

Bayesian Network Classifiers versus k-NN Classifier using Sequential Feature Selection

Franz Pernkopf
pernkopf@tugraz.at

University of Washington, Department of Electrical Engineering
M254 EE/CSE Building, Box 352500, Seattle, WA, USA

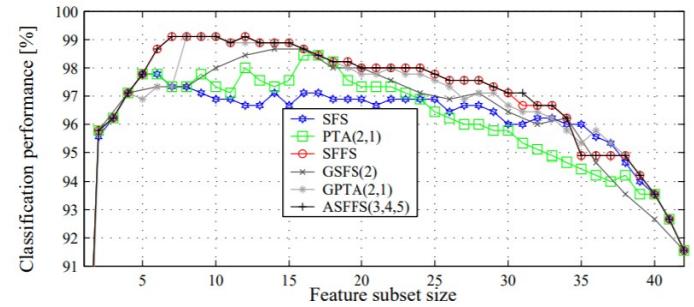
In the broad field of Machine Learning there are numerous classification techniques based on various metrics. The main motivation behind this paper was to use two classification techniques namely Bayesian Network Classifier and k-NN Classifier and show the optimum technique or algorithm with context to sequential feature selection.

The problem statement identified in this paper was to compare the Bayesian Network Classifier with the k-NN classifier and view the performance of both the classification methods on a subset of features using sequential feature selection.

This problem statement was approached in a three fold manner. Firstly the types of Bayesian classifiers were looked at. Three main Bayesian Classifiers were discussed the first being the Naive Bayes Classifier. This classifier assumes that all the attributes are conditionally independent given the class label. It can be viewed as a single Class node with many child attribute nodes connected only to the parent(Class) node. The performance of this classifier has been found to be good despite the fact of independence of the attributes and the unrealistic assumption it carries. The second being the Tree Augmented Naive Bayes Classifier. This classifier assumes that all the feature attributes are at most correlated to one other feature attribute(child node). It can be viewed as a single Class node with many child feature attribute nodes connected only to the parent(Class) node and one other attribute(child) node. The performance of this classifier is better than the Naive Bayes Classifier and is more realistic and implementable due to correlation with other attributes. The third is the Selective Unrestricted Bayesian Network Classifier. The Class node is treated equally as the attribute node and may be considered to have attribute node as its parent. As a result it does not follow a definite parent-child node structure and the attribute node need not be connected to the Class node. However this Classifier brings with it the drawback of requiring high degree of computation and may increase the probability of unreliable outcomes.

The next step was to determine the sequential feature selection algorithms and determine the most appropriate fit for carrying out the classification. Various techniques were studied to determine the best feature subset. The first technique was Sequential Forward Selection(SFS) and Sequential Backward Selection(SBS) which either in each iteration adds one more feature to the subset or removes one more feature from the subset respectively. This method would however cause nesting of feature attributes hence another technique was looked on. Generalized Sequential Forward Selection (GSFS(r)) and Generalized Sequential Backward Selection (GSBS(r)) at each stage add or remove r features simultaneously from the subset of feature attributes. This brings down the amounting of nesting attributes in the subset. Sequential Forward Floating Selection (SFFS) and Sequential Backward Floating Selection (SBFS) are the more preferred techniques which employ floating method techniques to rectify any error or wrong decision taken in the previous steps and the solution is as a result much more optimised. SFFS is an iterative forward selection technique while SBFS is an iterative backward selection technique. The most optimal solution however is the Adaptive Sequential Forward Floating Selection(ASFFS(r)) and Adaptive Sequential Backward Floating Selection(ASBFS(r)) which utilise both the principles of floating methods to correct any past decision and take or remove r features respectively,to avoid nesting of attribute features.

The final step was to carry out the experiment and in this paper a study was taken on a surface inspection task. In order to learn the various Bayesian models used in the experiment five-fold cross validation was used to improve accuracy. In order to judge the performance of the Bayesian networks five-fold cross validation was also applied with k-NN to act as a scoring function for all the Bayesian Networks and how they were trained via numerous different feature selection methods as discussed above. The classifiers used included: Naive Bayes classifier(NB), NB using the classical floating search (CFS-SNB), Tree augmented NB using hill-climbing search(HCS-TAN), Tree augmented NB using CFS (CFS-TAN), Unrestricted Bayesian network using CFS(CFS-SUN), k-NN using continuous-valued data and the SFFS method(SFFS-k-NN-C), k-NN using discrete-valued data and the SFFS method(SFFS-k-NN-D).



	%CV5	%H	#Arcs	#Features	#Evaluations	#Parameters
NB	89.11 ± 2.8727	95.45	40	40	1	275
CFS-SNB	96.44 ± 2.11	96.96	20	20	2098	122
HCS-TAN	97.11 ± 2.02	96.96	52	40	17893	533
CFS-TAN	97.11 ± 2.02	96.96	52	40	17958	533
CFS-SUN	98.66 ± 0.81	98.48	14	12	4097	230
SFFS-1-NN-C	99.33 ± 0.60	98.48	-	8	2873	450 samples × 8 features
SFFS-3-NN-C	99.11 ± 0.49	98.48	-	8	5086	450 samples × 8 features
SFFS-5-NN-C	98.44 ± 0.60	98.48	-	7	4346	450 samples × 7 features
SFFS-9-NN-C	98.44 ± 0.99	98.48	-	9	5086	450 samples × 9 features
SFFS-1-NN-D	96.22 ± 2.02	86.36	-	15	6003	450 samples × 15 features
SFFS-3-NN-D	96.44 ± 2.53	90.90	-	23	4803	450 samples × 23 features
SFFS-5-NN-D	96.22 ± 2.02	93.94	-	26	5302	450 samples × 26 features
SFFS-9-NN-D	96.00 ± 1.26	93.94	-	20	4745	450 samples × 20 features

Figure 1: Cross-validation classification performance of different sequential forward feature selection methods for a subset size up to 42.

Figure 2: Results of the experiment with various classification approaches

The data set consisted of 516 surface segments from a surface inspection task, and were distributed among 3 classes. 42 features were present in every data point or surface segment in this case.

The algorithms are then compared with the 3-NN classifier to judge cross validation performance. As seen from figure 1 different algorithms perform differently for various subsets. However the floating algorithms are shown to perform better than non floating algorithms. SFS and PTA algorithms are the most sub-optimal performing algorithms for various subset sizes. As seen from the figure the best performing algorithm is the SFFS algorithm and the remaining part of the experiment was chosen to be carried out with this algorithm primarily.

The next set of evaluations are carried out with metrics being (CV5) as the 5-fold cross validation classification accuracy estimate and (H) being the performance of algorithm on holdout data set. Figure 2 also shows the number of Evaluations, Parameters, Features and Arcs used in the classification accuracy estimate. The bold entries show the best performing classification accuracy result. The NB classification fares the most poorly while the CFS-SUN algorithm achieves the best classification accuracy estimate among all the Bayesian Network classifiers. However the selective k-NN (SFFS) classifier for continuous attributes performs better than the CFS-SUN algorithm as shown the bold accuracy estimates. For discretized feature spaces the selective k-NN (SFFS) classifier for discrete attribute's performance degrades. k-NN algorithms are time consuming as well as memory consuming, while the Bayesian network classifiers outperform selective k-NN methods in these aspects while the CFS-SUN algorithm maintains a high and comparable predictive accuracy.

Overall Bayesian network classifiers more often achieve a better classification rate compared to selective k-NN classifiers. The k-NN classifier performs well in the case where the number of samples for learning the parameters of the Bayesian network is small as the Bayesian network classifiers outperform selective k-NN methods in terms of memory requirements and computational demands for larger parameters. Especially, the performance of the selective unrestricted(CFS-SUN) Bayesian network classifier demonstrates the strength of Bayesian network classifiers.

[1] Friedman, N.; Geiger, D.; and Goldszmidt, M. 1997. Bayesian network classifiers. *Machine Learning* 29:131–163.

[2] Pernkopf, F. 2003. 3D surface analysis using Bayesian network classifiers. Technical report, Graz University of Technology.