

Ranking Based Feature Selectors for Effective Data Classification

¹B. Amarnath and ²S. Appavu Alias Balamurugan

¹Research Scholar, M.S.University, Tirunelveli, India

²Professor and Head/I.T, KLNCIT, Sivagangai, India

Abstract: Ranking is the attribute selection technique used in the pre-processing phase to emphasize the most relevant attributes which allow models of classification simpler and easy to understand. In recent years, there has been growing interest in learning to rank. It is a very important and a central task for information retrieval, such as web search engines, recommendation systems and advertisement systems. In this work we presented a comparison between several feature ranking methods used in and we considered eight ranking methods and adopted ten different learning algorithms, namely, NaiveBayes, J48, SMO, JRIP, Decision table, RandomForest, Multilayerperceptron and Kstar to test the accuracy. We also compare the results induced by several ranking methods with these algorithms. In our experiments, ranking methods with different supervised learning algorithms give quite different results for balanced accuracy. This study shows that the selection of ranking methods could be important for classification accuracy and the proposed method are able to reduce disaster impact through correctly identifying the potential risk features for disaster management.

Key words: Ranking Methods • Feature Selection • Classification Algorithms • Classification Accuracy

INTRODUCTION

Ranking is a crucial part of information retrieval. It is able to compute sorted score when given document as objects. Ranking is a central issue in information retrieval, in which, given a set of objects (e.g., Documents), a score for each of them is computed and the objects are sorted according to the scores. Depending on the applications the scores may represent the degrees of relevance, preference, or importance. Ranking is a very important topic in feature selection. While algorithms for learning ranking models have been intensively studied, this is not the case for feature selection, despite of its importance. The reality is that many feature selection methods used in classification are directly applied to ranking. We argue that because of the striking differences between ranking and classification, it is better to develop different feature selection methods for ranking.

Feature selection has emerged as a successful mechanism in many machine learning applications. Feature selection is also desirable for learning to rank. First, as the numbers of useful features for ranking are continuously growing, the time of extracting such high-dimensional features has become a bottleneck in ranking. Second, high-dimensional features may be

redundant or noisy, which results in poor generalization performance. Lastly, a ranking model with only a small set of features has less computational cost in prediction. Recently, considerable efforts have been made on feature selection for ranking. The main aim of this paper was to experimentally verify the impact of different ranking methods on classification accuracy. We have shown that there is no better ranking index for different data sets and different classifiers accuracy curves, as the function of the number of features used may significantly differ. The only way to be sure that the highest accuracy is obtained in practical problems is testing a given classifier on a number of feature subsets, obtained from different ranking indices. Diverse feature ranking and feature selection techniques have been proposed in the machine learning literature. The purpose of these techniques is to discard irrelevant or redundant features from a given feature vector. In this paper, we consider evaluation of the practical usefulness of the following commonly used ranking methods in different datasets.

- Relief.
- GainRatio (GR)
- InformationGain (IG).
- One-R

- SymmetricalUncertainty (SU)
- Chi-Squared
- SupportVectorMachine (SVM).
- Filter.

The results were validated using different algorithms for classification. A wide range of classification algorithms is available, each with its strengths and weaknesses. There is no single learning algorithm that works best on all supervised learning problems.

Literature Survey: A Ranking is a task that applies machine learning techniques to learn good ranking predictors. It is a relationship between a set of items and a unit that refer to different values. Many learning-to-rank algorithms have been proposed. The two prime functions of ranking are to deliver highly relevant search results and to be fast in ranking results. Many feature selection and feature ranking methods have been proposed. In 1989 Fuhr and Norbert introduced a Ranking OPRF [1] method which uses the idea of Polynomial regression. Cooper and William.S proposed a point wise SLR (Staged logistic regression ranking) method in 1992 [2]. A RELIEF ranking algorithm was proposed by Kira and Rendell in the year 1992 [3]. The strengths of relief is that, it is not dependent on heuristics, it requires only linear time in the number of given features and training instances and it is noise-tolerant and robust to feature interactions, as well as being applicable for binary or continuous data. However, it does not discriminate between redundant features and low numbers of training instances fool the algorithm. Kononenko et al. proposed some updates to the algorithm (RELIEF-F) in order to improve the reliability of the probability approximation, make it robust to incomplete data and generalizing it to multi-class problems [4]. Then the original support vector machine algorithm (SVM) was invented by Vladimir N. Vapnik in 1992 [5]. This SVM is supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. SVMs are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. SVMs deliver state-of-the-art performance in real-world applications such as text categorization, hand-written character recognition, image classification, bio sequences analysis, etc. and are now established as one of the standard tools for machine learning and data mining.

Information Gain: Another ranking method called as Information gain (IG) evaluates the worth of an attribute by measuring the information gain with respect to the class. An attribute selection measure, based on pioneering work by Claude Shannon on information theory, which studied the value of the information content of messages.

Gain Ratio: Gain Ratio is the extension of IG which attempts to overcome this bias. It evaluates the worth of an attribute by measuring the gain ratio with respect to the class. The Gain Ratio is the non-symmetrical measure that is introduced to compensate for the bias of the IG [6-17].

Symmetrical Uncertainty: The Symmetrical Uncertainty criterion compensates for the inherent bias of IG by dividing it by the sum of the entropies of X and Y [17]. SU takes values, which are normalized to the range [0, 1] because of the Correction factor 2. A value of SU = 1 means that the knowledge of one feature completely predicts and the other SU = 0 indicates, that X and Y are uncorrelated. Similar to GR, the SU is biased toward features with fewer values

Chi-squared: Feature Selection via chi square test is another very commonly used method [18]. Chi-squared attribute evaluation evaluates the worth of a feature by computing the value of the chi-squared statistic with respect to the class. The initial hypothesis H_0 is the assumption that the two features are unrelated and it is tested by chi squared

One-R: OneR builds one rule for each attribute in the training data and then selects the rule with the smallest error. It treats all numerically valued features as continuous and uses a straightforward method to divide the range of values into several disjoint intervals. It handles missing values by treating "missing" as a legitimate value. This is one of the most primitive schemes. It produces simple rules based on feature only. Although it is a minimal form of classifier, it can be useful for determining a baseline performance as a benchmark for other learning schemes.

A pairwise RankSVM [6] method which was devised in the year 2000 out performs more naive approaches to ordinal regression such as Support Vector Classification and Support Vector Regression in the case of more than two ranks. In the year 2003, 2005 and 2006 a pairwise RankBoost, RankNet [7] and IR-SVM, Lambda Rank

methods were developed. Subsequently, in 2007, the ranking methods Frank, GB Rank, ListNet, McRank, QBRank, RankCosine, RankGP and RankRLS were innovated. In the year 2007 a listwise ranking methods ListNet, RankCosine, RankGPand, SVMmap [8] were introduced. Ranking Refinement method (2008) is a semi-supervised approach to learning to rank that uses Boosting. Then a list wise ranking methods LambdaMART [9], ListMLE, PermuRank, SoftRank and a pairwise ranking methods Ranking Refinement [10] SSRankBoost [11], SortNet [12] were developed in 2008. In 2009 MPBoost, BoltzRank and BayesRank [13] later in 2010 NDCG Boost [14], Gblend, IntervalRank [15] and CRR [16] were discovered.

Point Wise Approach: In this case it is assumed that each query-document pair in the training data has a numerical or ordinal score. Then learning-to-rank problem can be approximated by a regression problem-given a single query-document pair, predict its score [19].

Pairwise Approach: In this case learning-to-rank problem is approximated by a classification problem- learning a binary classifier that can tell which document is better in a given pair of documents. The goal is to minimize the average number of inversions in ranking [20].

List Wise Approach: These algorithms try to directly optimize the value of one of the above evaluation measures, averaged over all queries in the training data. This is difficult because most evaluation measures are not continuous functions with respect to ranking model's parameters and so continuous approximations or bounds on evaluation measures have to be used.

Proposed Work and Experimental Results: Data mining or “Knowledge Discovery in Databases” is the process of discovering patterns in large data sets with artificial intelligence, machine learning, statistics and database systems. The overall goal of a data mining process is to extract information from a data set and transform it into an understandable structure for further use. In its simplest

form, data mining automates the detection of relevant patterns in a database, using defined approaches and algorithms to look into current and historical data that can then be analyzed to predict future trends. A data mining tools predict future trends and behaviours by reading through databases for hidden patterns; they allow organizations to make proactive, knowledge-driven decisions and answer questions that were previously too time-consuming to resolve. With Weka, Open Source software, you can discover patterns in large data sets and extract all the information. It is a Comprehensive tool for machine learning and data mining for predictive analytics. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a data set or called from your own java code. It is also well suited for developing new machine learning schemes. It also brings great portability, since it was fully implemented in the Java programming language, plus supporting several standard data mining tasks. It contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. Different ranking methods can also be implemented using the data pre-processing tool which is available in Weka. It is also well-suited for developing new machine learning schemes.

In our experiments, we have used five datasets namely, diabetes, segment-challenge, soybean, vote and ionosphere from the UCI data repository [20]. The first dataset is the diabetes data which has 768 instances and 9 attributes. The second data set segment-challenge has 1500 instances and 20 attributes. Similarly soybean, vote and ionosphere datasets have 683,435,351 instances and 36, 17, 35 attributes respectively. In weka a wide range of classification algorithms is available for data analysis. From this wide range of learning algorithms, eight different algorithms are chosen and applied on all the five datasets for our study.

Summary and Finding: In our work we found that, ranking from datasets is indeed a very important problem from both the algorithmic and performance perspective in data mining. Ranking methods with different classification algorithms gives different accuracy. Hence selection of

Table 1: Datasets used in the Experiment

Sl.No	Name of the Dataset	No. of attributes	No. of Instances
1	Diabetes	9	768
2	segment-challenge	20	1500
3	Soybean	36	683
4	Vote	17	435
5	Ionosphere	35	351

Table 2: Classification accuracy of different Classification algorithm without Ranking

S. No.	Dataset	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul.pr	Kstar
1	Diabetes	76.3	73.82	77.34	76.04	71.22	73.82	75.39	69.14
2	Segment-challenge	81.06	95.73	91.93	93.73	87.4	96.93	96.73	96.6
3	Soybean	92.97	91.5	93.85	91.94	84.33	92.09	93.41	87.99
4	Vote	90.11	96.32	96.09	95.4	94.94	95.63	94.71	93.33
5	Ionosphere	82.62	91.45	88.6	89.74	89.45	92.87	91.16	84.61
Classification Average		84.61	89.76	89.56	89.37	85.47	90.27	90.28	86.33

Table 3: Processing Time of different Classification algorithm without Ranking

S. No.	Dataset	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul.pr	Kstar
1	Diabetes	0.02	0.04	0.26	0.06	0.09	0.13	1.96	0.0
2	Segment-challenge	0.02	0.09	1.85	0.55	0.49	0.26	17.06	0.0
3	Soybean	0.0	0.03	4.77	0.11	0.81	0.33	97.25	0.0
4	Vote	0.0	0.0	0.04	0.01	0.09	0.07	2.41	0.0
5	Ionosphere	0.01	0.04	0.08	0.07	0.15	0.01	6.59	0.0
Average Processing Time		0.01	0.04	1.4	0.16	0.32	0.16	25.05	0

Table 4: Classification accuracy on selected features for Diabetes dataset

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul.pr	Kstar	F.S.Avg.
Relief	75.4	74.3	76.4	74.1	73.0	73.4	74.7	69.0	73.8
GainRatio	75.5	74.9	76.2	75.9	72.4	72.0	76.3	71.4	74.3
InfoGain	75.4	74.3	76.0	75.1	72.1	72.0	77.2	71.6	74.2
OneR	75.5	74.9	76.2	76.2	72.4	72.6	76.0	71.4	74.4
SU	75.4	74.3	76.0	75.1	72.1	72.0	77.2	71.6	74.2
Chi-squared	75.4	74.3	76.0	74.9	71.6	71.2	76.7	71.6	74.0
SVM	77.2	74.9	76.8	74.2	72.7	72.4	75.1	71.9	74.4
Filter	75.4	74.3	76.0	75.1	72.1	72.0	77.2	71.6	74.2
Classification Avg.	75.7	74.5	76.2	75.1	72.3	72.2	76.3	71.3	

Table 5: Classification accuracy on selected features for segment-challenge dataset

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul.pr	Kstar	F.S.Avg
Relief	73.3	94.6	83.1	93.8	87.0	96.2	95.6	96.9	90.1
GainRatio	66.4	89.2	77.4	86.6	82.8	90.6	86.3	92.1	84.3
InfoGain	76.9	94.8	89.6	93.9	87.0	96.2	85.3	97.1	91.4
OneR	75.0	94.9	87.6	93.6	87.0	96.4	95.5	97.0	90.9
SU	76.9	94.9	89.6	93.2	87.0	96.8	95.5	97.1	91.3
Chi-squared	66.4	89.2	77.6	88.0	95.6	82.8	88.9	95.1	85.5
SVM	82.0	94.6	90.7	93.4	88.2	96.7	96.0	95.1	92.2
Filter	76.9	94.8	89.6	93.9	87.0	96.2	95.3	95.7	91.4
Classification Avg.	74.2	93.4	85.7	92.1	87.7	94.4	93.9	96.0	

Table 6: Classification accuracy on selected features for soybean dataset

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul Pr	Kstar	F.S.Avg.
Relief	89.5	88.6	92.8	87.8	80.1	89.0	92.1	88.3	88.5
GainRatio	85.8	85.2	86.2	84.9	82.7	87.4	87.4	86.1	85.7
InfoGain	89.9	88.3	93.0	88.7	80.1	86.8	93.3	88.9	88.6
OneR	83.6	85.4	87.1	84.8	83.9	86.5	87.3	86.4	85.6
SU	89.8	90.3	93.4	89.8	82.4	88.3	93.6	90.5	89.8
Chi-squared	89.2	89.8	93.9	89.6	81.3	91.4	93.7	90.0	89.8
SVM	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Filter	89.9	88.3	93.0	89.9	80.1	86.8	93.3	88.9	88.8
Classification Avg.	88.2	88.0	91.3	87.9	81.5	88.0	91.5	88.4	

Table 7: Classification accuracy on selected features for vote dataset

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul Pr	Kstar	F.S.Avg.
Relief	90.3	96.3	95.6	95.9	95.9	95.9	93.8	94.9	94.8
GainRatio	91.3	95.2	95.6	95.6	95.6	94.5	95.2	92.9	94.5
InfoGain	91.3	95.2	95.6	95.6	95.6	94.5	95.2	92.9	94.5
OneR	90.6	94.7	95.6	95.4	95.4	95.2	94.0	92.9	94.2
SU	91.3	95.2	95.6	95.6	95.6	94.1	95.2	92.9	94.4
Chi-squared	91.3	95.2	95.6	95.6	95.6	93.6	94.0	92.9	94.2
SVM	91.5	96.3	95.9	96.3	94.7	95.9	94.9	94.0	94.9
Filter	91.3	95.2	95.6	95.6	95.6	94.3	95.2	92.9	94.5
Classification Avg.	91.1	95.4	95.7	95.7	95.5	94.7	94.7	93.3	

Table 8: Classification accuracy on selected features for ionosphere dataset

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul Pr	Kstar	F.S.Avg.
Relief	86.3	92.9	87.7	90.9	89.5	93.2	90.9	84.6	89.5
GainRatio	87.5	90.3	87.7	91.7	89.5	93.4	92.6	85.2	89.7
InfoGain	88.0	92.0	87.7	90.9	89.5	93.4	94.0	86.6	90.3
OneR	88.0	92.0	87.7	90.9	89.5	93.4	91.5	84.6	89.7
SU	88.0	92.0	87.7	90.9	89.5	93.4	92.0	86.3	90.0
Chi-squared	88.0	92.0	87.7	90.9	89.5	93.4	94.6	86.6	90.3
SVM	88.0	92.0	87.7	90.9	89.5	93.4	91.1	87.2	90.0
Filter	88.0	92.0	87.7	90.9	89.5	93.4	94.0	86.6	90.3
Classification Avg.	87.7	91.9	87.7	91.0	89.5	93.4	92.6	86.0	

Table 9: Average Classification accuracy on Full set with ranking

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul. Pr	Kstar	F.S.Avg
Relief	84.51	89.65	89.56	89.062	85.65	89.89	72.53	86.82	85.96
GainRatio	84.61	89.65	89.53	88.77	85.084	91.00	90.52	87.082	88.28
InfoGain	84.61	89.70	89.37	89.46	85.22	90.62	90.33	86.82	88.27
OneR	84.61	89.76	89.55	88.91	85.38	90.91	90.25	85.16	88.07
SU	84.61	89.71	89.53	88.92	85.25	90.48	90.41	86.82	88.22
Chi-squared	84.61	89.71	89.56	89.19	85.35	90.55	90.47	86.38	88.23
SVM	82.52	89.33	88.49	88.60	85.81	90.45	89.39	85.81	87.55
Filter	84.61	89.71	89.56	89.46	85.22	90.62	90.34	86.56	88.26
Classification average		84.34	89.65	89.39	89.05	85.37	90.57	88.03	86.43

Table 10: Average Classification accuracy on selected features with Ranking

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul. Pr	Kstar	F.S.Avg.
Relief	84.77	89.25	86.33	88.68	84.82	89.44	89.34	86.83	87.43
GainRatio	81.34	86.95	84.82	86.57	84.33	87.29	88.06	85.52	85.61
InfoGain	83.84	88.81	88.24	88.46	84.76	88.63	91.00	87.42	87.65
OneR	82.68	88.70	86.31	87.61	85.52	88.90	88.52	86.66	86.86
SU	84.10	89.30	88.55	88.70	85.23	88.79	90.70	87.68	87.88
Chi-squared	81.71	88.11	85.91	87.91	84.25	87.72	89.58	86.74	86.49
SVM	84.17	88.67	87.86	87.69	86.45	88.86	89.32	87.27	87.54
Filter	83.84	88.81	88.24	88.46	84.76	88.63	91.00	87.42	87.65
Classification average		83.31	88.58	87.03	88.01	85.02	88.53	89.69	86.94

Table 11: Average processing time with ranking on Full set

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul.Pr	Kstar	F.S.Avg
Relief	0.02	0.06	2.67	0.22	0.34	0.19	24.90	0.00	3.55
GainRatio	0.00	0.04	1.25	0.18	0.32	0.18	24.93	0.00	3.36
InfoGain	0.01	0.04	1.39	0.16	0.33	0.18	24.99	0.00	3.39
OneR	0.01	0.04	1.05	0.17	0.37	0.17	25.02	0.00	3.35
SU	0.01	0.04	1.16	0.21	0.33	0.18	24.96	0.00	3.36
Chi-squared	0.01	0.04	1.15	0.22	0.36	0.17	24.97	0.00	3.37
SVM	0.01	0.03	0.41	0.18	0.17	0.11	5.61	0.00	0.82
Filter	0.00	0.04	0.88	0.19	0.35	0.17	24.87	0.00	3.31
Classification average		0.01	0.04	1.25	0.19	0.32	0.17	22.53	0.00

Table 12: Average processing time with on selected features

Ranking Method	NB	J48	SMO	JRIP	DT	Rd.Frt	Mul. Pr	Kstar	F.S.Avg
Relief	0.00	0.02	1.45	0.13	0.13	0.10	12.15	0.00	1.75
GainRatio	0.00	0.04	0.93	0.13	0.12	0.10	9.36	0.00	1.34
InfoGain	0.00	0.02	0.99	0.12	0.17	0.14	13.06	0.00	1.81
OneR	0.00	0.02	0.98	0.12	0.17	0.10	10.72	0.00	1.51
SU	0.00	0.02	1.23	0.11	0.14	0.13	13.04	0.00	1.83
Chi-squared	0.00	0.02	1.00	0.13	0.13	0.11	12.59	0.00	1.75
SVM	0.00	0.02	0.30	0.08	0.07	0.09	2.55	0.00	0.39
Filter	0.00	0.02	0.80	0.11	0.13	0.11	12.99	0.00	1.77
Classification average		0	0.0225	0.96	0.11625	0.1325	0.11	10.8075	0

Table 13: Average Classification Accuracy and Processing Time for classification Algorithms

Classification Algorithms	Without Ranking on Full set		With Ranking On Full set		With ranking On selected set	
	F.S Avg.	Processing Time(S)	F.S Avg.	Processing Time(S)	F.S Avg.	ProcessingTime(S)
Naive Bayes	84.61	00.01	84.34	0.01	83.31	00.00
J48	89.76	00.04	89.65	0.04	88.58	00.02
SMO	89.56	01.40	89.39	01.25	87.03	00.96
JRIP	89.37	00.16	89.05	00.19	88.01	00.11
Decision Tree	85.47	00.32	85.37	00.32	85.02	00.13
Random Forest	90.27	00.16	90.57	00.17	88.53	00.11
Multilayer Perceptron	90.28	25.05	88.03	22.53	89.69	10.80
Kstar	86.33	00.00	86.43	00.00	86.94	00.00

ranking method is an important task for improving the classification accuracy. Not choosing the right ranking method for a dataset introduces bias towards selecting the best features. Furthermore predictive accuracy is not a useful measure when evolutionary classifiers learned on datasets. In this study, out of eight ranking methods SVM scores the maximum accuracy for three datasets (vote, segment-challenge and diabetes) Chi-square scores for two datasets (ionosphere and soybean) and Filter, OneR, InfoGain scores for one datasets (ionosphere, diabetes). But we found that Symmetrical Uncertainty (SU) which does not scores the maximum accuracy for any datasets give the maximum accuracy of 87.88 percentages comparing with other conventional ranking methods. The overall time taken by SU is higher when comparing with other ranking methods.

From this study we found the following significant observations:

- Multilayer Perceptron, Random Forest, J48, SMO and JRIP perform better than other classification algorithms with and without ranking and also on selected features.
- SVM ranking method will take only minimal processing time period than other ranking methods with reasonable classification accuracy.
- The selected features by Relief ranking method will provide better performance comparing with ranking with full dataset.

- With selected features, the performance of Gain Ratio is poor than other ranking methods.
- SU based ranking method will able to reduce the number of initial attributes with maximum time period and also increase the classification performance than other methods.
- addition the development of disaster risk management framework through data analytics by the use of the proposed ranking based feature selector can be considered as one of our future work.

CONCLUSION

Random Forest, Multilayerperceptron and Kstar to test the accuracy. We also compare the results induced by several ranking methods with these algorithms. In our experiments, ranking methods with different supervised learning algorithms give quite different results for balanced accuracy. This study shows that the selection of ranking methods could be important for classification accuracy and the proposed method are able to reduce disaster impact through correctly identifying the potential risk features for disaster management.

REFERENCES

1. Fuhr Norbert, 1989. Optimum polynomial retrieval functions based on the probability ranking principle, ACM Transactions on Information Systems.

2. Cooper William, S., C. Frederic Gey and Daniel P. Dabney, 1992. Probabilistic retrieval based on staged logistic regression, SIGIR '92 Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, pp: 198-210.
3. Kira, K. and L.Rendell, 1992. A practical approach to feature selection. In *ML92: Proceedings of the ninth international workshop on Machine learning*. Morgan Kaufmann Publishers Inc., pp: 249-256.
4. Robnik-Sikonja, M. and I. Kononenko, 2003. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 53: 23-69.
5. Cortes, C. and V. Vapnik, 1995. Support-vector networks, *Machine Learning*, 20(3): 273.
6. Large Margin Rank Boundaries for Ordinal Regression by Ralf Herbrich, Thore Graepel and Klaus Obermayer, pp: 2000.
7. Burges Chris and Tal Shaked, 2005. Learning to Rank using Gradient Descent, Proceedings of the 22nd International Conference on Machine Learning, Bonn.
8. Yue Yisong and Thomas Finley, 2007. A Support Vector Method for Optimizing Average Precision-SIGIR'07, Amsterdam, The Netherlands. Copyright 2007 ACM 978-1-59593-597-7/07/0007.
9. Wu Qiang, Chris J.C. Burges and Krysta M. Svore, 2008. Ranking, Boosting and Model Adaptation and JianfengGaoMicrosoft Research One Microsoft Way Redmond, WA 98052, Microsoft Research Technical Report MSR-TR-2008-109-October 15.
10. Leonardo Rigutini, Tiziano Papini, Marco Maggini and Franco Scarselli, 2008. SortNet: learning to rank by a neural-based sorting algorithm, SIGIR workshop: Learning to Rank for Information Retrieval.
11. Massih-Reza Amini, Vinh Truong and Cyril Goutte, 2008. A Boosting Algorithm for Learning Bipartite Ranking Functions with Partially Labeled Data, International ACM SIGIR conference.
12. Rigutini Leonardo, Tiziano Papini, Marco Maggini and Franco Scarselli, 2008. SortNet: learning to rank by a neural-based sorting algorithm, SIGIR 2008 workshop: Learning to Rank for Information Retrieval.
13. Jen-Wei Kuo and Pu-Jen Cheng, 2009 Learning to Rank from Bayesian Decision Inference"-CIKM'09, November 2-6, Hong Kong, China. Copyright 2009 ACM 978-1-60558-512-3/09/1.
14. Hamed Valizadegan, Rong Jin, Ruofei Zhang and Jianchang Mao, 2010. Learning to Rank by Optimizing NDCG Measure, in Proceeding of Neural Information Processing Systems (NIPS).
15. Moon Taesup, Alex Smola, Yi Chang and Zhaohui Zhen, 2010. Yahoo! Labs.IntervalRank, Isotonic Regression with Listwise and Pairwise Constraint-WSDM'10, February 4-6, 2010, New York City, New York, USA. Copyright 2010 ACM 978-1-60558-889-6/10/02.
16. Combined Regression and Ranking, D. Sculley Google, 2010. Inc. Pittsburgh, PA USA- KDD'10, Washington DC, USA. Copyright 2010 ACM 978-1-4503-0055-1/10/07.
17. Hall, A., M.A. and L.A. Smith, 1998. Practical feature subset selection for machine learning, Proceedings of the 21st Australian Computer Science Conference, pp: 181-191.
18. Liu, H. and R. Setiono, 1995. Chi2: Feature selection and discretization of numeric attributes, Proc. IEEE 7th International Conference on Tools with Artificial Intelligence, pp: 338-391.
19. Kuramochi B. and G. Karypis, 2005. Gene classification using expression profiles: a feasibility study, *International Journal on Artificial Intelligence Tools*, 14(4): 641-660.
20. UCIRepository: <http://archive.ics.uci.edu/ml>.