PrefRank: fair aggregation of subjective user preferences

Anirban Basu
KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino-shi,
Saitama 356-8502, Japan
basu@kddilabs.jp

Juan Camilo Corena
KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino-shi,
Saitama 356-8502, Japan
corena@kddilabs.jp

Shinsaku Kiyomoto
KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino-shi,
Saitama 356-8502, Japan
kiyomoto@kddilabs.jp

Jaideep Vaidya
MSIS Department, Rutgers
1, Washington Park, Newark,
NJ 07102, USA
jsvaidya@rutgers.edu

Stephen Marsh
Faculty of Business and IT,
UOIT
2000 Simcoe St N, Oshawa,
ON L1H 7K4, Canada
stephen.marsh@uoit.ca

Yutaka Miyake
KDDI R&D Laboratories, Inc.
2-1-15 Ohara, Fujimino-shi,
Saitama 356-8502, Japan
miyake@kddilabs.jp

ABSTRACT

Ranking vast amounts of user-contributed content, such as digital photographs, is handled well through user-driven ranking, but user-driven ranking is often subjective and difficult to compare. The analytic hierarchy process helps making sense of subjective opinion, whereas finding a global ranking is a problem of rank aggregation of partially ranked lists. In this position paper, we propose a solution – PrefRank – based on eigenvector centrality that helps aggregating partially ranked lists. Our proposed approach can be used in other application scenarios involving qualitative judgement and ranking, such as reviewing academic papers for a conference.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering

General Terms

Human Factors, Ranking, Algorithms

Keywords

Analytic Hierarchy Process, Ranking, Recommendation

1. INTRODUCTION

Ranking of ever increasing digital content has become indispensable, which can be made more efficient with user-driven ranking. A fundamental problem of user-driven opinion is its subjective nature leading to difficulties with comparability. The other significant problem is the aggregation of such user-driven rankings where each user ranks only a very small subset of the entire item space, thus generating essentially incomparable partially ranked lists. In this position paper, we tackle these two problems in the light of a realistic application scenario: ranking of digital photographs. Apart from ranking digital photographs, our proposal is applicable to other areas involving subjective decision making, such as ranking research papers based in the review process for a research conference.

Jagabathula and Shah [4] argued that pairwise comparison of items form a better basis for recommender systems than standalone ratings. The Analytic Hierarchy Process due to Thomas L. Saaty [8] is a well-known pairwise comparison process for making complex decisions. The problem of rank aggregation is well studied, especially in the context of meta search engines. Distance measures such as the Spearman's footrule distance and Kendall's τ are used to compare and aggregate two ordered lists. Borda count [1] is a positional method for ordering lists and has been extended [7] for use with partial lists. Markov chains have been used [2] to develop ranking. PageRank [5] is a similar concept. Markov chain based methods require either row-stochastic or column-stochastic matrix representations of a graph. However, none of the existing academic work has dealt with the problem and the application scenario (i.e., photograph ranking) that we are focussing in this paper. Existing commercial websites apply user-driven photograph ranking but in most cases, the ranking procedure does not make use of pairwise comparisons. The only commercial service that allows for pairwise comparison is Pixoto[1].

The contributions of this paper are: (1) the use of the analytic hierarchy process to solve the problem of subjective ranking; and (2) a novel use of eigenvector centrality as a solution to the problem of disjoint ranked lists.

2. USER-DRIVEN RANKING

A major challenge with user-driven ranking is the high degree of inconsistent subjective opinion. Even with a preset criteria, users do not necessarily understand each criterion the same way. This problem is solved using a pairwise comparison technique called the analytic hierarchy process (AHP), proposed by Thomas L. Saaty [8]. The key idea behind AHP is the notion in psychology that we, humans,

[1]See: `http://www.pixoto.com/`

SAC'14 March 24-28, 2014, Gyeongju, Korea.

tend to make judgement by comparing things in pairs [9]. Without loss of generality, let us also assume that based on certain decision criteria, one user's local ranking of a sample of five items (images A through E) is:

$$R_I = \begin{pmatrix} 0.23932 \\ 0.21082 \\ 0.10822 \\ 0.20039 \\ 0.24125 \end{pmatrix} \tag{2.1}$$

This puts E (the fifth image) at the top spot, followed by A, B, D and C respectively. This relative ordering of the images only is important for rank aggregation. Not all items will be ranked by all users leading to disjoint partially ranked lists. The global set of items can be represented as a graph structure where each item should be connected to every other item by directional links. Every egress link signifies a loss while an ingress link signifies a win. If A has won against B three times and lost nine times, then A will have three ingress edges from B while having nine egress edges to B. With this notation, the local ranking obtained in the eigenvector R_I (equation 2.1) can be written as the degree matrix of graph (with rows signifying ingress, and columns signifying egress edges) is:

$$G_{R_I} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} \tag{2.2}$$

At the point of initialisation, each new vertex in the graph is connected to every other in either direction exactly once. This requirement follows from the Perron-Frobenius theorem [3,6] in linear algebra. When a user submits her locally ranked sub-graph, the cloud increments the relevant link counts in the global matrix. Periodically, the cloud runs an eigenvector decomposition. The final ranking order is obtained from the normalised principal eigenvector. For instance, a snapshot of a five-images graph (in reality, it will contain many more) changed sometime after initialisation may look like the one shown in figure 2.1. The degree matrix of the graph is written as:

$$G = \begin{pmatrix} 0 & 393 & 445 & 894 & 27 \\ 5 & 0 & 79 & 400 & 449 \\ 97 & 338 & 0 & 475 & 464 \\ 254 & 130 & 313 & 0 & 143 \\ 20 & 382 & 206 & 39 & 0 \end{pmatrix} \tag{2.3}$$

The normalised principal eigenvector of G is given as:

$$R = \begin{pmatrix} 0.32722 \\ 0.14596 \\ 0.22339 \\ 0.18806 \\ 0.11537 \end{pmatrix} \tag{2.4}$$

Sorting this vector, the rank order is inferred, i.e., image A is the best followed by, in order: C, D, B and E. Next, we present some points of interest in relation with the aforementioned proposal. (1) *Not all users may be considered equal*; hence the local sub-graphs may be weighted by some factors before joining them into the global graph. (2) *Time slicing* should be used to guarantee the consistency of the rank of an item within the set of its contemporaries. This resists *rank reversals* due to new items. (3) *Scheduling what to display to the user to rank* should attempt to ensure that each item receives a fair share of comparisons. (4) The items being compared must be actually comparable, thus *topic sensitivity* should be considered. (5) The local ranking generated by a user is, to a certain extent, private data because it reveals how much the user prefers each item; thus *privacy* is an open issue.

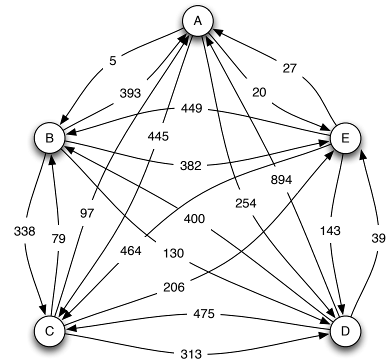

Figure 2.1: A state of ingress and egress edges of the global graph containing five images.

3. CONCLUSION AND FUTURE WORK

In this position paper, we have proposed a novel rank aggregation technique for user-driven subjective ranking of user-contributed content. We are currently developing a prototype of the system for experimentation on a real cloud platform and large volumes of data in order test its scalability.

4. REFERENCES

[1] J. C. de Borda. Mémoire sur les élections au scrutin. 1781.

[2] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

[3] G. Frobenius. *Über Matrizen aus nicht negativen Elementen*. Reimer, 1912.

[4] S. Jagabathula and D. Shah. Inferring rankings using constrained sensing. *Information Theory, IEEE Transactions on*, 57(11):7288–7306, 2011.

[5] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[6] O. Perron. Zur theorie der matrices. *Mathematische Annalen*, 64(2):248–263, 1907.

[7] D. G. Saari. The mathematics of voting: Democratic symmetry. *Economist*, 83, 2000.

[8] T. L. Saaty. The analytic hierarchy process: planning, priority setting, resources allocation. *McGraw-Hill*, 1980.

[9] L. Thurston. A law of comparative judgement. *Psychological Review*, 34:273–286, 1927.