

On Mining Rating Dependencies in Online Collaborative Rating Networks

Hady W. Lauw¹, Ee-Peng Lim², and Ke Wang³

¹ School of Computer Engineering, Nanyang Technological University, Singapore.

`hadylauw@pmail.ntu.edu.sg`

² School of Information Systems, Singapore Management University, Singapore.

`eplim@smu.edu.sg`

³ Department of Computing Science, Simon Fraser University, Burnaby, BC, Canada.

`wangk@cs.sfu.ca`

Abstract. The trend of social information processing sees e-commerce and social web applications increasingly relying on user-generated content, such as rating, to determine the quality of objects and to generate recommendations for users. In a rating system, a set of reviewers assign to a set of objects different types of scores based on specific evaluation criteria. In this paper, we seek to determine, for each reviewer and for each object, the dependency between scores on any two given criteria. A reviewer is said to have high dependency between a pair of criteria when his or her rating scores on objects based on the two criteria exhibit strong correlation. On the other hand, an object is said to have high dependency between a pair of criteria when the rating scores it receives on the two criteria exhibit strong correlation. Knowing reviewer dependency and object dependency is useful in various applications including recommendation, customization, and score moderation. We propose a model, called *Interrelated Dependency*, which determines both types of dependency simultaneously, taking into account the interrelatedness between the two types of dependency. We verify the efficacy of this model through experiments on real-life data.

1 Introduction

Rating is one of the key ingredients in social web (as well as many non-web settings). Rating plays an important role in influencing decision making and people's choices. In a rating system, reviewers (users) assign rating scores to objects (products, content items, etc.). Multi-criteria ratings on objects in social web have become very common as users often rate objects based on different evaluation criteria, derived from important object features. For example, to evaluate a digital camera on a product review site, a reviewer may give scores to different camera features (e.g., ease of use, battery life, memory size), in addition to an overall score.

By observing the correlation between criterion-level scores and overall scores (or between scores of two different criteria), we can derive insights about the

dependency behavior of reviewers and objects, which are useful in many applications. For example, a recommender system [1] may want to recommend to a reviewer who exhibit high dependency between the memory size and overall scores a camera that has a large memory size.

Given a set of rating data, we seek to determine the *reviewer dependency* of every reviewer, and the *object dependency* of every object from the rating data.

- *Reviewer Dependency.* A reviewer is said to have high reviewer dependency between a pair of criteria when his or her ratings on various objects based on the two criteria exhibit strong correlation. For example, a camera reviewer may value a camera’s memory size so much so that s/he assigns high overall scores to cameras with high memory size scores (also assigned by the reviewer), and low overall scores to cameras with low memory size scores.
- *Object Dependency.* For an object, its object dependency represents the extent to which the object shows correlation in ratings between the two criteria by various reviewers. For example, a well-written conference paper may introduce an idea that is so unique that its reviewers either like it due to the idea being novel (high novelty and overall scores), or dislike it as the idea does not look novel to them (low novelty and overall scores).

A *Naive* approach to determine reviewer dependency and object dependency is to simply apply a standard correlation measure, such as Pearson correlation [2]. For instance, in this approach, a reviewer dependency is equated to the correlation between her scores on the first criterion and her scores on the second criterion. However, this approach fails to recognize the relationship between reviewer dependency and object dependency.

Highly correlated rating scores may be due to reviewer dependency or object dependency. For example, *Naive* may conclude that a reviewer places a premium on a camera’s memory size if her memory size scores correlate with her overall scores (*high reviewer dependency*). However, this would be justified only if the other reviewers of the camera do not show similar correlations (*low object dependency*). Otherwise, the correct conclusion should be that memory size is a dominant selling feature of this camera such that any reviewer is bound to judge it mainly by its memory size (*high object dependency*).

Interrelated Dependency Principle. Therefore, we propose the following principle relating reviewer dependency and object dependency.

- A reviewer has *high reviewer dependency* on a pair of evaluation criteria when his or her rating scores on various objects based on the two criteria exhibit *high correlation*, and these rated objects exhibit *low object dependency* on the same two criteria.
- An object has *high object dependency* on a pair of evaluation criteria when the rating scores it receives from various reviewers based on the two criteria exhibit *high correlation*, and these reviewers exhibit *low reviewer dependency* on the same two criteria.

The rest of the paper is organized as follows. In Section 2, we describe our proposed *Interrelated Dependency* (or *ID*) model to determine reviewer dependency and object dependency between two given criteria. In Section 3, we present experimental results on a real-life dataset. These are followed by an overview of related work in Section 4 and conclusion in Section 5.

2 Interrelated Dependency (ID) Model

Before describing the *Interrelated Dependency* (or *ID*) model, we first review the notations to be used. We model a rating system as a bipartite network with reviewers and objects forming the two distinct sets of entities. A reviewer r_i may assign to an object o_j two rating scores $a_{ij}, b_{ij} \in [0, 1]$ based on two different criteria a and b . Without any loss of generality, the overall score is treated as a criterion. For each r_i , we want to determine r_i 's **reviewer dependency** between a and b , denoted by $d_{r_i}(a, b) \in [0, 1]$. For each o_j , we want to determine o_j 's **object dependency** between a and b , denoted by $d_{o_j}(a, b) \in [0, 1]$. When a and b are implicit, we may further simplify the reviewer dependency and object dependency notations as d_{r_i} and d_{o_j} respectively.

Our proposed *ID* model consists of a pair of equations: Equation 1 to determine reviewer dependency d_{r_i} and Equation 2 to determine object dependency d_{o_j} . To determine d_{r_i} with Equation 1, we compute the correlation observed on r_i 's scores $[\mathcal{F}(\mathbf{a}_{r_i}, \mathbf{b}_{r_i})]$, and reduce it by the aggregate object dependency of r_i 's objects $[\mathcal{A}gg_j(1 - d_{o_j})]$. \mathcal{F} is a correlation measure, which will be described later in this section. \mathbf{a}_{r_i} and \mathbf{b}_{r_i} are the vectors of r_i 's a_{ij} and b_{ij} scores across different o_j 's rated by r_i . $\mathcal{F}(\mathbf{a}_{r_i}, \mathbf{b}_{r_i})$ is higher when there is greater correlation between r_i 's a_{ij} and b_{ij} scores. $\mathcal{A}gg$ represents a function to aggregate $(1 - d_{o_j})$ across various o_j 's. $\mathcal{A}gg_j(1 - d_{o_j})$ is higher when r_i 's rated objects generally have low dependency. Thus, a reviewer has high d_{r_i} if she assigns highly correlated scores on objects with low d_{o_j} .

$$d_{r_i} = \mathcal{F}(\mathbf{a}_{r_i}, \mathbf{b}_{r_i}) \times \underset{j}{\mathcal{A}gg}(1 - d_{o_j}) \quad (1)$$

$$d_{o_j} = \mathcal{F}(\mathbf{a}_{o_j}, \mathbf{b}_{o_j}) \times \underset{i}{\mathcal{A}gg}(1 - d_{r_i}) \quad (2)$$

Equation 2 to determine object dependency is symmetrical to Equation 1. An object has high d_{o_j} if it exhibits highly correlated scores by reviewers with low d_{r_i} . In both Equations 1 and 2, we require $d_{r_i}, d_{o_j} \in [0, 1]$ as well as $\mathcal{F}(\mathbf{a}_{r_i}, \mathbf{b}_{r_i}), \mathcal{F}(\mathbf{a}_{o_j}, \mathbf{b}_{o_j}) \in [0, 1]$.

In this paper, we use the *average* function for $\mathcal{A}gg$, which is suitable as it takes into account d_{o_j} of all the o_j 's concerned. Another possible function is *median*. Examples of unsuitable functions include *summation*, which unfairly penalizes active reviewers (with many objects) and actively-rated objects (with many reviewers); *minimum*, which unfairly penalizes occasional dependency; and *maximum*, which unfairly rewards occasional lack of dependency.

The above equations are based on the principle of interrelated dependency between d_{r_i} and d_{o_j} . To determine d_{r_i} , we need to know d_{o_j} (vice versa). Moreover, d_{r_i} and d_{o_j} are inversely related, with d_{r_i} higher when d_{o_j} is lower. The reviewer dependency of various reviewers are related by co-rating common objects, and the object dependency of various objects are related by having common reviewers. As reviewers and objects are inter-connected within the rating data, the interrelatedness extends to all reviewers and objects, which means all d_{r_i} 's and d_{o_j} 's need to be solved simultaneously. Given *average* as the *Agg* function, the set of Equation 1 for each r_i and Equation 2 for each o_j forms a system of linear equations. Such systems can be solved using either linear algebra or iterative methods, provided the system is uniquely determined [3].

To determine reviewer dependency, we rely more on objects with low dependency. Ideally we should have as many low dependency objects per reviewer as possible. Given that low dependency objects are usually the majority, a reviewer will likely rate more low dependency objects when s/he has rated many objects. Correlation measure is also generally more meaningful for larger "sample" size.

Correlation Measure. \mathcal{F} measures the correlation between two sets of scores. The measure we use in this paper is a slightly modified version of the *Pearson product-moment correlation coefficient* [2]. For two vectors \mathbf{a} and \mathbf{b} of N real-valued elements each, the correlation coefficient $\mathcal{F}(\mathbf{a}, \mathbf{b}) \in [0, 1]$ is determined by Equation 3. $\mathbf{a}[n]$ is the n^{th} element of \mathbf{a} , while $\mu_{\mathbf{a}}$ and $\sigma_{\mathbf{a}}$ are the mean and the standard deviation of \mathbf{a} 's elements respectively. Similar notations apply to \mathbf{b} . Higher $\mathcal{F}(\mathbf{a}, \mathbf{b})$ value indicates greater correlation between \mathbf{a} and \mathbf{b} . This measure is also symmetric, i.e., $\mathcal{F}(\mathbf{a}, \mathbf{b}) = \mathcal{F}(\mathbf{b}, \mathbf{a})$.

$$\mathcal{F}(\mathbf{a}, \mathbf{b}) = \left| \frac{\sum_n^N (\mathbf{a}[n] - \mu_{\mathbf{a}})(\mathbf{b}[n] - \mu_{\mathbf{b}})}{(N - 1) \times \sigma_{\mathbf{a}} \times \sigma_{\mathbf{b}}} \right| \quad (3)$$

This measure is slightly different from the original *Pearson correlation*. While *Pearson correlation* ranges from -1 to 1, Equation 3 ignores the sign, thus confining the range to 0 to 1. Secondly, *Pearson correlation* is undefined for $\sigma_{\mathbf{a}} = 0$ or $\sigma_{\mathbf{b}} = 0$. We choose to define $\mathcal{F}(\mathbf{a}, \mathbf{b}) = 0$ for these cases. $\sigma_{\mathbf{a}} = 0$ or $\sigma_{\mathbf{b}} = 0$ happens when either \mathbf{a} or \mathbf{b} has uniform elements, in which case the two vectors can reasonably be considered independent. $\mathcal{F}(\mathbf{a}, \mathbf{b})$ is high when elements of \mathbf{a} and \mathbf{b} rise and fall together, and low when they do not vary or vary independently.

Other possible correlation measures include mutual information [4], or rank correlations [5]. We will study these alternatives as part of future work.

Naive Model as a Special Case. A suitable baseline alternative to the *ID* model should be the basic model which does not incorporate the interrelated dependency between d_{r_i} and d_{o_j} . This baseline model, called the *Naive* model, consists of the pair of Equation 4 to determine d_{r_i} and Equation 5 to determine d_{o_j} . *Naive*'s equations compute d_{r_i} and d_{o_j} respectively using only the correlation measure $\mathcal{F}(\mathbf{a}, \mathbf{b})$. Given *average* as the *Agg* function, *ID* would degenerate into *Naive* when all d_{o_j} 's (or d_{r_i} 's) are uniform, in which case *ID*'s Equation 1 (or Equation 2) practically degenerates into *Naive*'s Equation 4 (or Equation 5).

$$d_{r_i} = \mathcal{F}(\mathbf{a}_{r_i}, \mathbf{b}_{r_i}) \quad (4)$$

$$d_{o_j} = \mathcal{F}(\mathbf{a}_{o_j}, \mathbf{b}_{o_j}) \quad (5)$$

3 Experiments

Our experimental objective is to verify the effectiveness of the proposed model on a real-life dataset. First, we compare the reviewer dependency and object dependency ranks produced by the *Naive* and *ID* models. Next, we highlight a specific case example that illustrated how the contribution of interrelated dependency principle that makes the *ID* model more effective.

3.1 Dataset

Our dataset was obtained from *Epinions*, a product review Web site. We crawled the site over three days (April 20-23, 2007), beginning from a seed page⁴. The collected dataset consisted of a subset of all products, reviewers, and scores.

Table 1. Data Size

	<i>Reviewers</i>	<i>Objects</i>	<i>Overall Scores</i>	<i>Weight Scores</i>
<i>Original</i>	781	428	2288	2087
<i>Filtered</i>	102	184	1163	1163

In this dataset, each product belongs to a category and each category has its specific rating criteria. Each reviewer may assign one or more scores based on these criteria. These scores, which ranged from 1 to 5, were normalized to a range from 0.2 to 1 by a simple division by 5. We chose to focus on the *Beers* category, as within the collected dataset, this category was relatively large and had more active reviewers and more actively-rated objects. There are four rating criteria for *Beers* category: *Overall*, *Weight*, *Flavor*, and *Complexity*. In the following experiments, we mainly focus on the (*Overall*, *Weight*) pairing. We have also separately carried out other experiments with (*Overall*, *Flavor*) and (*Overall*, *Complexity*) pairings with similar results. For the (*Overall*, *Weight*) pairing, the original data size collected is shown in the first row of Table 1.

We further filtered the dataset to make it more suitable for experiments. First, we considered a reviewer as having rated an object only if the reviewer had assigned both *Overall* and *Weight* scores. Second, we ensured that each reviewer must have at least 3 objects and each object must have at least 3 reviewers, by iteratively removing reviewers and objects not meeting the condition until the condition was met. The final data size is shown in the second row of Table 1.

⁴ http://www.epinions.com/rest-Restaurants-All/ViewAll_~1/Grp_~14529

3.2 Rank Comparison

This part of the experiment compares the ranked lists produced by *Naive* and *ID* for the *(Overall, Weight)* pairing. We construct a ranked list of reviewers and a ranked list of objects for each solution. Reviewers and objects are ranked in decreasing order of d_{r_i} and d_{o_j} respectively. The highest d_{r_i} or d_{o_j} value is given rank 1. Same values share the same rank.

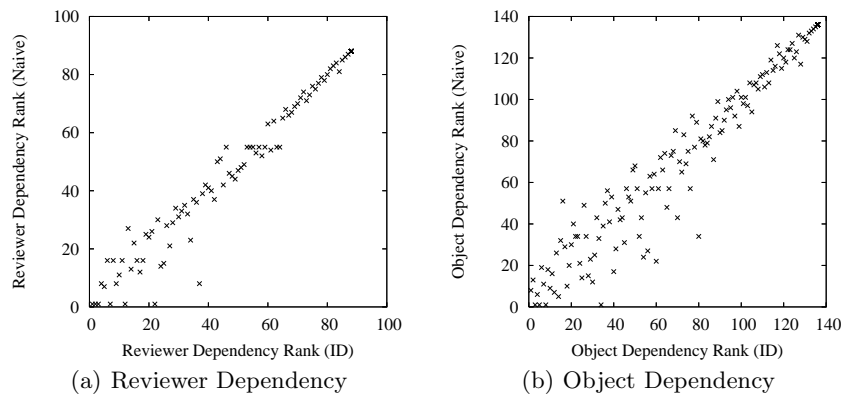


Fig. 1. Rank Scatterplots: *ID* vs. *Naive*

Figure 1(a) is a scatterplot of reviewer dependency ranks. A point in each scatterplot represents a reviewer, with the x -value being the rank assigned by *ID* and the y -value being the rank assigned by *Naive*. The reviewer dependency ranks assigned by *ID* and *Naive* are positively correlated in general, but are not identical. The positive correlation is expected as both are based on the same \mathcal{F} function. However, *ID* also takes into account object dependency, resulting in a different ranking. Figure 1(a) shows large variances around the diagonal, indicating that *ID* and *Naive*'s rankings are quite different. For example, there are instances where a significant number of reviewers are tied according to *Naive*, but are differentiated by *ID*. In Figure 1(a), 6 reviewers share rank 1 by *Naive*, but are given ranks ranging from 1 to 22 by *ID*.

Figure 1(b) is the corresponding scatterplot for object dependency ranks. As this figure shows much resemblance to the earlier Figure 1(a), similar observations as for reviewer dependency ranks can also be made for object dependency ranks. This itself is an important observation, for it highlights the interrelatedness between reviewer dependency and object dependency. If *ID*'s object dependency ranks were identical to *Naive*'s, so would the reviewer dependency ranks be. Instead, the variance in reviewer dependency ranks results from the variance in object dependency ranks (and vice versa).

3.3 Case Example

Object *boulevard_dry_stout*'s profile is given in Table 2(a). *Naive* assigns this object high object dependency of 1.00 (rank 1). Meanwhile, *ID* assigns it much lower object dependency of 0.33 (rank 34). The rating scores assigned by its three reviewers indeed suggest high object dependency. However, Table 2(a) also shows that *boulevard_dry_stout*'s three reviewers (*lafeet*, *wingdman*, *impydykiechick*) all have high reviewer dependency. For instance, *wingdman* has high reviewer dependency of 0.71 (rank 17). We take a more detailed look at *wingdman*'s rating scores in Table 2(b). Visual inspection reveals a trend whereby *wingdman*'s Overall (*Ove*) score is high (or low) when his Weight (*Wei*) score is high (or low). Given the high reviewer dependency of its reviewers, it is expected that *boulevard_dry_stout* would show high correlation between its *Ove* and *Wei* scores. However, this correlation should be attributed more to its reviewers. Thus, *ID* justifiably assigns lower object dependency (and ranks) to *boulevard_dry_stout*.

Table 2. Case Example

(a) Profile of Object <i>boulevard_dry_stout</i>					(b) Objects reviewed by <i>wingdman</i>			
<i>Object</i>		<i>d_{o_j}</i> (rank)			<i>Objects</i>		<i>Score</i>	
		<i>Naive</i>	<i>ID</i>				<i>Ove</i>	<i>Wei</i>
boulevard_dry_stout		1.00 (1)	0.33 (34)		beamish_irish_stout	1.0	1.0	
<i>Reviewers</i>		<i>d_{r_i}</i> (rank)			boulevard_bully_porter		1.0	1.0
		<i>Ove</i>	<i>Wei</i>	<i>Naive</i>	<i>ID</i>	boulevard_dry_stout		1.0
lafeet	1.0	1.0	0.83 (25)	0.69 (19)	dixie_blackened_woodoo_lager	1.0	1.0	
wingdman	1.0	1.0	0.91 (12)	0.71 (17)	sierra_nevada_stout	1.0	1.0	
impydykiechick	0.8	0.8	0.84 (23)	0.60 (34)	grolsch_premium_lager	0.8	0.6	
					hacker-pschorr-weisse	0.6	0.6	
					murphy_s_irish_amber	0.6	0.6	
					samuel_adams_summer_ale	0.6	0.2	
					tequila_beer	0.6	0.2	

4 Related Work

The work on *systematic* dependency uses correlation analysis [6] to test the dependency between a given factor and the assigned scores, which affect all reviewers and objects in general. For example, [7] investigates whether reviewers' scores on peer review submissions are dependent on the reviewers' also being authors; [8] studies whether students tend to assign higher scores to instructors who have given them higher grades; [9] shows that venture capitalists' evaluation of start-up teams is dependent on their similarity to these teams. In contrast, we are concerned with measuring the *specific* dependency of individual reviewers and objects, as reviewers and objects may exhibit varying dependency with respect to two given rating criteria.

The problem of multiple criteria decision making (MCDM) [10] is concerned with making an optimal decision, in view of two or more potentially conflicting criteria. The optimality of a decision necessarily depends on the preferences of the

decision-maker. For example, a user may wish to select among several hotels one that scores high on both *luxury* and *affordability*. This may involve comparing in some way the scores assigned to an object on various criteria in aggregate. This problem is different from our problem in that the optimal decision is not necessarily the object with the highest or lowest dependency.

5 Conclusion

In this paper, we address the problem of determining reviewer dependency and object dependency with respect to two rating criteria. The key principle in our *ID* model is modeling the interrelatedness between reviewer dependency and object dependency, which is ignored by the *Naive* model. Experiments on real-life data show that our approach is more effective than *Naive*. Moreover, although our modeling is based on two given rating criteria, without loss of generality, it can be applied in a pairwise manner for multiple criteria, for instance in learning of the dominant rating criteria affecting the overall scores.

Acknowledgment

This work was supported by A*STAR Public Sector R&D, Singapore, Project Number 062 101 0031.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* **17**(6) (2005) 734–749
2. Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K.: *Probability & Statistics for Engineers & Scientists*. 7th edn. Prentice-Hall, Inc. (2002)
3. Anton, H., Dorres, C.: *Elementary Linear Algebra with Applications*. John Wiley & Sons, Inc. (1987)
4. Dionisio, A., Menezes, R., Mendes, D.A.: Mutual information: A measure of dependency for nonlinear time series. *Physica A* **344** (2004) 326–329
5. Fagin, R., Kumar, R., Sivakumar, D.: Comparing top k lists. *SIAM Journal on Discrete Mathematics* **17**(1) (2003) 134–160
6. Cohen, J., Cohen, P., West, S.G., Aiken, L.S.: *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd edn. Lawrence Erlbaum Associates, Inc. (2003)
7. Blackburn, J.L., Hakel, M.D.: An examination of sources of peer-review bias. *Psychological Science* **17**(5) (2006) 378–382
8. Greenwald, A.G., Gillmore, G.M.: Grading leniency is a removable contaminant of student ratings. *American Psychologist* **52**(11) (1997) 1209–1217
9. Franke, N., Gruber, M., Harhoff, D., Henkel, J.: What you are is what you like – similarity biases in venture capitalists’ evaluations of start-up teams. *Journal of Business Venturing* **21**(6) (2006) 802–826
10. Figueira, J., Greco, S., Ehrgott, M., eds.: *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer Science+Business Media, Inc. (2005)