Validation-based normalization and selection of interestingness measures for association rules

WALEED A. ALJANDAL	WILLIAM H. HSU	VIKAS BAHIRWANI	DOINA CARAGEA	TIM WENINGER

Department of Computing and Information Sciences
Kansas State University
Manhattan, KS, USA
{waleed, bhsu, vikas, dcaragea, weninger}@cis.ksu.edu

Abstract – We investigate the problem of tuning and selecting among interestingness measures for association rules. We first derive a parametric normalization factor for such measures that addresses imbalanced itemset sizes, and show how it can be generalized across many previously derived measures. Next, we develop a validation-based framework for both the normalization and selection tasks, based upon mutual information measures over attributes. We then apply this framework to market basket data and user profile data in weblogs, to automatically choose among or fine-tune alternative measures for generating and ranking rules. Finally, we show how the derived normalization factor can significantly improve the sensitivity of interestingness measures when used for pure association rule mining and also for a classification task. We also consider how this data-driven approach can be used for fusion of association rule sets: either those elicited from subject matter experts, or those found using prior background knowledge.

INTRODUCTION

One of the most important aspects of association rule mining is ranking rules by their significance, according to some quantitative measure that expresses their interestingness with respect to a decision support or associative reasoning task. Rules take the form X → Y, where both X and Y are subsets of an observed itemset L = $\{I_1, I_2, \ldots, I_k\}$. Two well-known measures for association rule interestingness are the support, P(X) and the confidence, P(Y | X). These probabilistic measures have been used with other statistical formulae to derive compound measures used in discovering the most significant rule. One limitation of existing binary measures of rule interestingness is that they do not account for the relative size of the itemsets to which each candidate pair of associated subsets (X, Y) belongs. Moreover, there are some hidden associations related to candidates appearing in small groups. Thus, giving some attention and weight to these small groups may lead us to a different relationship perspective. This kind of data behavior can be seen, for example, in social network data where each user record consists of features such as interests, communities, schools attended, etc. In particular, user's list of interests, each of which corresponds to a list of interest holders. Some interests such as "DNA replication" have low membership; whether this is because the interests are less popular or more specialized, it often suggests a more significant association between users naming them than between those who have interests such as "Music" or "Games" in common. In general, an extremely large number of interest holders tends to correspond to a more tenuous link. The size of the itemsets produces further information that can be used to increase the sensitivity of measures applicable to a candidate association.

In this paper, we propose an itemset size-sensitive joint probability estimator by derving a normalization factor which is expressed by the size n_i of each itemset L_i to which the rule antecedent X and consequent Y belong. The new size-sensitive measure is flexible and applicable to a wide range of previously developed interestingness measures. The advantage of itemset size normalization factor is more prominent in domains where common occurrence of sets X and Y within very large itemsets are poor predictors of true association between X and Y.

OBJECTIVE INTERESTINGNESS MEASURES

Deriving an objective interestingness measure usually involves estimating some aspect of a candidate rule's structure, analytical performance and statistical significance with respect to observed itemset data. Compound measures are based on primitive measures grounded in probability density functions, with some – such as the normalization approach described in this paper – based on parametric fusion of these primitive measures, while others are based on more *ad hoc* rules of combination.

Piatetsky-Shapiro (1991) first proposed using statistical independence of rules as an interestingness measure. More methods have since been proposed using different statistical approach. Brin, *et al.* (1997) proposed lift and $\chi 2$ (chi-squared) as correlation measures and developed an efficient mining method. Hilderman, *et al.* (2001) and Tan, *et al.* (2002) have comparative studies of different interestingness measures and address the concept of null-transactions. Since the probability of an item appearing in a particular transaction is usually very low, it is desirable that a correlation measure should not be influenced by these transactions which they call it "null-transactions".

Three measures for capturing relatedness between item pairs are proposed by Shekar (2004). These measures use the concept of function embedding to appropriately weigh the relatedness contributions due to Mutual Interaction, complementarity and substitutability between items. At the end they propose interestingness coefficient by combining the three relatedness measures. All the three measures are calculated based on the probability without taking into account the transaction itself (large or small).

Following these studies, Tan, Kumar and Srivastava (2002) discussed the properties of twenty-one objective interestingness measures and analyze the impacts of support-based pruning and contingency table standardization. This study ends with conclusion that there is no measure that is consistently better than others in all application domains. However, using the new concept of itemset size-sensitive joint probability change the way of how these measures capture the co-occurrence membership relation.

Table 1 Probability based objective interestingness measures

Measure	Formula
Accuracy	$P(AB) + P(\neg A \neg B)$
Lift/Interest	$P(B\|A)/P(B)$ -or- $P(AB)/P(A)P(B)$
Leverage	$P(B\|A) - P(A)P(B)$
Relative Risk	$P(B\|A)/P(B\backslash\neg A)$
Jaccard	$P(AB)/(P(A) + P(B) - P(AB))$
Certainty Factor	$\frac{P(B\|A) - P(B)}{1 - P(B)},$

From the previous studies we can see how the Probability Based objective interestingness measures (some of them in Table 1) contain joint probability (co-occurrence) as important part which may affect measurement value. However, in the calculation of joint probability, items relations in the dataset tuples are treated equally in all interestingness measures without differentiation between one tuple and others. Even if some of interestingness measures adopt attributes values to adjust the final result, there is still no change in the value of the joint probability. We are looking for making the joint probability more sensitive to items relation with each tuple using itemset size to reflect the real relation between items which is going to be the first step to make the interestingness measures more promising.

ITEMSET SIZE-SENSITIVE JOINT PROBABILITY

Normalization knowledge has been used to discover the correlation between a single numeric feature and multiple intermediate concepts. This concept will make the difference of the results' order which improves the measurer quality. Normalization Knowledge reduce unrelated correlation making axis-parallel division in the instance space more useful (Steven 1996).

The main concept of the current objective measures is based on the probability. When Hilderman, *et al.* (1998) proposed a concept of share-confident and support, they believe that involve the quantity and price of the items in the confident and support computation will improve the measurement quality.

Statistical joint probability has been used as important part in the interestingness measures. In the data mining applications, this may not reflect the true relationship interpretation between some items if there is a large variance among the number of items in each tuple. Frequent items in small-sized item sets may be more informative about the relationship between their constituent items than the large ones. We now consider how to extend the interestingness factor to take in account the size of the tuple to construct a new concept of size-sensitive probability. Let m be a constant such as the minimum given tuple size (we can also use a trim-mean[1]), so for each tuple with size n_i there is a real number $C_i \geq 1$ such that:

$$m^{c_i} = n_i$$

$$m = \sqrt[c_i]{n_i} = n_i^{\frac{1}{c_i}} \therefore m \propto C^{-1} \qquad (1)$$

Let $R_i = 1 / C_i$. Then:

$$m = n_i^{R_i}, \qquad 0 < R_i \leq 1$$

$$R_i = \frac{\log m}{\log n_i} \quad \text{----->} \quad R_i = \log_{n_i} m \qquad (2)$$

R represents a relational factor that describes the relationship between m and the size n_i of each tuple. Moreover, the value of R_i will become more efficient if we involve the number of target items in the equation. For example, if the target items are (x_1, x_2, x_3) the value of R_i for tuple K should be slightly larger than the value of R_i of the same tuple when the target items are (x_1, x_2). Therefore, the q^{th} root is used to adjust the value of R_i based on the number of target items q in $X \cup Y$. Then:

$$R_{i_q} = \sqrt[q]{\log_{n_i} m} \qquad (3)$$

If we consider R_{i_q} in calculating the joint probability, we can define an *itemset size-sensitive* joint probability. Let $L \equiv \{x_1, x_2 \ldots x_k\}$ be the set of items. Let D be a set of transactions ($|D| = N$), where each transaction T is a set of items such that $T \subseteq L$. Then:

---

[1] Trim-mean is the average which can be obtained by trimming the largest and the smallest cretin percentage (this percentage can vary) of the numbers in a series and then calculating the arithmetic mean for the remaining numbers.

$$\hat{p}(x_1, x_2, \ldots, x_q) \triangleq \frac{1}{N}\sum_{i=1}^{N} R_{i_q} \triangleq \frac{1}{N}\sum_{i=1}^{N} \sqrt[q]{\log_{n_i} m} \tag{4}$$

The lower bound of this equation is achieved where the number of items q in one tuple is 2, which is also the smallest possible tuple size.

From Equation 3, Figure 1 illustrates the curve of the normalization factor under three assumptions of the size of the target items. When the size of tuple n_i = 5 the value of R_{i_q}= 1 (the max for R_{i_q}) which is exactly equal to the normal value when we compute the normal joint probability. Moreover, the R_{i_q}has correlated relation with the size of target items. We can see from Figure 1 that the value of R_{i_q} increased with the increasing of the number of target items q which makes R_{i_q} get close to 1. This can give a logical explanation of the relation and interpretation of the tuple size and the target items' size in the new normalization factor equation.

Based on Equation 3, let m = 5 and $5 \leq n_i \leq 100$, q = {2, 3, 4}

CONCEPT HIERARCHY (ONTOLOGY)

A concept hierarchy or ontology is an explicit description (similar to the formal specification of a program) of the concepts and relationships that exist in a domain (Gruber 1994). Ontologies can be seen as metadata that are used provide a better understanding of the data. In social networks, ontologies can provide a crisp semantic organization of the knowledge available in the domain. In particular the interest ontology can be used to make explicit the relationships between various interests, thus helping in the process of understanding the data. Moreover, the accuracy can be improved if an interest ontology is exploited (Bahirwani, *et al.* 2008) when constructing features using association rule measures.

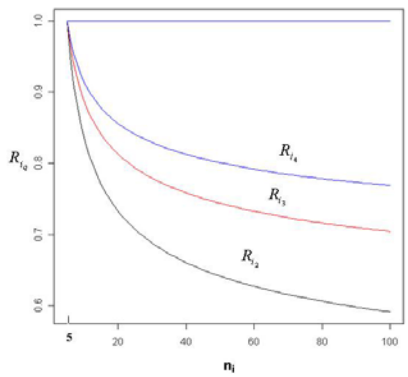

Fig. 1 The normalization factor Curve

EXPERIMENTAL RESULTS

In this section we illustrate the result of three experiments. First one uses original lift measures of users' common interests as friendship prediction feature and compares it with the normalized lift (using size-sensitive probability). The second experiment demonstrates the improvement of the classification measures when we use some interestingness measures for users' common communities as new features with graph features as were used in Hsu et al. (2007). The last experiment shows the advantage of using an ontology and the effect on the classification result of the normalized and unnormalized measures in small datasets

First Experiment: Link Mining in Social Networks using User Interests. The first experiment presents the results of classification using some measures with and without the normalization factor for users' common interest on the *LiveJournal* data set.

This experiment uses users' common interest measures to predict link existence (friendships). This dataset was developed by Hsu *et al.* (2007) for link prediction based on graph features where they found that using mutual interests alone results in relatively poor prediction accuracy. Uncategorized interests in *LiveJournal* (each user indicates his/her own interests) increase the weakness of the mutual interests feature because of misspellings, or the addition of stop words such as “the” or “of”, or by adding symbols such as underscores. However, by using our new normalization factor we show that we can improve prediction results over previously poor results (Claim 1).

To evaluate our new normalization factor, we designed a three phase framework as shown in Figure 2 for comparing the classification results of two different methods i.e. Interestingness Measure (IM) withR and withoutR (for user-interests information).

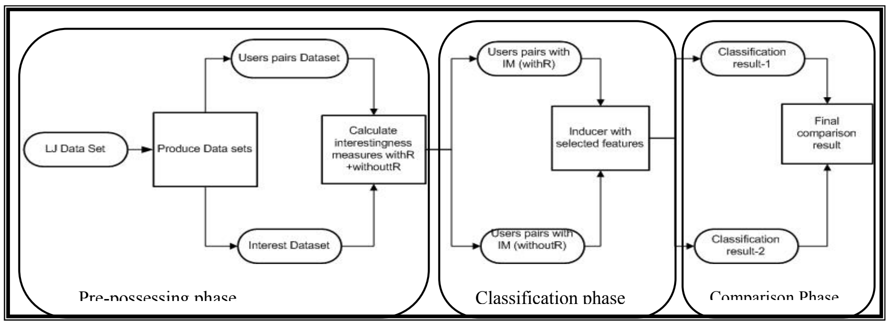

Figure 2: Three phases of evaluation process (using Classification)

We produced two random 10,000 user-pair datasets: one with original measures (support, lift) features and another with normalized measures features (normalized support, normalized lift) where the number of actual friends is ≈2.2%. These files contain the following attributes: identification numbers of *u and v*, Support and Lift of common interests (for the rule $v \rightarrow u$, which the same as $u \rightarrow v$) .

Claim (1):

From the *LiveJournal* dataset we can construct feature baskets $\boldsymbol{B_i}$ such that
$B_i \;\; 1 \leq i \leq n_\tau \quad where\ \tau \in \{interests, communties, schools \ldots.\}$
$u \overset{c_j}{\rightarrow} v \quad For\ high\ c_j\ given\ \{B_i\} \Rightarrow (u,v) \in E\ with\ high\ probability$
where c_j is association rule measure for some B_i
E is a set of connected user pairs (u, v) represent the actual friendship relation.

Table 2 presents results using two different inducers: Random Forest and IB1. In each case, the normalization factor boosted the accuracy measures, which was a result of improving the sensitivity of interestingness measures when used as features for link prediction.

Table 2: Classification results (10-fold CV), with 10,000 user pairs

Inducer	Measure	Accuracy	Precision	Recall	F-measure
IB1	Normalized lift	77.56%	0.491	0.532	0.510
	Original lift	74.2%	0.418	0.438	0.428
Random Forest	Normalized support	80.41%	0.599	0.333	0.428
	Original support	78.73%	0.659	0.070	0.126
IBl	Normalized support	76.02%	0.450	0.401	0.424
	Original support	68.59%	0.275	0.260	0.267

This improvement was achieved across all accuracy measures, with different ranges of improvement depending on the base inducer used except the precision of Random Forest which has recovered by higher improvement in Recall and F-measures where the last one is a combination of precision and recall. For example, the best accuracy improvement in this experiment occurred when we used the IB1 inducer with one attribute, support of common interests. This attribute improves classification accuracy by 10.83% (from 68.59 to 76.02).

In the classification process we collected accuracy measures for each 10-fold cross-validated run to illustrate the significance of normalized measures. In the detailed results of IB1 with one attribute (lift), all measures (Precision, Recall and F-measure) are improved using normalized lift. Moreover, we used a *T*-test to evaluate the significance of the results at 95% level of confidence (the alpha level is 0.05). Table 3 shows the p-value of the T-test results for validation set precision, recall and F-measure.

Table 3: p-value of T-test for precision, recall and f-measure on IB1

	Precision	*Recall*	*F-measure*
P(T<=t) one-tail	3.87E-05	1.51E-05	1.05E-05

In all measures, the results of T-test show that there is a significant improvement ($t < .05$) when we use normalized lift.

Second Experiment: Link Mining with Graph Features and Community Associations. The second experiment shows that by selecting proper interestingness measures we can improve the link prediction. In previous research, Hsu, *et al.* (2007) found that using mutual interests as the only features results in very poor prediction accuracy using any inducer, because of the limited information contained in interests alone. In this experiment, we conceder all graph features with the user communities membership information measures as new feature. Table 4 shows the J48 classification results for 5,980 user pairs with 10-fold cross validation.

Table 4: Result of accuracy, precision, recall, F-measure for J48

Feature	Accuracy (%)	Precision	Recall	F-measure
GF	92.977	0.932	0.903	0.918
GF+AR	93.780	0.936	0.919	0.927
GF + N-AR	94.081	0.941	0.921	0.931

(GF: Graph Feature, AR: Interestingness Measures, N-AR: Normalized Interestingness Measures)

The classification results show how link prediction improved when we used the interestingness measures of user interests as new features. This improvement was further augmented using normalized interestingness measures.

In the classification process we collected accuracy measures for each 10-fold cross-validations run to illustrate the significance of normalized measures. In the detailed results of J48 with different selection of features (GF and GF+N-AR), a significant improvement was observed across all measures (Accuracy, Precision, Recall and F-measure), especially when we use the normalized interestingness measures.

We again use a *T*-test to evaluate the significance. Table 5 shows the p-value of T-test result for accuracy, precision, recall and F-measure.

For all measures, the results of the *T*-test reflect a significant improvement attained by using normalized association rules measures with graph features.

Table 5: P-value of T-test for accuracy, precision, recall and f-measure on J48

	Accuracy	*Precision*	*Recall*	*F-measure*
P(T<=t) one-tail	2.81E-03	3.84E-02	4.85E-03	2.54E-03

Third Experiment: Ontology-Based Refinement of User Interests. Our third experiment addresses the prediction of friendships in *LiveJournal* using association rule based measures for users' common interests with the use of the ontology. In related work, Bahirwani *et al.* (2008) have implemented a hybrid clustering algorithm, HAD, to automatically extract the concept hierarchy of interests. Like mentioned earlier, the accuracy of predicting friendship links in a social network, for instance, in absence of graph features is very low (Hsu, *et al.* 2006). In the paper (Bahirwani, *et al.* 2008), explore how ontologies can be used to improve this performance.

We are going to use eight normalized association rule measures as new friendship prediction features (plus number of common interest): Support, Confidence, Confidence, Lift, Conviction, Match, Accuracy, and Leverage.

We use six different inducers (we show only three because of space in Table 6 and Table 7). In this experiment, the training and test data set consists of 1000 user pairs. Training data set consists of about 50% friend pairs and 50% non-friend pairs while the test data consists of randomly selected user pairs to preserve the original distribution of positive-negative instances in *LiveJournal*.

Table 6: Classification result- without ontology (for normalized and unnormalized)

Inducer	Method	Accuracy (%)	Precision	Recall	F-Measure	ROC
Random Forest	unnormalized	67.5	0.015	0.556	0.030	0.688
	normalized	65.3	0.014	0.556	0.028	0.605
Logistic	unnormalized	74.4	0.019	0.556	0.038	0.678
	normalized	85.5	0.034	0.556	0.065	0.68
ADTree	unnormalized	73.7	0.019	0.556	0.037	0.671
	normalized	78.8	0.023	0.556	0.045	0.694

Table 7: Classification result- with ontology (for normalized and unnormalized)

Inducer	Method	Accuracy (%)	Precision	Recall	F-Measure	ROC
Random Forest	unnormalized	67.6	0.018	0.857	0.036	0.773
	normalized	70	0.020	0.857	0.038	0.829
Logistic	unnormalized	86.8	0.037	0.714	0.070	0.912
	normalized	89.7	0.056	0.857	0.104	0.894
ADTree	unnormalized	77.8	0.026	0.857	0.051	0.90
	normalized	82.7	0.034	0.857	0.065	0.925

Even though in this experiment we use only 1000 user pairs, the normalized measures improve the classification measurers when we use the ontology. These normalized measures take into account the popularity that particular interests hold in common, where the most popular interests (held by a significant proportion of users) being slightly less revealing than rarer interests. Furthermore, we investigate how the ontology improves the classification measurers especially when we use normalized measures which boost the measures sensitivity regarding to interest popularity. When computing the measures, we modify the interests of users by viewing the interests at the "best" level of abstraction as suggested in by Bahirwani, *et al.* (2008). This process will give some enhancement to the normalized measures apically with small datasets.

Table 6 and Table 7 summarize the expected improvements. For example, without an ontology, better results are observed for unnormalized measures using the *Random Forest* inducer (in most of the classification measures) but when we modify the data according to a concept hierarchy, the improvement using normalized measures consistently exceeds that achieved using unnormalized measures (even with the small dataset).

CONCLUSIONS

In this paper, we showed that normalized measures (using itemset size-sensitive joint probability) increase the sensitivity of the interestingness measure to the distribution of data in the context of item set size, thus improving upon measures such as unnormalized lift. We have used this method with several datasets, and have obtained statistically significant improvements in link prediction problems using classification methods.

In future work we will investigate the impact of this normalization approach on other interestingness measures, datasets, and association and classification tasks. In addition, we plan to apply ontologies with graph features and both user interest *and* community associations to larger datasets. We hypothesize that this will reduce semantic ambiguities in the relational data model of user profile data, which may lead to further improvement in link prediction.

ACKNOWLEDGEMENTS

The authors wish to thank Jing Xia for his assistance with this paper.

REFERENCES

Bahirwani, Vikas, Waleed Aljandal, William H. Hsu, and Doina Caragea. "Ontology Engineering for Social Network Data Mining." Under Preparation. 2008.

Blanchard, J, F Guillet, R Gras, and H Briand. "Using information-theoretic measures to assess association rule interestingness." *Proceeding of the 2005 international conference on datamining (ICDM'05).* Houston, TX, 2005. 66–73.

Brin, S, R Motwani, JD Ullman, and S Tsur. "Dynamic itemset counting and implication rules for market basket analysis." *Proceeding of the 1997 ACM-SIGMOD international conference on management of data (SIGMOD'97).* Tucson, AZ, 1997. 255–264.

Geng, L., and H.J. Hamilton. " Interestingness measures for data mining: A survey." *ACM Comput. Surv.* Sep. 2006. 38, 3.

Gruber, Thomas R. "Towards Principles for the Design of Ontologies Used for Knowledge Sharing." *International Journal of Human and Computer Studies.* 1994. 907-928.

Hilderman, RJ, and HJ Hamilton. *Knowledge discovery and measures of interest.* Boston : Kluwer Academic, 2001.

Hilderman, Robert, Colin Carter, Howard J Hamilton, and Nick Cercone. "Mining Market Basket Data Using Share Measures and Characterized Itemsets." *Pacific-Asia Conference on Knowledge Discovery and Data Mining.* 1998.

Hsu, W. H., J. Lancaster, M. S. R. Paradesi, and T. Weninger. "Structural link analysis from user profiles and friends networks: A feature construction approach." *ICWSM-2007.* Boulder, CO, 2007. 75-80.

Hsu, William H, Andrew L. King, Martin S. R. Paradesi, Timothy Weninger, and Tejesvi Pydimarri. "Collaborative And Structural Recommendation Of Friends Using Weblog-Based Social Network Analysis." *AAAI '06: Proceedings of Computational Approaches to Analyzing Weblogs - AAAI.* 2006.

Piatetsky-Shapiro, G. " Notes of AAAI'91 workshop knowledge discovery in databases (KDD'91)." *AAAI/MIT Press.* Anaheim, CA, 1991.

Shekar, B., and R. Natarajan. "A Transaction-based Neighborhood-driven Approach to Quantifying Interestingness of Association Rules." *In Proc. Fourth IEEE Int. Conf. on Data Mining.* 2004.

Steven, Donoho. "Knowledge-Guided Constructive Induction." *Ph.D. thesis.* University of Illinois, 1996.

Tan, P-N, V Kumar, and J Srivastava. "Selecting the right interestingness measure for association patterns." *Proceeding of the ACM SIGKDD international conference on knowledge discovery in databases (KDD'02).* Edmonton, Canada: ACM SIGKDD, 2002. 32–41.