A General Model for Mutual Ranking Systems

Vu Le Anh[1], Hai Vo Hoang[2][*], Kien Le Trung[3], Vinh Phan Cong[1], Hieu Le Trung[4],
and Jason J. Jung[1,5]

[1] Nguyen Tat Thanh University, Ho Chi Minh city, Vietnam,
[2] Information Technology College, Ho Chi Minh city, Vietnam,
[3] Hue University of Sciences, Hue City, Vietnam,
[4] Duy Tan University, Da Nang city, Vietnam
[5] Yeungnam University, Gyeongsan, Korea

Abstract. Ranking has been applied in many domains using recommendation systems such as search engine, e-commerce, and so on. We will introduce and study N-tier ranking, which can rank n classes of objects at once. The ranking scores of these classes are dependent to the others. For instance, PageRank by Google is a 2-tier ranking, which ranks the webpages and links at once. Particularly, we are demonstrating N-tier ranking in ranking conference and journal problems. We have conducted the experiments for the models in which the citations are not considered. The experimental results are based on the DBLP dataset, which contains more than one million papers, authors and thousands of conferences and journals in computer science. Finally, N-tier ranking is a very strong ranking algorithm can be applied in many real-world problems.

Key words: N-tier ranking, Markov chain, PageRank, Academic ranking, Conference ranking, Ranking algorithms, Prolific ranking, Recommendation systems, Bibliographical database, DBLP.

1 Introduction

Ranking is an interesting but difficult problem on many information processing systems. With a large amount of information, the systems need to adapt efficient ranking schemes to sort out (or to select) only the information which are highly relevant to the users' contexts. Particularly, in the context of *bibliometrics*, a set of given entities can be quantified to compare several evaluation indicators (e.g., popularity and reputation). For example, impact factors (IF) of international journals can be measured by taking into account how many times the papers in the corresponding journals have been cited.

However, there are some problems on the ranking schemes of traditional systems (i.e., web searching and impact factor). These systems are based on simple links among only the target entities (e.g., web pages and papers). We assume that this type of referencing process is not enough for measuring the evaluation indicators more precisely.

In this work, we formulate a mathematical model of N-tier ranking scheme, which can represent a generalized referencing process. The N-tier ranking is based on N-linear mutual ranking system in which N classes are ranked at once. Especially, as

---

[*] Corresponding Author: Hai Vo Hoang, Information Technology College, Ho Chi Minh, Vietnam. Phone/Fax: +84 938 642 717. Email: vohoanghai2@gmail.com.

a real-world problem, conference ranking has been regarded as an interesting issue on academic communities. The conference publication (e.g., proceedings) is related with various factors such as authors (and their affiliation), papers (and their citations), and conferences.

The remainder of this paper is as follows. In Sect. **??**, we describe the backgrounds of the N-tier ranking (e.g., the rank scoring and conceptual modeling on relationships between classes). Sect. **??** defines the N-tier ranking as a N-linear mutual ranking system in which there exists a core class. We study two N-tier ranking models for the author, publication and conference ranking problem in different contexts in Sect. **??**. Also, Sect. **??** shows the experiments for the simple N-tier ranking model of authors, publications and conference ranking. In Sect. **??**, we exhibit the related work for comparison with the proposed N-tier ranking model. Finally, Sect. **??** draws a conclusion of this work.

2 Backgrounds

2.1 N-linear mutual ranking system

The couple $(\mathcal{A}, R)$ is called a *ranking system* if (i) $\mathcal{A} = \{a_1, \ldots, a_n\}$ is a finite set, and (ii) $R : \mathcal{A} \to [0; +\infty)$. $\mathcal{A}$ is called a *class*, $a \in \mathcal{A}$ is called an *object of the class* $\mathcal{A}$, and R is called a *score* of $\mathcal{A}$. R is *positive* if $R(a) > 0 \ \forall a \in \mathcal{A}$. $n = |\mathcal{A}|$ is the *size* of $\mathcal{A}$.

Definition 1. $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^N$ *is called a* N-mutual ranking system *described by a system of functions* $F = (f_1, \ldots, f_N)$ *if* $(\mathcal{A}_i, R_i)$ *is a ranking system and*

$$R_i = f_i(R_1, \ldots, R_N) \qquad\qquad i = 1, \ldots, N.$$

F is called a *constraint system* of Ω, and f_i is called a *constraint function* of $(\mathcal{A}_i, R_i)$. f_i is called a *linear constraint function* if $\exists \alpha_{ij}, \beta_i \in [0; +\infty)$, $I_i = (t_u^{(i)})_{n_i}$ is a n_i-dimensional normalized nonnegative real number vector , $W_{ij} = (\omega_{kl}^{(ij)})_{n_i \times n_j}$ is a nonnegative real number and normalized columns matrix:

$$\sum_j \alpha_{ij} + \beta_i = 1 \quad and \quad f_i(R_1, \ldots, R_N) = \sum_{j=1}^N \alpha_{ij} W_{ij} R_j + \beta_i I_i.$$

$F = (f_1, \ldots, f_N)$ is called *linear constraint system* if f_i is linear constraint function for all $i \in \{1, \ldots, N\}$.

Definition 2. *The* N-mutual ranking system Ω described by F is called N-linear mutual ranking system *if* F is a linear constraint system.

Let a_{iu}, a_{jv} be objects in $\mathcal{A}_i, \mathcal{A}_j$ respectively. Suppose $\mathcal{C}^*(a_{iu}, a_{jv}) = \alpha_{ij}\omega_{uv}^{(ij)}$. From the definitions, we have:

$$R_i(a_{iu}) = \sum_{j=1}^{N}\sum_{v=1}^{n_j} \alpha_{ij}\omega_{uv}^{(ij)} R_j(a_{jv}) + \beta_i t_u^{(i)} = \sum_{j=1}^{N}\sum_{v=1}^{n_j} \mathcal{C}^*(a_{iu}, a_{jv})R_j(a_{jv}) + \beta_i t_u^{(i)}$$

a_{jv} is called *affect to* a_{iu} (denoted by $a_{jv} \to a_{iu}$) if $\mathcal{C}^*(a_{iu}, a_{jv}) > 0$. Class $\mathcal{A}_i$ is called *total affect and reflect directly to* class $\mathcal{A}_j$ (denoted by $\mathcal{A}_i \to \mathcal{A}_j$) if $\forall a_{jv} \in \mathcal{A}_j$: $\exists a_{iu_1}, a_{iu_2} \in \mathcal{A}_i$: $a_{jv} \to a_{iu_2} \wedge a_{iu_1} \to a_{jv}$.

Definition 3. *Class $\mathcal{A}_i$ is called* total affect and reflect to *class $\mathcal{A}_j$, denoted by $\mathcal{A}_i \rightsquigarrow \mathcal{A}_j$, if $\mathcal{A}_i \to \mathcal{A}_j$ or $\exists \mathcal{A}_k : \mathcal{A}_i \to \mathcal{A}_k \wedge \mathcal{A}_k \rightsquigarrow \mathcal{A}_j$.*

2.2 PageRank

We rewrite the PageRank into a 2-linear mutual ranking system as follows:

$\mathcal{W} = \mathcal{A}_1$ is the class representing for the set of webpages. $\mathcal{L} = \mathcal{A}_2$ is the class representing for hyperlinks. For each hyperlink $l \in \mathcal{L}$ from web $u \in \mathcal{W}$ to web $v \in \mathcal{W}$, we denote $u = in(l)$ and $v = out(l)$. For each $v \in \mathcal{W}$, we denote: $IN(v) = \{l \in \mathcal{L} | v = out(l)\}$ and $N_{out}(v) = |\{l \in \mathcal{L} | v = in(l)\}|$.

PageRank[**?**] determined the ranking system of webpages by the following formula: $\forall v \in \mathcal{W}$,

$$R_w(v) = d \sum_{l \in IN(v), u=in(l)} \frac{R_w(u)}{N_{out}(u)} + \frac{1-d}{|\mathcal{W}|} \tag{1}$$

where $d \in (0,1)$ is a constant.

Suppose $W_{21} = (\delta_{kt})_{|\mathcal{L}| \times |\mathcal{W}|}$ is a matrix in which $\delta_{kt} = \frac{1}{N_{out}(w_t)}$ if $l_k \in \mathcal{L} \wedge w_t = in(l_k) \in \mathcal{W}$, otherwise 0. W_{21} is a nonnegative real number and normalized columns matrix. Suppose $W_{12} = (\gamma_{tk})_{|\mathcal{W}| \times |\mathcal{L}|}$ is a matrix in which $\gamma_{tk} = 1$ if $l_k \in L \wedge w_t = out(l_k) \in \mathcal{W}$, otherwise 0. We construct a 2-linear mutual ranking system on two classes $\mathcal{W}$ and $\mathcal{L}$ as follows: Let $\bar{R}_w$ and $\bar{R}_l$ be scores on the classes $\mathcal{W}$, $\mathcal{L}$ respectively. They are satisfied:

$$\bar{R}_l = W_{12}\bar{R}_w \qquad and \qquad \bar{R}_w = dW_{21}\bar{R}_l + (1-d)I_{|\mathcal{W}|}, \tag{2}$$

where $I_{|\mathcal{W}|}$ denotes the $|\mathcal{W}|$-dimensional vector in which all its elements are $1/|\mathcal{W}|$. It is not difficult to see that (**??**) confirms: for all web $v \in \mathcal{W}$,

$$\bar{R}_w(v) = d \sum_{l \in IN(v), u=in(l)} \frac{\bar{R}_w(u)}{N_{out}(u)} + \frac{1-d}{|\mathcal{W}|}.$$

Since the equation (**??**) has the unique solution which is the PageRank score (see in [**?**]), $\bar{R}_w$ is the PageRank score R_w. Vice versa, if R_w is a solution of (**??**), R_w should be $\bar{R}_w$. Thus, the PageRank score R_w is totally determined by the equation (**??**), or in other words, PageRank can be presented as the two-linear ranking system described by (**??**).

Note that, since for each link $l \in \mathcal{L}$, let $u = in(l)$ and $v = out(l)$ then web u affects to link l, $(u \to l)$ and link l affects to web v, $(l \to v)$. Therefore, the class $\mathcal{W}$ total affect and reflect directly to the class $\mathcal{L}$, $\mathcal{W} \rightsquigarrow \mathcal{L}$.

3 N-tier Ranking Model

Definition 4. *Let $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^N$ be a N-linear mutual ranking system. Ω is called a N-tier ranking if $\exists \mathcal{A}_i : (\beta_i > 0) \wedge (I_i$ is positive $) \wedge (\forall \mathcal{A}_j (j \neq i) : \mathcal{A}_i \rightsquigarrow \mathcal{A}_j)$. $\mathcal{A}_i$ is called a* core *of the system Ω.*

Proposition 1. *PageRank is a 2-tier ranking system.*

Proof. Because $1 - d > 0$ and $\mathcal{W} \rightsquigarrow \mathcal{L}$, $\mathcal{W}$ is the core of PageRank. Therefore, PageRank is a 2-tier ranking system.

The ranking scores are determined by the linear constraint system and the classes.

Proposition 2. *Suppose the classes $\{\mathcal{A}_i\}_{i=1}^N$ and the linear constraint system $F = (f_1, \ldots, f_N)$ are given. There exists a unique $\{R_i\}_{i=1}^N$ in which R_i is a score on $\mathcal{A}_i$, such that $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^N$ is a N-tier ranking described by F and $\sum_{i, a \in \mathcal{A}_i} R_i(a) = 1$.*

Proof. Assuming the linear constraint functions f_i is given by:

$$f_i(R_1, \ldots, R_N) = \sum_{j=1}^N \alpha_{ij} W_{ij} R_j + \beta_i I_i \qquad with \qquad \sum_j \alpha_{ij} + \beta_i = 1.$$

The system $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^N$ is a N-tier ranking described by F if the sequence of scores $R_1, \ldots, R_N$ satisfy the following equations:

$$R_i = \sum_{j=1}^N \alpha_{ij} W_{ij} R_j + \beta_i I_i \qquad \forall \, i = 1, \ldots, N. \qquad (3)$$

Let $R = \frac{1}{N}[R_1, \ldots, R_N]$ be a n-dimensional vector with $n = |\mathcal{A}_1| + \cdots + |\mathcal{A}_N|$, and

$$W = \begin{pmatrix} \alpha_{11} W_{11} & \cdots & \alpha_{1N} W_{1N} \\ \vdots & \ddots & \vdots \\ \alpha_{N1} W_{N1} & \cdots & \alpha_{NN} W_{NN} \end{pmatrix} + \begin{pmatrix} \beta_1 \mathbf{I}_{n_1} & \cdots & \mathbf{O}_{n_1 \times n_N} \\ \vdots & \ddots & \vdots \\ \mathbf{O}_{n_N \times n_1} & \cdots & \beta_N \mathbf{I}_{n_N} \end{pmatrix},$$

where $\mathbf{I}_{n_i}$ is the $(n_i \times n_i)$-matrix which its columns are I_{n_i}, and $\mathbf{O}_{n_i \times n_j}$ is the $(n_i \times n_j)$-zeros matrix, **(??)** implies

$$R = WR. \qquad (4)$$

Since W is a nonnegative normalized columns matrix and R is a nonnegative normalized vector, R can be considered as a stationary distribution of a Markov chain in which its transition probability matrix is W. Because Ω is a N-tier ranking system, i.e. there exists a core $\mathcal{A}_i$, this Markov chain is ergodic (irreducible and aperiodic) (see in [**?**]). Thus, there exists a unique stationary distribution R satisfying the equation (**??**); in other words, there exists a unique sequence $\{R_i\}_{i=1}^N$ satisfying the sequence equations (**??**), or making $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^N$ be a N-tier ranking system described by F. The proposition is proven.

The ranking scores are computed by following algorithm:

Algorithm : Finding the sequence scores $\{R_i\}_{i=1}^N$

   Input : $\alpha_{ij}, \ \beta_i, \ W_{ij}, I_i$

   Output : $\{R_i\}_{i=1}^N$

1. **begin**
2.   Initialize $R_1^{(0)}, \ldots, R_N^{(0)}$: uniform distributions, $k = 0$
3.   **repeat**
4. $k = k + 1$
5.   **for** $i = 1$ to N **do**
6.   Update $R_i^{(k)} \leftarrow \sum_j \alpha_{ij} W_{ij} R_j^{(k-1)} + \beta_i I_i$
7.   **end**
8.   **until** $\sum_i \| R_i^{(k)} - R_i^{(k-1)} \| \leq$ a stopping criterion
9. **end**

---

4 Ranking Authors, Publications and Conferences

In this section, we apply the N-tier ranking model for constructing a model to evaluate authors, publications and conferences (journals) in the world of science. Concretely, we consider a four-tier ranking model corresponding with four ranking systems: $(\mathcal{A}, R_a)$ - $\mathcal{A}$ is a set of all scientists which has publications, $(\mathcal{P}, R_p)$ - $\mathcal{P}$ is a set of all publications, $(\mathcal{C}, R_c)$ - $\mathcal{C}$ is a set of all sciential conferences and sciential journals, and $(\mathcal{L}, R_l)$ - $\mathcal{L}$ is a set of all citations between publications. $R_a, \ R_p, \ R_c$ and R_l are the scores for each classes $\mathcal{A}, \ \mathcal{P}, \ \mathcal{C}$ and $\mathcal{L}$, respectively.

For each citation $l \in \mathcal{L}$, $u = in(l)$ and $v = out(l)$ if l is from $u \in \mathcal{P}$ to $v \in \mathcal{P}$. For each publication $v \in \mathcal{P}$, we denote: $IN(v) = \{l \in \mathcal{L}|v = out(l)\}$; $OUT(v) = \{l \in \mathcal{L}|v = in(l)\}$ $N_{out}(v) = |OUT(v)|$; $C(v) \in \mathcal{C}$ is the conference of v; $A(v) \subseteq \mathcal{A}$ is the set of authors of v. For each author $a \in \mathcal{A}$, $P(a) = \{v \in \mathcal{P}|a \in A(v)\}$ is a set of publications of a. For each conferenece $c \in \mathcal{C}$, $P_c(c) = \{v \in \mathcal{P}|c = C(v)\}$ is a set of publications published in c.

The 4-tier ranking system model for ranking authors, publications, conferences and citations is constructed based on some following ideas:

1. The score of an author depends only on his publications, and each publications affects to all of its authors:

$\forall a \in \mathcal{A}, p \in \mathcal{P}$:

$$R_a(a) = \sum_{p' \in P(a)} \mathcal{C}^*(a, p') R_p(p') \quad and \quad \sum_{a' \in A(p)} \mathcal{C}^*(a', p) = 1 \qquad (5)$$

If *a publication affects equally to its authors* (a), (??) is rewritten as follows:

$\forall a \in \mathcal{A}, p \in \mathcal{P}, a' \in A(p)$:

$$\mathcal{C}^*(a', p) = \frac{1}{|A(p)|} \quad and \quad R_a(a) = \sum_{p' \in P(a)} \frac{R_p(p')}{|A(p')|} \qquad (6)$$

2. The score of a conference depends only on its publications:

$$\forall c \in \mathcal{C} : R_c(c) = \sum_{p' \in P_c(c)} R_p(p'). \qquad (7)$$

3. The score of a citation depends on the citing publication, and each publications affects to all of its citations:

$\forall l \in \mathcal{L}, p \in \mathcal{P}, p' = in(l)$:

$$R_l(l) = \mathcal{C}^*(l, p') R_p(p') \quad and \quad \sum_{l' \in OUT(p)} \mathcal{C}^*(l', p) = 1 \qquad (8)$$

If *a publication affects equally to its citations* (b), (??) is rewritten as follows:

$\forall l \in \mathcal{L}, p \in \mathcal{P}, p' = in(l), l' \in OUT(p)$:

$$\mathcal{C}^*(l', p) = \frac{1}{N_{out}(p)} \quad and \quad R_l(l) = \frac{R_p(p')}{|N_{out}(p')|} \qquad (9)$$

4. The score of a publication depends on its citations, its authors and its conference and randomly finding by some reader. Each conference affects to all of its publications. Each author affects to all of its publications. Hence :

$\forall p \in \mathcal{P}, c \in \mathcal{C}, a \in \mathcal{A}, c' = C(p)$:

$$R_p(p) = \alpha_1 \sum_{l' \in IN(p)} R_l(l') + \alpha_2 \sum_{a' \in A(p)} \frac{\mathcal{C}^*(p, a')}{\alpha_2} R_a(a') + \alpha_3 \frac{\mathcal{C}^*(p, c')}{\alpha_3} R_c(c') + \beta_p I_p,$$

$$\qquad (10)$$

$$\sum_{p' \in P(a)} \mathcal{C}^*(p', a) = \alpha_2 \quad and \quad \sum_{p' \in P_c(c)} \mathcal{C}^*(p', c) = \alpha_3 \qquad (11)$$

where $\alpha_1, \alpha_2, \alpha_3, \beta_p > 0$ and $\alpha_1 + \alpha_2 + \alpha_3 + \beta_p = 1$, I_p is a $|\mathcal{P}|$-dimensional normalized uniform random vector.

If *a conference affects equally to its publications* (c) and *an author affects equally to its publications* (d), the equation (??) and (??) are rewritten as follows:

$$\forall p \in \mathcal{P}, c \in \mathcal{C}, a \in \mathcal{A}, c' = C(p), p' \in P(a), p'' \in P_c(c):$$

$$\mathcal{C}^*(p',a) = \frac{\alpha_2}{|P(a)|} \quad and \quad \mathcal{C}^*(p'',c) = \frac{\alpha_3}{|P_c(c)|} \tag{12}$$

$$R_p(p) = \alpha_1 \sum_{l' \in IN(p)} R_l(l') + \alpha_2 \sum_{a' \in A(p)} \frac{R_a(a')}{|P(a')|} + \alpha_3 \frac{R_c(c')}{|P_c(c')|} + \beta_p I_p, \tag{13}$$

The model which is described by equations (**??**), (**??**), (**??**), (**??**) and (**??**) is called *the general 4-tier ranking model for the ranking authors, publications and conferences problem.* And the model which is described by equations (**??**), (**??**), (**??**), (**??**) and (**??**) is called *the simple 4-tier ranking model for the ranking authors, publications and conferences problem.* Both of the general and simple 4-tier ranking models are really N-tier ranking in which the publication class is the central.

5 Experiments

5.1 Dataset, NPC and SD4R Model

Datasets. We exploited the DBLP data set[2], which provides bibliographic information on major computer science journals and proceedings, to conduct experiment for the simple 4-tier ranking model for the ranking authors, publications and conferences problem on fields of computer science. A publication can be both in a conference and another journal, too. To avoid this gap, we keep only the publications related to conferences and ignore the journals's ones. We has built program to parse DBLP dataset in XML format to extract the authors, title, and publication venue information from the guides [**?**, **?**]. We mention our readers that DBLP has no information about the citations between publications. Finally, we generate two datasets D_1 and D_2 for the experiments. D_1 contains all authors, publications ($\|\mathcal{P}\| = 1253997, \|\mathcal{A}\| = 845295, \|\mathcal{C}\| = 3351$). D_2 contains only the authors, and the publications belong to *not small* conferences which have over 300 publications ($\|\mathcal{P}\| = 1046030, \|\mathcal{A}\| = 753896, \|\mathcal{C}\| = 949$).

NPC Model. Because there is no information about the citations between publications, we adopt a *naive model based on publication counting (NPC)* for ranking author and conference. In NPC model, for each author $a \in \mathcal{A}$ $R'_a(a) ::= |P(a)|$ and for each conference $c \in \mathcal{C}$ $R'_c(c) ::= |P_c(c)|$.

SD4R Model. We implemented a *simple DBLP 4-tier ranking (SD4R) model* with central publication class follow adapted versions of equations (**??**), (**??**), (**??**), (**??**) and (**??**). Because we do not have any information about the citations and for the simplicity, we omitted equations (**??**), (**??**) and set $\alpha_1 = 0, \alpha_2 = \alpha_3$ and $\beta_p = 1 - 2\alpha_2$. The equation (**??**) are rewritten as follows :$\forall p \in \mathcal{P}, c' = C(p)$

$$R_p(p) = \alpha_2 \sum_{a' \in A(p)} \frac{R_a(a')}{|P(a')|} + \alpha_2 \frac{R_c(c')}{|P_c(c')|} + (1 - 2\alpha_2)I_p, \tag{14}$$

5.2 Experimental results

We scale the rank scores of SD4R such that:$\sum_{a \in \mathcal{A}} R_a(a) = \sum_{a \in \mathcal{A}} R'_a(a)$ and $\sum_{c \in \mathcal{C}} R_c(c) = \sum_{c \in \mathcal{C}} R'_c(c)$. For each author $a \in \mathcal{A}$ and conference $c \in \mathcal{C}$, we measure the difference between ranking scores of two models: (i)$\Delta(a) = R_a(a) - R'_a(a)$, $\%\Delta(a) = \frac{\Delta(a)}{R'_a(a)}$ and (ii)$\Delta(c) = R_c(c) - R'_c(c)$, $\%\Delta(c) = \frac{\Delta(c)}{R'_c(c)}$. By examining the Δ and $\%\Delta$ functions, we found following interesting results.

– *SD4R ranking scores reflect how hot the conferences are.* Table **??** has show the top 5 conferences most increasing and decreasing the ranking score over D_2. All *increasing* conferences are young, annual events and get hot topics now. Their topics are about remote sensing (IGARSS - IEEE International Geoscience and Remote Sensing Symposium, from 2005), computer human interaction (CHI - Computer Human Interaction, from 1990), medical image computing (MICCAI - Medical Image Computing and Computer-Assisted Intervention, from 1998), robot (IROS - International Conference on Intelligent RObots and Systems, from 1998), solid-state circuits (ISSCC - International Solid-State Circuits Conference, from 2009)...The top 5 *decreasing* conferences are held for over a long time or biennial events, in local community (GI-Jahrestagung - language spoken is Germany, from 1972) or with less interesting topics such as artificial intelligence (IJCAI - International Joint Conference on Artificial Intelligence, biennial from 1969), image processing(ICIP - International Conference on Image Processing, from 1994; IFIP - International Federation for Information Processing, biennial from 1959), computational linguistics (ACL - Meeting of the Association for Computational Linguistics, from 1979),...

Table 1. Top 5 conferences most differences of ranking value of conferences using SD4R vs. NPC over D_2; (a) increasing and (b) decreasing ranking values

Conference name	nPubs	SD4R val	DIFF val	Δ %	Conference name	nPubs	SD4R val	DIFF val	Δ %
IGARSS	9691	10173	482	4.97	IJCAI	5635	5321	-314	5.57
CHI	8737	9218	481	5.51	ICIP	15125	14821	-304	2.01
MICCAI	3778	4246	468	12.39	IFIP	2796	2529	-267	9.55
ISSCC	1185	1552	367	30.97	GI	4349	4132	-217	4.99
IROS	7906	8218	312	3.95	ACL	3489	3278	-211	6.05

– *SD4R affects much on authors and little on conferences.* The average of $|\%\Delta|$ of the conferences is 2.7%. And the average of $|\%\Delta|$ of the authors is 23%. It implies that SD4R does not change the conference ranking scores but do change the author ranking scores.

– *SD4R reflects the contribution of the author betters than the NPC.* Table **??** has show the top 5 authors most increasing and decreasing the ranking score over D_2. We observe that all top *decreasing* authors are have a large number of publications. Their publications have a large number of co-authors. All top *increasing* authors are really key-person in their research topics.

Table 2. Top 5 authors most differences of ranking value of authors using SD4R vs. NPC over D_2; (a) increasing and (b) decreasing ranking values

Author name	nPubs	SD4R val	DIFF val	Δ %		Author name	nPubs	SD4R val	DIFF val	Δ %
Debenham, J.K.	132	247	115	87.3		Gao, W.	488	367	-121	-121
Tsumoto, S.	204	307	103	50.5		Barolli, L.	283	200	-83	-83
Yama kami, T.	55	156	101	183		Takanishi, A.	166	90	-76	-76
Hancock, E.R.	396	485	89	22.6		Catthoor, F.	241	168	-73	-73
Kamimura, R.	55	139	84	153		Liu, W.	426	354	-72	-72

6 Related work

Link-based ranking has been applied in many application in various domains. In this section, we want to describe such applications and compare them.

- Webpages: One of the most well-known example is PageRank utilized by Google search engine. Since the hyperlink structure among the webpages are easily represented as a web graph, the PageRank of each webpage can be measured (see Sect. **??** for more detail).

- Named entities: In Natural Language Processing (NLP) communities, named entity recognition is an important problem. Ranking scheme has been applied to solve the problem. Collins [**?**] proposes a ranking method based on a maximum-entropy tagger. Also, Vercoustre et al. [**?**] has presented how to apply the ranking method to Wikipedia.

- Scientific articles: In bibliometrics, the scientific articles (e.g., research papers, technical reports, and so on) are evaluated with respect to the quality (e.g., novelty and originality) as well as academic influence to the communities (e.g., impact) by relaying on the citations (e.g., references and quotation) [**?**].

- Researchers: Also, Researchers has been ranked by citation analysis (e.g., how many papers has he/she published, how many times have his/her papers cited, and so on). More interestingly, H-index (Hirsch index) has been designed to measure both the productivity and impact of the published work of the researchers.

So far, conferences have been ranked by subjective opinions and consensus among well known experts in a domain. Such lists in computer science area are compiled here[3]. In this work, we have proposed a novel conference ranking framework to integrate all possible evidence.

7 Conclusion and future works

We have introduced and studied N-tier ranking for mutual ranking systems. The mutual relationships between ranking objects are described by a system of linear equations,

---

[3] http://intelligent.pe.kr/ConfRank.txt

N-mutual ranking system. A N-mutual ranking system is a N-tier ranking systems if it has a core class which affects and reflects all other classes in the system. The rank scores of the N-tier ranking system are unique and computed by a Markov chain. We have pointed out that PageRank is a 2-tier ranking. It has two classes: the webpages (a core class) and links.

We have introduced and studied a general and a simple 4-tier ranking models for ranking authors, publications, conferences. A general model is a generic one. In a simple model, we consider each publication, author, conference, citation is equally. We have conducted the experiments for the models in which the citations are not considered. The experimental results are based on the DBLP dataset. By comparing the difference between the SD4R vs. NPC models, we have shown that our ranking system can reflect how hot the conference are and record the contribution of the authors better than the naive ranking system. Moreover, our ranking system makes a big change on author ranking.

As future work, we are planning to i) get the citations between the publication to upgrade the quality of our ranking system, ii) study how to combine a N-tier ranking systems with a given ranking systems, iii) investigate the time series in N-tier ranking and the trend prediction problem, and iv) apply N-tier ranking systems in various ranking problems, e.g., business ranking, event ranking, and so on.

References

1. S. Brin and L. Page: The anatomy of a large-scale hypertextual web search engine. In Proceedings of the 7th International World Wide Web Conference, 107-117 (1998)
2. L. T. Kien, L. T. Hieu, T. L. Hung, L. A. Vu: MpageRank: The Stability of Web Graph. Vietnam Journal of Mathematics Vol. 37, 475-489 (2009)
3. A. Sidiropoulos, D. Katsaros, Y. Manolopoulos: Generalized Hirsch h-index for disclosing latent facts in citation networks. Scientometrics, Volume 72, Number 2, 253-280 (2006).
4. Michael Ley, Patrick Reuther: Maintaining an Online Bibliographical Database: The Problem of Data Quality. EGC 2006, 5-10 (2006).
5. Michael Ley: DBLP - Some Lessons Learned. PVLDB 2(2), 1493-1500 (2009).
6. T. Furukawa, S. Okamoto, Y. Matsuo, M. Ishizuka: Prediction of social bookmarking based on a behavior transition model. Proceedings of the 2010 ACM Symposium on Applied Computing, 1741-1747 (2010).
7. S. Rendle, C. Freudenthaler, L. S. Thieme: Factorizing personalized Markov chains for next-basket recommendation. Proceedings of the 19th international conference on World wide web, 811-820 (2010).
8. S. Rendle, C. Freudenthaler, L. S. Thieme: Factorizing Markov Models for Categorical Time Series Prediction.
9. Microsoft Corporation: Microsoft Academic Search. http://academic.research.microsoft.com/ (June -26 -2013).
10. Michael Collins: Ranking algorithms for named-entity extraction: boosting and the voted perceptron. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 489-496 (2002)

11. Anne-Marie Vercoustre, James A. Thom, Jovan Pehcevski: Entity ranking in Wikipedia. Proceedings of the 2008 ACM symposium on Applied computing (SAC'08), 1101-1106 (2008).
12. Blaise Cronin: Bibliometrics and beyond: some thoughts on web-based citation analysis. Journal of Information Science, Vol. 27, No. 1, 1-7 (2001).