MEETING ABSTRACT Open Access

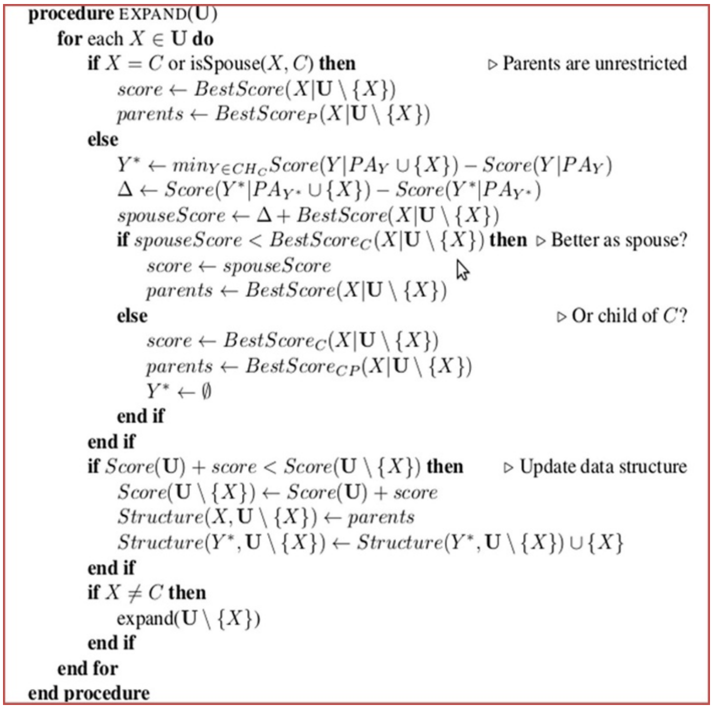
Feature selection and prediction with a Markov blanket structure learning algorithm

Yuan Tan[1], Zhifa Liu[2*]

From 12th Annual UT-ORNL-KBRIN Bioinformatics Summit 2013
Buchanan, TN, USA. 22-24 March 2013

Background

Classification and prediction are common tasks in machine learning. For example, many studies have attempted to predict gene expression given information, such as DNA sequence, expression of other genes or epigenetic modifications. Many existing methods, such as neural networks and support vector machines, have been used to make these predictions. Unfortunately, these black box techniques offer little insight into the reasoning behind the predictions. In many cases, relatively few attributes contribute to the classification accuracy. Bayesian networks explicitly encode the relationships among attributes to make predictions. In a Bayesian network, the Markov blanket (MB) of the class variable gives all of the



```
procedure EXPAND(U)
  for each $X \in \mathbf{U}$ do
    if $X = C$ or isSpouse$(X, C)$ then                          ▷ Parents are unrestricted
      $score \leftarrow BestScore(X|\mathbf{U} \setminus \{X\})$
      $parents \leftarrow BestScore_P(X|\mathbf{U} \setminus \{X\})$
    else
      $Y^* \leftarrow min_{Y \in CH_C} Score(Y|PA_Y \cup \{X\}) - Score(Y|PA_Y)$
      $\Delta \leftarrow Score(Y^*|PA_{Y^*} \cup \{X\}) - Score(Y^*|PA_{Y^*})$
      $spouseScore \leftarrow \Delta + BestScore(X|\mathbf{U} \setminus \{X\})$
      if $spouseScore < BestScore_C(X|\mathbf{U} \setminus \{X\})$ then  ▷ Better as spouse?
        $score \leftarrow spouseScore$
        $parents \leftarrow BestScore(X|\mathbf{U} \setminus \{X\})$
      else                                                       ▷ Or child of $C$?
        $score \leftarrow BestScore_C(X|\mathbf{U} \setminus \{X\})$
        $parents \leftarrow BestScore_{CP}(X|\mathbf{U} \setminus \{X\})$
        $Y^* \leftarrow \emptyset$
      end if
    end if
    if $Score(\mathbf{U}) + score < Score(\mathbf{U} \setminus \{X\})$ then   ▷ Update data structure
      $Score(\mathbf{U} \setminus \{X\}) \leftarrow Score(\mathbf{U}) + score$
      $Structure(X, \mathbf{U} \setminus \{X\}) \leftarrow parents$
      $Structure(Y^*, \mathbf{U} \setminus \{X\}) \leftarrow Structure(Y^*, \mathbf{U} \setminus \{X\}) \cup \{X\}$
    end if
    if $X \neq C$ then
      expand$(\mathbf{U} \setminus \{X\})$
    end if
  end for
end procedure
```

Figure 1 The implementation of Markov Blanket Feature Selection Algorithm.

* Correspondence: Zhifa.Liu@stjude.org
[2]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA
Full list of author information is available at the end of the article

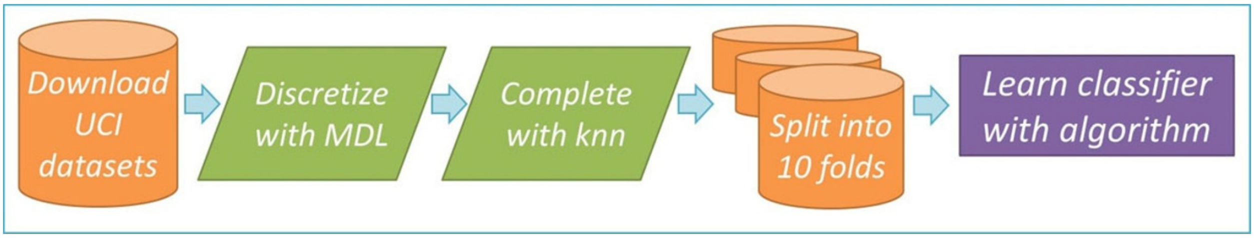


Figure 2 The procedure of data processing and classification.

information necessary to predict its value. In this work, we propose an algorithm to learn only the MB of the class variable; other attributes are removed. Therefore, our algorithm combines classification and feature selection. Results on benchmark machine learning datasets indicate that our feature selection technique usually reduces the size of the dataset more than 80% on some datasets. Accuracy results suggest that the classification ability of our algorithm is competitive with existing state of the art techniques.

Materials and methods

In a classification problem, we are given a dataset consisting of a set of attributes $\mathbf{A}$ and a class variable C. Furthermore, the dataset is split into a training set D_{tr} and a testing set D_{te}. The goal is to learn a classifier from D_{tr} that correctly predicts C in D_{te}. In this study, we compared the performance of our Markov Blanket structure with other classical classifiers such as C4.5 [1] , optimal Bayesian network [2], and Tree Augmented Naïve Bayes Network [3] and Markov Blank Hill Climbing [4]. Here is a general introduction for those classifiers.

Markov blanket feature selection algorithm

The intuition of this algorithm is that an attribute is either a parent, child or spouse of C, or the attribute is not in C's MB. Hence we only add each attribute to the MB according to an ordering and score for the new network. And we do not add attributes that make the score worse. Meanwhile, we keep the MB with the best score among all orderings. To this end, our algorithm performs feature selection and finds MB structure which has the maximum classification ability. The return structure is a Bayesian classifier for classification variable C. The general idea of the algorithm is shown in the Figure 1.

Experiments

We compared our feature selection algorithm to several state of the art classification methods on several benchmark datasets. All of the classification methods we selected learn a “human readable” model. In order to represent a wide variety of data domains, we downloaded 14 datasets from UCI machine learning repository [5]. The data processing and the classification steps in Figure 2 was followed a similar data procedure in [6].

Results

As shown in Figure 3 for compression ratio of these benchmark datasets, our feature selection often achieved quite high compression ratios by ignoring attributes which do not help predict C. From this, we infer that only a few attributes are necessary to predict C.

The accuracy results in Figure 4 demonstrate that, despite compressing the data over 80% in some cases, MB feature selection is still competitive in terms of accuracy with state of the art methods.

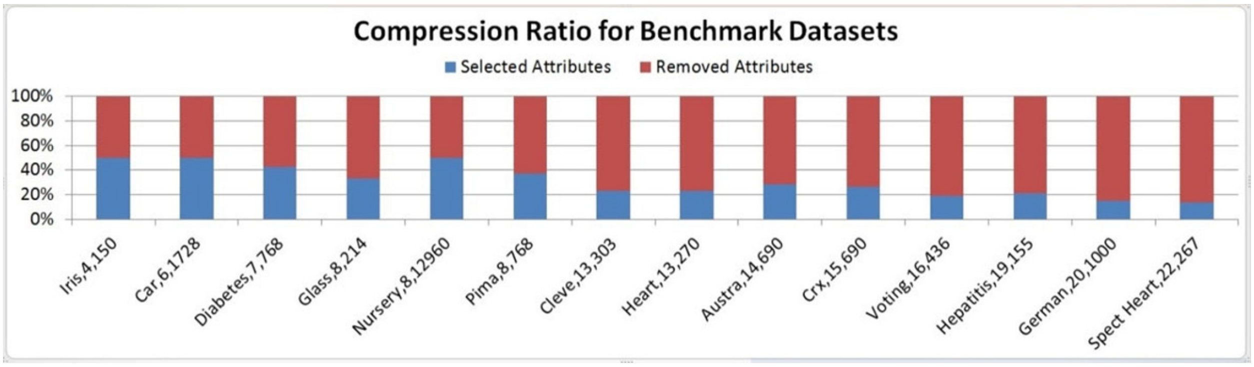


Figure 3 The Compression Ratio for Benchmark Datasets by Markov Blanket Feature Selection Algorithm.

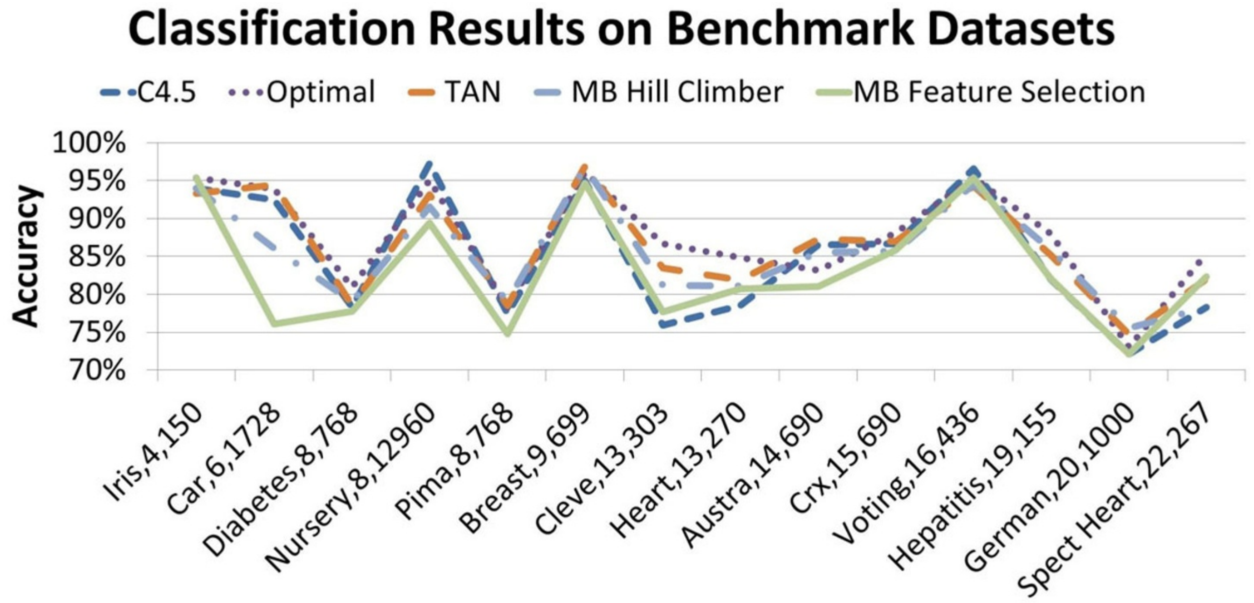


Figure 4 The classification results for Benchmark Datasets by Markov Blanket Feature Selection Algorithm.

Discussion and conclusions

The compression ratio decreases as the number of variables in the dataset increases. This suggests that, even as dataset sizes increase, only a few attributes are helpful in predicting the class variable. The compression ratio is unaffected by the number of records in the dataset. This suggests that even when given many records, our algorithm does not pick many attributes in an attempt to overfit the dataset. Ignoring unimportant attributes does not significantly affect the classification accuracy. Despite compressing the data on average more than 70%, the classification accuracy is rarely more that 5% below the best classifier. Identifying MB variables could significantly reduce the cost of diagnostic lab tests by focusing interest on only the most relevant attributes.

Authors' details

[1]Department of Statistics, Mississippi State University, Starkville, MS 39759, USA. [2]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.

Published: 22 October 2013

References

1. Quinlan JR: **C4.5: programs for machine learning.** *Machine Learning* 1994, **16(3)**:235-240.
2. Malone B, Yuan C, Hansen E, Bridges S: **Improving the scalability of optimal Bayesian network learning with external-memory frontier breadth-first branch and bound search.** In *Proceedings of the Twenty-Seventh Annual Conference on Uncertainty in Artificial Intelligence. Volume 2011.* Barcelona: AUAI Press;Fabio G. Cozman and Avi Pfeffer 2011:479-488.
3. Friedman N, Geiger D, Goldszmidt M: **Bayesian network classifiers.** *Machine Learning* 1997, **29**:131-163.
4. Tsamardinos I, Brown LE, Aliferis CF: **The max-min hill-climbing Bayesian network structure learning algorithm.** *Machine Learning* 2006, **65(1)**:31-78.
5. Bache K, Lichman M: **UCI Machine Learning Repository.**[http://archive.ics.uci.edu/ml].
6. Liu Z, Malone B, Yuan C: **Empirical evaluation of scoring functions for Bayesian network model selection.** *BMC Bioinformatics* 2012, **13(Suppl 15)**: S14.

doi:10.1186/1471-2105-14-S17-A3
Cite this article as: Tan and Liu: **Feature selection and prediction with a Markov blanket structure learning algorithm.** *BMC Bioinformatics* 2013 **14** (Suppl 17):A3.