

Bipartite Weighted Graph Access for Optimal Label Prediction

Jyothi Puligadda

ABSTRACT---This article portrayed a novel bipartite weighted graph strategy for Feature optimization for machine learning models. Unlike many of the existing optimization techniques of diverse categories such as evolutionary computation techniques, diversity assessment strategies, the proposal is deterministic approach with minimal computational overhead, which has referred further as Bipartite Weighted Graph Approach for Optimal label prediction (BWG-OLP). The proposed model is about to derive a given feature is optimal or not by the respective feature's correlation with the records and the correlation with the fellow features. The experimental study has carried on benchmark datasets to estimate the significance of the proposed method.

Keywords—Mutual Information, Bipartite Weighted Edge Graph, LASSO, Particle-Swarm Optimization, Degree of Positive Label Scope.

1. INTRODUCTION

The prominent challenge that recognition of pattern has is high dimensional data. The well-known models for lessening the dimensionality were subspace models based on variance like PCA. Nevertheless, the vectors of PCA feature extracted the features capture sets with an important integrated variance. This might render them correspondingly ineffective for the tasks of classification. Therefore, it could be important in detecting the minimum features subset, which are informative towards clustering & classification. The thought of underpinning selection of feature is to choose the features, which are mostly associated to classification at the time of lessening redundancy. Shared information gives an ethical way of evaluating the combined dependence of 2 variables. It has been utilized by various researchers for enhancing the theoretic information criteria of feature selection. For instance, the work [1] developed MIFS (Mutual information-based feature selection) criteria. Here, the features were chosen in a voracious approach. The existing features selected set S is provided, at every stage it locates x_i feature, which increase the significance to $I(x_i; C)$ class. Here, selection is controlled by $\beta I(x_i; S)$ proportional term, which evaluates the information of overlap among the existing features & candidate feature; whereas, the β parameter might impact the selected features significantly, and its regulation remains as an open issue. The work [2] presents that on other dimension, it utilizes MRMR (Maximum Relevance Minimum Redundancy criteria that is similar to the MIFS through $\beta = 1/n-1$). The work [3] presents that JMI (Joint mutual information) criteria is on the basis of conditional-MI, and features are selected by examining whether they bring extra information to the

contemporary feature-set. Here, this model evades the redundant features effectively. The work [4] presents that MIFS is improved by enhancing the MIFS-U under the estimation of uniform information distribution for the input-features. Here, it computes the MI on the basis of Parzen window that is computationally less demanding and delivers better predictions. Nevertheless, there were 2 confines for MI-feature selection models. Initially, they estimate that every single related feature needs to be relied with desired class. It means, if one feature is deliberated as relevant, then it needs to be associated with required class, or else, the feature could be irrelevant [5]. Therefore, only small relevant features set is chosen and huge combinations of feature are not deliberated. The 2nd confine is that, various models consider simply pairwise dependencies of feature and would not examine for 3rd or higher sequence dependencies among existing & candidate features.

2. RELATED WORK

To attain diminished set of features for the issue of classification, several feature selections models (wrapper & filter) is implemented by various researchers to enhance the accuracy of classifiers. The work [6] proposed a feature selection algorithm known as particle-swarm optimization (PSO) to cluster the text. The work [7] presents that the model based on wrapper to choose optimal genes from the dataset of microarray is projected with an approach of Markova blanket. Here, they attained effective outcomes by their approach. Other approach is the embedded selection of feature for wrapping and filter that also applied extensively for the classification issues. Some of well-known or prominent embedded FS models are Random forest, LASSO and SVM-RFE. The method SVM-RFE is implemented for the classification of cancer [8]. The work [9] presents that LASSO method is proposed with the help of various researchers for drawing minimum features towards effective outcomes by using stability arguments. The work [10] presents that the random forest model is implemented for the classification of land coverage.

The work [11] implemented wrapper & filter mechanisms for resolving protein disordered area prediction problem. The dataset that deliberated has 440 attributes initially. Primarily, F-Score & IG could be employed over dataset, and further wrapper model is implemented for identifying the better performance of classification. Contemporarily, evaluated the projected framework through some of significant conventional feature-selection methods (IG, GR,

Revised Manuscript Received on August 19, 2019.

Dr. JyothiPuligadda, Assistant Professor in Mathematics, Anurag Engineering College, (An UGC Autonomous Institution) Kodad, Suryapet Dist., Telangana, India.(Jyothi.puligadda.77@gmail.com)

Chi&Rel). This notion of IG is based on information concept. It analyzes the relationship among classes & features for eradicating redundant attributes and enormous independent attributes through class-label. The feature selection model based on IG is implemented over voting dataset & kidney disease [12]. The work [13] proposed FS model based on mutual information. The work [14] proposed Max Dep concept. The statistical subset dependency is calculated by target class-label. Here, this model aims for selecting n features, which jointly possess the highest dependency with target class label.

For combined high dimensional data protein, the MCI-RFE (Maximum Correlation Information-Recursive Feature Elimination) model is proposed [15]. Here, in MCI-RFE model, the importance of each attribute could be computed by MCI. Later, MCI is combined with RFE for producing the robust feature subset. The MCI-RFE could be competitive highly with Relief-RFE, Random Forest & SVM-RFE.

The FS is an alternative challenge for various researchers in addressing the crucial problems associated with huge amount of the attributes in pattern recognition field for securing effective outcomes [16]. The CFS (correlation-based selection) method is implemented by various researchers for several reasons. The CFS is implemented for estimating the demand of electricity in Australia [17]. Here, they implemented tree, neural networks-based algorithms over 2 years' time load data series. They projected FAST algorithm on the basis of CCE & SU for attaining optimum subset. The FAST algorithm could be the algorithm based on clustering that works in 2 stages. In primary stage, the graph theory clustering model is implemented for producing attributes into groups. In record of 2nd stage, the prims algorithm is implemented for choosing optimum attributes [18].

To overcome confines noticed in existing feature optimization methods, this paper depicted weighted bipartite graph scheme for indicating optimum feature selection, which competes on both the continuous & categorical formats of data. Here, the projected method performs better when compared with other proposed methods.

3. BIPARTITE WEIGHTED GRAPH BASED FEATURE OPTIMIZATION

The proposal is a bipartite weighted graph that estimates the metric referred as Correlation Impact (CI) between the features, and between the features and labeled records. Partitioning the given labeled records such that positive and negative, from positive records, extracting the unique values observed for each attribute as a set, such that each entry of this set represents the attribute name and corresponding unique value. Build a bipartite weighted graph between the features such that the weighted edge connects two features as authorities, if both of these features exists in at least one record. In addition, the edge weight is the ratio of records contains both the authorities and total number of records. Then, build a bipartite weighted graph between records as hubs and features as authorities. An edge appears between a hub and authority, if the corresponding authority exists the respective hub. Each of these edges represents by the weight, which is the empirical probability of the weights

observed between the authority and the other features existing in the corresponding hub. Build a matrix using the depicted edge weights in bipartite graph, such that each row represents the edge weights observed for each record to all features and each column represents the edge weights of each feature to all records. Then apply HITS algorithm on the depicted matrix to identify each record weight as hub weight and each feature weight as authority weight. Afterwards, find the Correlation Impact of each feature from the record and feature weights discovered through HITS algorithm, which is the ratio between sum of the weights of the records those contains the respective feature and the sum of the weights of the all records. Then the Correlation Impact of these features is used to depict the degree of positive label scope (DPLS) of each record, which is the ratio between the sum of the Correlation Impact of the features exists in corresponding record and the total number of features. Find the threshold of degree of positive label scope (DPLST), which is the mean of the DPLS of the records. Find the lower and upper bound of the DPLST, where lower bound is the result obtained by subtracting the deviation of the DPLS observed for all records from the DPLST and the upper bound is the result obtained by summing the deviation of the DPLS observed for all records and the DPLST. Further, these DPLST, its lower bound and upper bound re used to assess the positive label scope of a given unlabeled record.

3.1 Data Formation

The notation $T = \{t_1, t_2, \dots, t_{|T|}\}$ denotes the set of transactions representing in the format of two-dimensional matrix, such that each transaction labeled as positive or negative. Each of the transaction $t_i = \{v_i, v_2, \dots, v_{|t_i|}\}$ is a vector of values.

Further, find all unique values of the given set T as a set F , such that an entry $\{f_j^i \exists f_j^i \in F\}$ of the set F denoted by a unique index j and the column index i .

Further, discover the unique bigrams from the set F and list them as a set B

Further, partition the given set T as two sets T_+, T_- , such that the set T_+ contains all the records labeled as positive, and the set T_- contains all the records labeled as negative.

Let the features of each bigram $\{bg \exists bg \in B\}$ connected with a positive weighted edge pwe , and a negative weighted edge, which have to scaled as follows

$\forall_{k=1}^{|B|} \{bg_k \exists bg_k \in B\}$ Begin // for each bigram listed in the set B

$$pew(bg_k) = \frac{1}{|T_+|} \sum_{m=1}^{|T_+|} \{1 \exists bg_k \subset t_m \wedge t_m \in T_+\} \quad //$$

positive weighted edge between the features in bigram bg_k

$$new(bg_k) = \frac{1}{|T_-|} \sum_{m=1}^{|T_-|} \{1 \exists bg_k \subset t_m \wedge t_m \in T_-\} \quad //$$

negative weighted edge $new(bg_k)$ between the features in bigram bg_k

End

Bipartite weighted edge graph

Further, for the set T_+ of positive labels, the proposal builds a bipartite weighted edge graph between the transactions T_+ and features exist in set F , which is as follows

Initial phase defines an edge e between a feature $\{f \exists f \in F\}$ and a transaction $\{t \exists t \in T\}$, if the corresponding feature $\{f \exists f \in F\}$ exists in respective transaction $\{t \exists t \in T\}$

Further, estimates the weights of the edges between transactions and features as follows.

A bipartite weighted edge graphs have to build between features F and both positive T_+ and negative T_- transactions F that represents positive transactions as hubs (left side elements of the bipartite weighted graph) and all possible features listed in the set F as authorities (right side of the bipartite weighted graph). There will be an edge between a hub (transaction) and authority (feature), if the authority exists in the corresponding hub. The edge has to estimate as follows

$$\forall_{k=1}^{|T_+|} \{t_k \exists t_k \in T_+\} \text{ Begin // for each positive transaction (hub)}$$

$$\forall_{i=1}^{|F|} \{f_i \exists f_i \in F\} \text{ Begin // for each feature (authority)}$$

$ew(t_k, f_i) = 0$ // the weight of the edge between hub t_k and authority f_i

$$\forall_{j=1}^{|F|} \{f_j \exists f_j \in F \wedge i \neq j \wedge (f_i, f_j) \subseteq t_k\} \text{ Begin // for each}$$

feature (authority)

$$ew(t_k, f_i) + = \{ew(bg) \exists (f_i, f_j) \subseteq bg\}$$

End

$ew(t_k, f_i)^* = (|t_k|)^{-1}$ // the edge weight $ew(t_k, f_i)$ of the edge between positive transaction (positive hub) t_k , and feature (authority) f_i

End

End

Similar approach should apply between features F and negative transactions T_- that delivers a bipartite weighted graph with edge weights between the features (authorities) F and negative transactions (hubs) T_- .

Feature Correlation Weights

Further, the correlation weight of each feature has to estimate as follows

$$\forall_{i=1}^{|F|} \{f_i \exists f_i \in F\} \text{ Begin // for each feature (authority)}$$

$$pcw(f_i) = \frac{1}{|T_+|} \sum_{k=1}^{|T_+|} \{ew(f_i, t_k) \exists f_i \in t_k \wedge t_k \in T_+\} \quad //$$

the positive correlation weight $pcw(f_i)$ of the feature (authority) f_i

$$ncw(f_i) = \frac{1}{|T_-|} \sum_{k=1}^{|T_-|} \{ew(f_i, t_k) \exists f_i \in t_k \wedge t_k \in T_-\} \quad //$$

the negative correlation weight $ncw(f_i)$ of the feature (authority) f_i

End

Degree of Label Scope

Further, Degree of positive and negative label scopes $dpls, dnls$ for each record has to estimate as follows,

$$\forall_{k=1}^{|T_+|} \{t_k \exists t_k \in T_+\} \text{ Begin // for each positive transaction (hub)}$$

$$dpls(t_k) = \left(\sum_{j=1}^{|F|} pcw(f_j) \exists f_j \subset t_k \wedge t_k \in T_+ \right) * (|F|)^{-1}$$

//degree of positive label scope $dpls(t_k)$ of the transaction t_k

End

$$\forall_{k=1}^{|T_-|} \{t_k \exists t_k \in T_-\} \text{ Begin // for each negative transaction (hub)}$$

$$dnls(t_k) = \left(\sum_{j=1}^{|F|} ncw(f_j) \exists f_j \subset t_k \wedge t_k \in T_- \right) * (|F|)^{-1}$$

//degree of negative label scope $dnls(t_k)$ of the transaction t_k

End

Label Scope Thresholds and Boundaries

Further, the range of label scope thresholds $plst, nlst$ of both positive and negative labels has to estimate as follows

$$plst = \frac{1}{|T_+|} \left(\sum_{k=1}^{|T_+|} \{dpls(t_k) \exists t_k \in T_+\} \right) \quad //$$

positive label scope threshold

$$d_{plst} = \frac{1}{|T_+|} \left(\sum_{k=1}^{|T_+|} \left\{ \sqrt{(plst - dpls(t_k))^2} \exists t_k \in T_+ \right\} \right) //$$

deviation of degree of positive label scope $dpls$ from the positive label scope threshold $plst$

$$nlst = \frac{1}{|T_-|} \left(\sum_{k=1}^{|T_-|} \{dnls(t_k) \exists t_k \in T_-\} \right) \quad //$$

negative label scope threshold

$$d_{nlst} = \frac{1}{|T_-|} \left(\sum_{k=1}^{|T_-|} \left\{ \sqrt{(nlst - dnls(t_k))^2} \exists t_k \in T_- \right\} \right) //$$

deviation of degree of negative label scope $dnls$



from the negative label scope threshold $nlst$

Further, discover the lower and upper bounds of the positive and negative label scope thresholds as follows

$$l_{plst} = plst - d_{plst} \quad // \text{ lower bound } l_{plst} \text{ and upper bound } u_{plst} = plst + d_{plst}$$

u_{plst} of the positive label scope threshold

$$l_{nlst} = nlst - d_{nlst} \quad // \text{ lower bound } l_{nlst} \text{ and upper bound } u_{nlst} = nlst + d_{nlst}$$

u_{nlst} of the negative label scope threshold

The usage of these thresholds to perform label prediction is the context of the domain expert.

Label Prediction Strategy

The given record t has to label as positive

- If degree of positive label scope $dpls(r)$ of the record r is greater than the upper bound of the positive label scope threshold u_{plst}
- Else if, degree of positive label scope $dpls(r)$ of the record r is greater than the positive label scope threshold $plst$ & the degree of negative label scope $dnls(r)$ is less than the negative label scope threshold $nlst$.
- Else if, degree of positive label scope $dpls(r)$ of the record r is greater than the lower bound of the positive label scope threshold l_{plst} & the degree of negative label scope $dnls(r)$ is less than the lower bound of the negative label scope threshold l_{nlst} .

The given record t has to label as negative

- If degree of negative label scope $dnls(r)$ of the record r is greater than the upper bound of the negative label scope threshold u_{nlst} , and the degree of positive label scope $dpls(r)$ of the record r is less than the positive label scope threshold $plst$
- Else if, degree of negative label scope $dnls(r)$ of the record r is greater than the negative label scope threshold $nlst$ & the degree of positive label scope $dpls(r)$ is less than the lower bound of the positive label scope threshold l_{plst} .
- In all other conditions, label is suspicious, can recommend for experts' opinion

4. EMPIRICAL ANALYSIS OF THE PROPOSED MODEL& RESULTS

The significance of the feature selection and label prediction strategy of the proposal has explored by conducting empirical analysis on heart disease dataset that available in UCI repository [19]. Chosen dataset comprise samples count of 294 (188: positives, 106: negatives), which has considered for analyzing the bipartite weighted graph strategy to perform optimal label prediction. Among these positive and negative labeled records 75% (141: positive,

and 80: negatives) has selected to derive the positive and negative label scope thresholds. The leftover 25% records have unlabeled and given as input to the proposed model that indeed predicts the label through scales derived. The results from the empirical study have delivered significant results. The scales derived from the explored study has detailed in Table 1.

Table 1: The Positive and negative label thresholds and their respective lower and upper bounds

	Label Scope Threshold	Lower Bound	Upper Bound
Positive	0.5142	0.4815	0.5469
Negative	0.4073	0.3777	0.4369

Performance Analysis

The precision, recall, sensitivity, and accuracy are the metrics to scale the prediction performance, which have used to scale the significance of the proposed model. The performance metrics and the values obtained for respective metrics have listed in Table 2.

Table 2: The performance metrics and the values obtained for respective metrics

count of positives used as input	47
count of Negatives used as input	26
TP	43
FP	2
TN	24
FN	4
Precision	0.955556
recall	0.923077
specificity	0.9248
accuracy	0.917808

The overall positive and negative records given to predict the label are 47 and 26 in respective order. The count of records labeled as positives and negatives are 45 and 28 in respective order. The resultant values of the assessment metrics have explored in following description.

The metric "precision" denotes the ratio of truly detected positives (true positives) against the total positives (aggregate of true positives and false positives) detected, which denotes as positive predictive value. The precision observed from the experimental study is 0.956 (rounded to three decimals).

The metric "recall" denotes the ratio of truly detected positive label records count against the count of positive labeled records given as input, which is often termed as true positive rate. The value observed for this metric is 0.923077.

The other metric "specificity" denotes the ratio of records truly labeled as negative against total number of negative records given as input. This metric often termed as true negative rate, which is 0.925 from the experiments.

The metric “accuracy” denotes the overall true prediction rate of the method, which denotes the ratio of the count of truly predicted positives and negatives against the count of the records given as input.

5. CONCLUSION

The contribution of this manuscript is a concept of graph-theory, which is a bipartite weighted graph strategy that enables optimal label prediction for supervised learning. The crucial objective the proposal is to estimate the correlation weights between bigrams (two feature), and the positive & negative correlation weights between the features and transactions. These correlation weights have discovered from the edge weights of the bipartite graphs build between features as hubs and authorities, and the features as authorities and transactions as hubs. Further, these correlation weights have used to discover the label prediction coefficients. The experimental study portrayed the significance of the proposed model to perform optimal label prediction. The future research can use the proposed model as member function to boost the performance of the evolutionary computation techniques.

REFERENCES

1. Battiti R. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on neural networks*. 1994 Jul;5(4):537-50.
2. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2005 Aug 1(8):1226-38.
3. Yang H, Moody J. Feature selection based on joint mutual information. In *Proceedings of international ICSC symposium on advances in intelligent data analysis 1999 Jun 22* (pp. 22-25).
4. Kwak N, Choi CH. Input feature selection by mutual information based on Parzen window. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2002 Dec 1(12):1667-71.
5. Cheng H, Qin Z, Qian W, Liu W. Conditional mutual information-based feature selection. In *2008 International Symposium on Knowledge Acquisition and Modeling 2008 Dec 21* (pp. 103-107). IEEE.
6. Abualigah LM, Khader AT. Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. *The Journal of Supercomputing*. 2017 Nov 1;73(11):4773-95.
7. Wang A, An N, Yang J, Chen G, Li L, Alterovitz G. Wrapper-based gene selection with Markov blanket. *Computers in biology and medicine*. 2017 Feb 1; 81:11-23.
8. Duan KB, Rajapakse JC, Wang H, Azuaje F. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE transactions on nano bioscience*. 2005 Aug 29;4(3):228-34.
9. Thakurta AG, Smith A. Differentially private feature selection via stability arguments, and the robustness of the lasso. In *Conference on Learning Theory 2013 Jun 13* (pp. 819-850).
10. Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2012 Jan 1;67:93-104.
11. Hsu HH, Hsieh CW, Lu MD. Hybrid feature selection by combining filters and wrappers. *Expert Systems with Applications*. 2011 Jul 1;38(7):8144-50.
12. Potharaju SP, Sreedevi M. A novel cluster of feature selection method based on information gain. *IJCTA*. 2017;10(14):9-16.
13. Maji P, Garai P. On fuzzy-rough attribute selection: criteria of max-dependency, max-relevance, min-redundancy, and max-significance. *Applied Soft Computing*. 2013 Sep 1;13(9):3968-80.
14. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*. 2005 Apr;3(02):185-205.
15. Yuan M, Yang Z, Huang G, Ji G. Feature selection by maximizing correlation information for integrated high-dimensional protein data. *Pattern Recognition Letters*. 2017 Jun 1;92: 17-24.
16. Partila P, Voznak M, Tovarek J. Pattern recognition methods and features selection for speech emotion recognition system. *The Scientific World Journal*. 2015;2015.
17. Koprinska I, Rana M, Agelidis VG. Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*. 2015 Jul 1;82:29-40.
18. Mudaliar PU, Patil TA, Thete SS, Moholkar KP. A Fast Clustering Based Feature Subset Selection Algorithm for High Dimensional Data. *International journal of emerging trend in engineering and basic science*. 2015 Feb;2(1):494-9.
19. hungarian.data. Retrieved from <https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/hungarian.data> (2017).