a prior for further analysis. Constant reviews and other platform misuse is to be recorded an incorporated into the model to generate a relevant posterior. Based on the original data given, certain ranks are to be computed. A personal rank, an educational rank and a rank based on behaviour on the platform will be computed. A stratified population will be first studied and ways to interconnect them will be studied.

Based on the original data, we can determine useful information that can be divided into:

1. Personal and educational data
2. Data on migration
3. Data based on platform behaviour

Using these we can determine whether or not to map a certain individual to another individual based on personal (not updated frequently), educational (moderate frequency) and platform behaviour datasets (updated frequently) via the inferential model.

Mathematical Model

The mathematical model of our project draws heavily on the PageRank algorithm by Brin and Page. For this we try to create a model of the user of the app.

Random Interaction Model:

We draw heavily from PageRank's random surfer model, with a few notable exceptions. Like PageRank, we assume that the person in question will randomly select a person in his/her recommendations, and that we can rank people based on common interests. However, unlike PageRank, we are not assuming an equal likelihood for all persons in the recommendations and will use personal information given to determine bias towards a particular set of individuals and try to integrate it into the model.

Brin and Page noted that simply mapping intensity alone does not give good results. However it could be still used as an indicator for the algorithm. A larger mapping intensity could mean more importance.

We define Primary rank of an individual u as R_u ,to be :

$$R(u)=c\sum_{v\in B(u)}\frac{R(v)}{N(r)}+cR_i(u)$$

Where c is a normalisation constant, such that $|R(u)|=1$ [3] and $N(r)$ is the number of pages that r links to.

We then define a matrix A=[a_{uv}]

$$a_{uv}=\{\frac{1}{N(u)}\ if u\ links\ to\ v$$

0, otherwise

then **A** is a Markov matrix and ***R*** is a probability density vector of the population, and upon multiple iterations we get,

$$\boldsymbol{R}=c\mathbf{A}\boldsymbol{R}$$

therefore R is an eigenvector of A with eigenvalue of $\frac{1}{c}$

Abuse ranking is also found as:

$$\boldsymbol{Ab}(u)=c\sum_{v\in B(u)}\frac{m(u,v)}{N(v)}$$

$m(u,v)$={1, if reported abuse with respect to this individual

0, otherwise

Also, complement is given by $\boldsymbol{Ab'}(u)=1-\boldsymbol{Ab}(u)$

Large values of R and Ab' indicate a larger ranking in the recommendations and converse also holds.

Our Markov matrix is given by

A-Ab'+E

where E is a supplementary matrix describing the ethnic linkings that are based on documented ethno-religious or social conflicts. This new matrix, when normalised with respect to rows, will act as a Markov matrix whose eigenvectors will describe the general perception to a particular individual given initial conditions. The reason we use a Markov matrix is to allow us to determine the probability of one person, in a certain set of iterations, being mapped to another person in the population. This stochastic process is best understood in matrix form, as a transformation from a default probability density vector to the system's state probability density vector.

For ensuring people are not segregated based on ethnic differences in our model, **E** was scaled by (1-γ), where γ is a time dependent function

γ:t→[0,1]

where "γ" is a function that will allow users to gradually move away from a ethnic based mapping system to a more open minded space. As γ→1, **E** becomes insignificant and can be left out after this happens.

A drawback of this is that we cannot ensure noncompliant behaviour. To solve this, we looked at work done by Imbens and Rubin [4]. Assuming that noncompliance is absolute for the dataset, we have two probability density vectors to represent both: $[\boldsymbol{\delta_{ij}}]_{\mathbf{1\times n}}$ represents absolute

noncompliance and $\frac{e}{n}$ represents absolute compliance to data given. A useful consequence of this is that as population **n** increases, $\frac{e}{n}$ tends to 0, resulting in no effect from noncompliance beyond a certain degree.

Defining $D(\boldsymbol{R_1}|\boldsymbol{R_2})$ for the noncompliant behavioural component, as a likelihood function, which shows us how likely the dataset could have given unusable data given we assume that the data is fully correct.

This can be represented with an n×2 matrix for $\boldsymbol{R}$ that can supplement the ranking matrix

$$\boldsymbol{R}= [\ \delta_{ab}\frac{e}{n}]_{1\times n} \quad [4]$$

where $1 \leq a \leq n$ and b is a random number between 1 and n

Results

Our Markov matrix will be defined as:

$$\mathbf{M}=\mathbf{A}-\mathbf{Ab'}+(1-\gamma)\mathbf{E}$$

and the Markov matrix itself can be written as

$$\mathbf{M'}_{ij}=\mathbf{M}_{ij}/\ (\sum_{j=1}^{n,\ i\neq j} \mathbf{M}_{ij})$$

and this may be used to evaluate for the population mapping probability vector as:

$$\boldsymbol{R}_{n+1}=\mathbf{M'}\boldsymbol{R}_n$$

where **R** will be evaluated to a certain number of iterations as dictated by the size of the population. [8]

Similar to the Google matrix, people with many common interests that match is less likely and hence this operation may be reduced to a $O(n)$ operation instead of the $O(n^2)$ as expected of the matrix. [8]

Design of the survey questionnaire model for the study

Survey Questionnaire: Jesuit Worldwide Learning Project

Questionnaire

(i) Please fill in the given space on the right your answers to the corresponding question

(ii) Please fill in the answers in block letters

(iii) This questionnaire has **one** printed side. Please answer all questions asked.

A. Profile Details:		
(i)	Name	
(ii)	Date of Birth (DD-MM-YYYY)	
(iii)	Nationality/ Country of current residence	
(iv)	State of residence in nation mentioned in (iii)	
(v)	Ethnicity	
(vi)	Religion practising/ Religion of family	
(vii)	Highest Level of education	
(viii)	Specialisation	
(ix)	Sex	
B. Additional Information		
(i)	Could you state the reasons as to why you came to Bangalore?	
(ii)	What is your state of origin?	
(iii)	Why did you choose Bangalore specifically?	
(iv)	Have you faced discrimination of any form during your stay in Bangalore?	
(v)	Could you state as to what the problem was specifically?	

Conclusion

The above Random Interaction model is an algorithm that was devised after extensive discussions, both with individuals who have worked with people affected by ethnic biases and other socio-cultural issues and with individuals who have also been affected by said bias. This algorithm will account for noncompliant data if supplied to it. The evolution of the model as ethnic identities are rendered insignificant to the ranking, will provide valuable insight to the nature of biases. The model should be valid for use in situations similar to those presented here. This algorithm will serve as a template for a larger mathematical model for a ranking based application to be used by the Jesuits for global reach studies to understand ethnic biases.

The design of the questions was done in such manner to prevent the questionnaire from seeming to intrude on the personal lives of the individuals concerned. Also the questions were designed keeping in mind certain situations that may occur, like

1. Feedback on the application itself (this would probably also provide information on the success of the questionnaire, if seeming to be intrusive a complaint to the same will be received).

2. Options to report a person of inappropriate behaviour (would aid in the ranking system itself) and options to report misuse of the platform by another user.

Having said this, the question on how to improve the presented solution depends on identifying other factors that may have an influence in the bias of a user

Acknowledgement

The authors thank the Department of Mathematics, St. Joseph's College (Autonomous), Shanthinagar, Bengaluru and the faculty of the Indian Social Institute, Benson Town, Bengaluru for their support and encouragement. Special thanks to Rev. Fr. O.P. Martin and Rev. Fr. J. Joye of ISIB, for their valuable insights into the issue of migration.

Conflict of Interest

No known conflict of interest.

Funding Source

No funding was availed for this study.

References

1. 2nd International Jesuit Networking Conference. Final Document - 2nd International Conference on Jesuit Networking :Prophetic Networks for a Universal Mission, Georgetown University 15-17 February 2018 [internet] Jesuit Network 2019. [cited 30 December 2020]. Available at: https://jesuit.network/wp-content/uploads/2019/02/Prophetic-Networks-for-a-Universal-Mission-05.pdf
2. United Nations. About the Sustainable Development Goals [internet] United Nations 2020. [cited 30 December 2020]. Available at: https://www.un.org/sustainabledevelopment/sustainable-development-goals/
3. United Nations. Sustainable development goals |UNDP. [internet] United Nations 2020. [cited 30 December 2020]. Available at: https://www.undp.org/content/undp/en/home/sustainable-development-goals.html
4. United Nations. Goal 16: Peace, justice and strong institutions |UNDP. [internet] United Nations 2020. [cited 30 December 2020]. Available at: https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-16-peace-justice-and-strong-institutions.html
5. Martin, P.O., Gonsalves, R., Sharma, S. Through the prism of labour: Gender and distress migration in contemporary India. Bangalore: Indian social institute BangalorePublications; 2017.
6. Page, L., Brin, S., Motwani, R., Winograd, T. The PageRank citation ranking: Bringing order to the web. [internet] Stanford InfoLabs 1998 [cited 10 September 2020] Available at: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf
7. Imbens, G. W., Rubin, D. B. Bayesian inference for casual effects in randomised experiments with noncompliance. The Annals of Statistics.Vol. 25,No. 1.pp. 305–327. Institute of Mathematical Sciences; 1997. Available at: https://doi.org/10.1214/aos/1034276631
8. Langville, A., Meyer, C. D. Google's PageRank and Beyond: the science of search engine rankings. Princeton and Oxford: Princeton University Press; 2011.