

Multi Criteria Wrapper Improvements to Naive Bayes Learning

José Carlos Cortizo

Jccp@ainetsolutions.com

Artificial Intelligence & Net. Solutions,
Universidad Europea de Madrid





OUTLINE

- Motivation
- FBL Algorithm
- Experiments
- Conclusions
- Current work



MOTIVATION

- Classification algorithms' performance depends on instances description
- Work in Feature Subset Selection
 - Filter Model
 - Wrapper Model

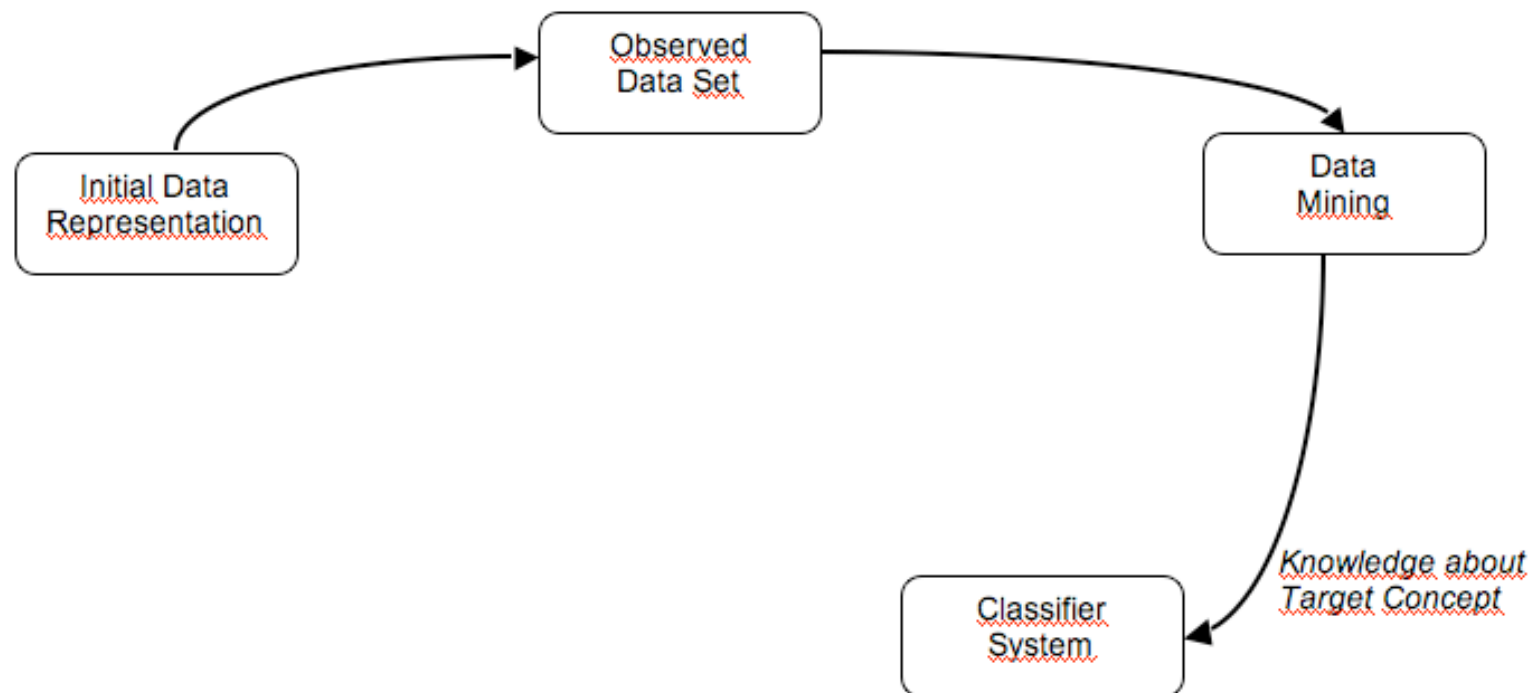


MOTIVATION

- Naive Bayes Classifier
 - Assuming independent attributes
 - $P(a_1, a_2 \dots a_n | v_j) = P(a_1 | v_j) \cdot P(a_2 | v_j) \dots P(a_n | v_j)$
 - $v_m = \operatorname{argmax}_{v_j \in V} P(v_j) \prod P(a_i | v_j)$
- Goal: *Obtain an extended-Naive Bayes algorithm robust to attribute dependencies*

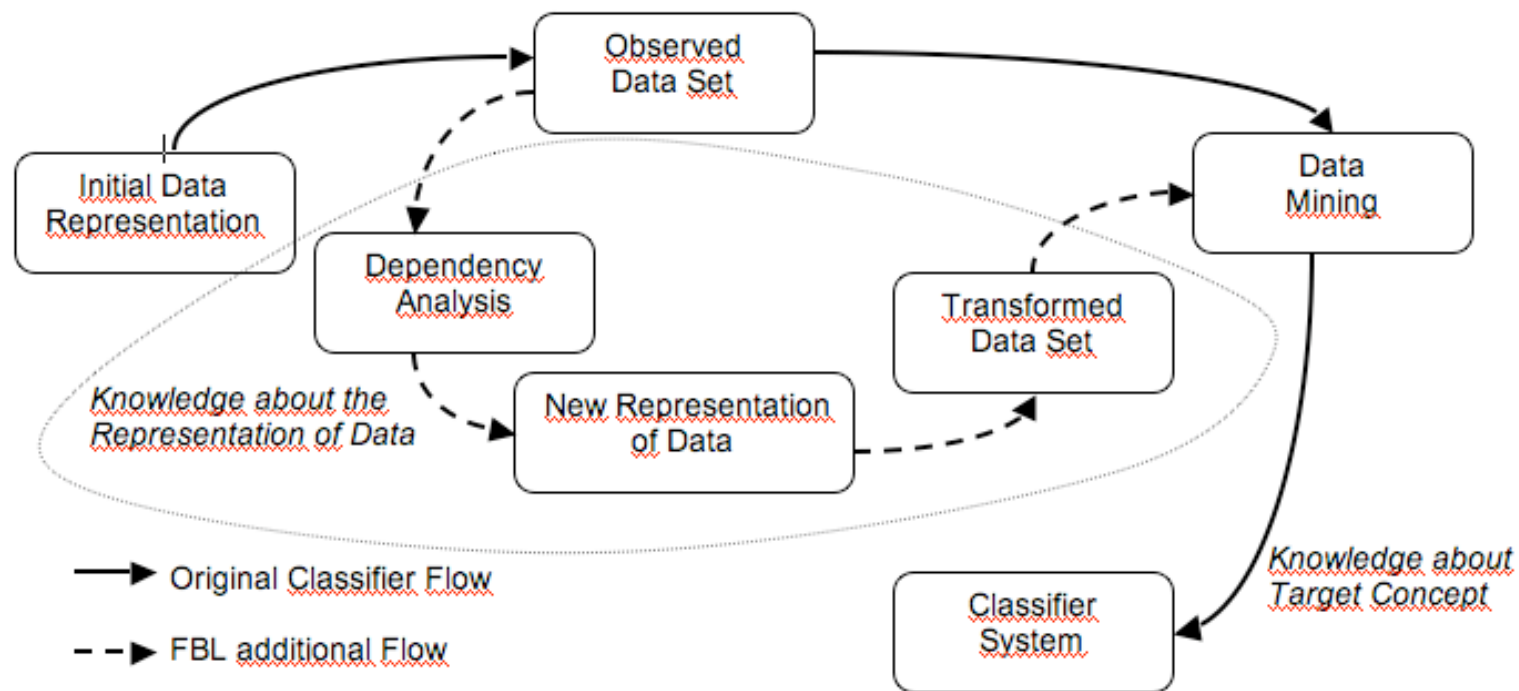
FBL ALGORITHM

- Naive Bayes data flow



FBL ALGORITHM

- FBL Extended data flow





FBL ALGORITHM

- FBL
 - Wrapper
 - Search guided by dependencies among attributes

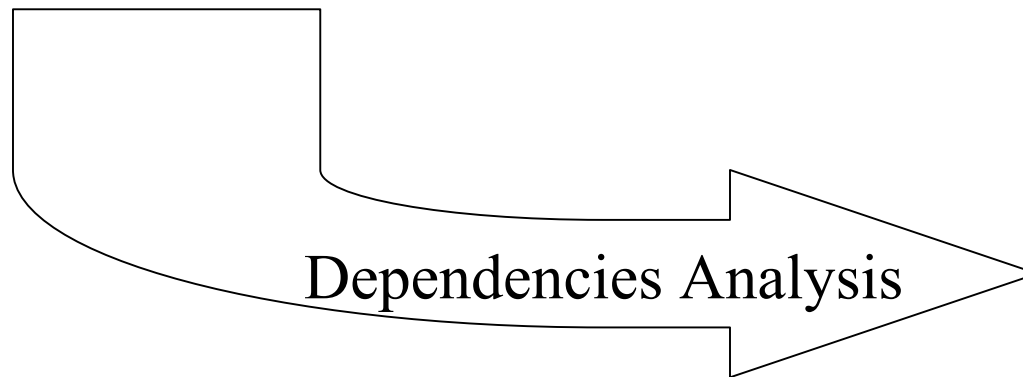


FBL ALGORITHM

- Steps
 - Initialization
 - Dependencies Analysis

FBL ALGORITHM

Attributes = {att1, att2...att6}



[att2,att5] : 0.97

[att1,att3] : 0.89

[att4,att3] : 0.88

[att5,att1] : 0.81

[att2,att1] : 0.68

[att5,att3] : 0.54

[att3,att2] : 0.53

...



FBL ALGORITHM

- Steps
 - Initialization
 - Dependencies Analysis
 - Dependency based wrapping

FBL ALGORITHM

1)

[att2,att5] : 0.97

[att1,att3] : 0.89

[att4,att3] : 0.88

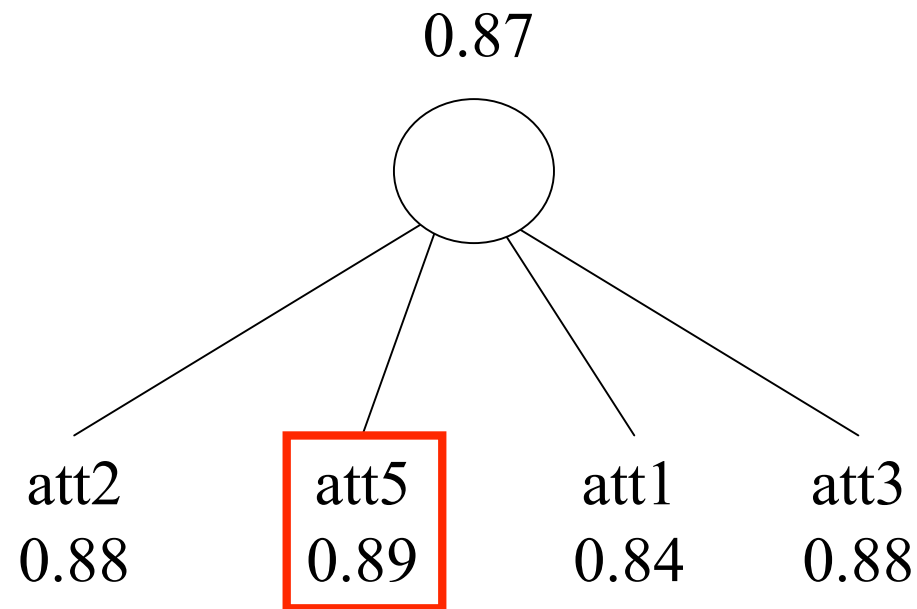
[att5,att1] : 0.81

[att2,att1] : 0.68

[att5,att3] : 0.54

[att3,att2] : 0.53

...



FBL ALGORITHM

2)

~~[att2,att5] : 0.97~~

[att1,att3] : 0.89

[att4,att3] : 0.88

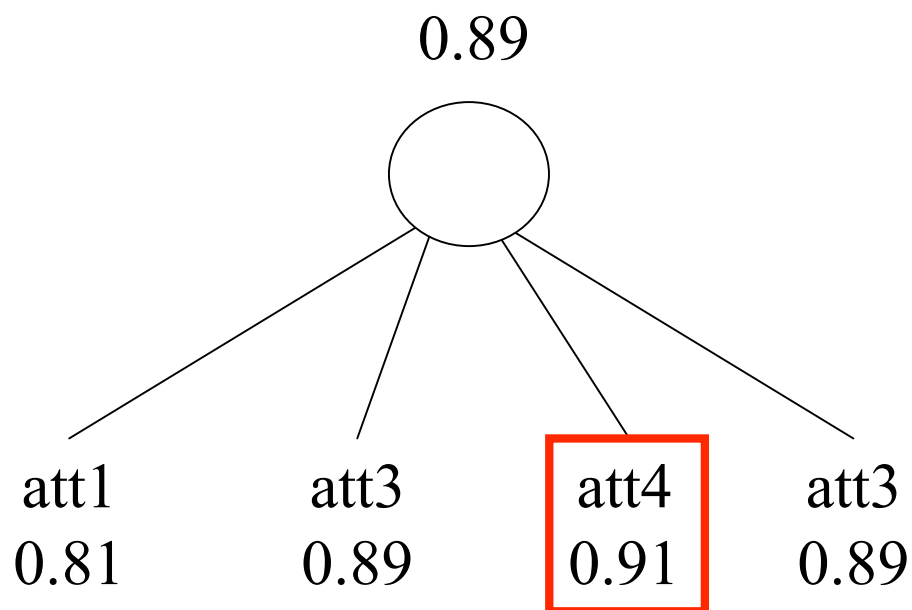
~~[att5,att1] : 0.81~~

[att2,att1] : 0.68

~~[att5,att3] : 0.54~~

[att3,att2] : 0.53

...



FBL ALGORITHM

3)

~~[att2,att5] : 0.97~~

~~[att1,att3] : 0.89~~

~~[att4,att3] : 0.88~~

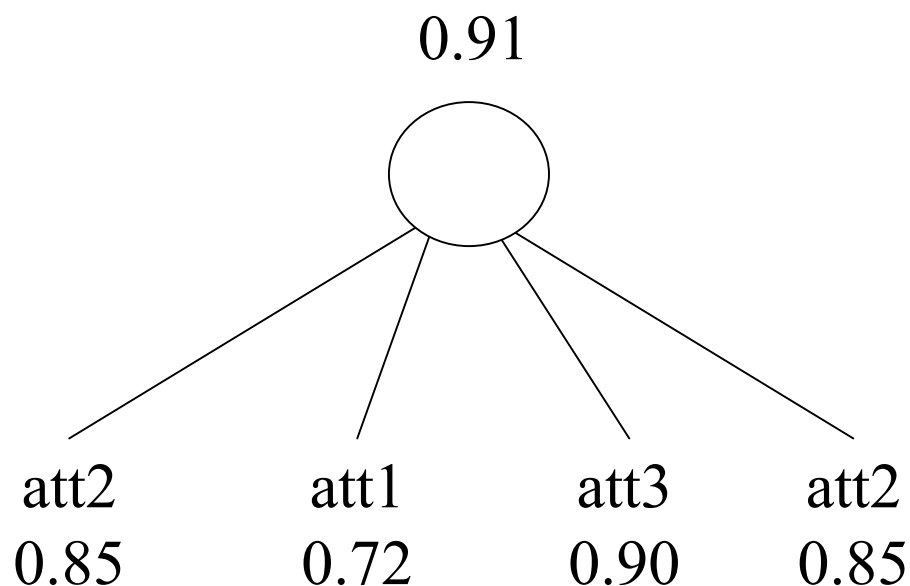
~~[att5,att1] : 0.81~~

[att2,att1] : 0.68

~~[att5,att3] : 0.54~~

[att3,att2] : 0.53

...



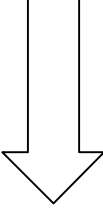
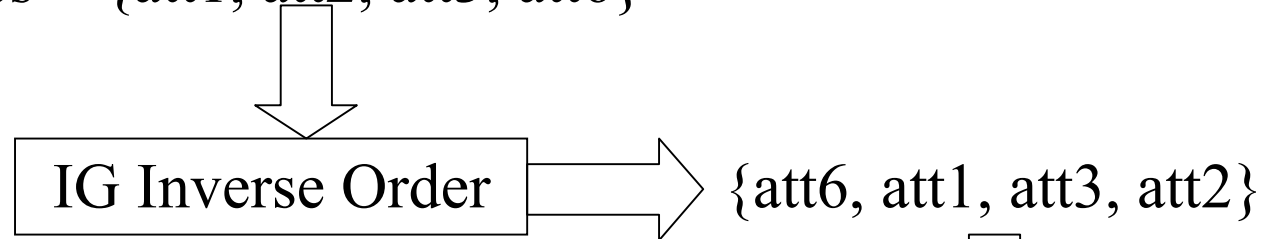


FBL ALGORITHM

- Steps
 - Initialization
 - Dependencies Analysis
 - Dependency based wrapping
 - IG based wrapping

FBL ALGORITHM

Attributes = {att1, att2, att3, att6}



rem att6? 0.92 OK
rem att1? 0.75 NO
rem att3? 0.68 NO
rem att2? 0.69 NO



EXPERIMENTS

- 2 stages
 - Is NB robust to dependencies?
 - Study the real performance



EXPERIMENTS

- First Stage
 - 5 original data sets
 - Add some synthetic dependencies
 - $\text{Synth} = a \cdot \text{Original} + b$
 - a, b random integers
 - 13 synthetic data sets
 - Classifiers (Weka): Naive Bayes, C4.5, C4.5 rules, KNN (N=1), KNN (N=4), SMO



EXPERIMENTS

Algorithm	CM1	CM2	CM3	BC1	BC2	BC3	BC4	BS1	BS2	BS3	GI1	WR1	TAE
Naive Bayes	-1.15	-2.45	-1.02	-0.14	-0.14	-0.14	-0.14	-9.11	-5.23	-0.68	-0.08	-0.07	-0.04
C4.5	-0.61	1.77	0.68	1.72	3.72	2.72	0.14	0.00	0.45	0.68	0.13	0.01	0.89
C4.5 rules	0.34	0.81	1.08	0.72	1.58	0.29	0.86	-2.74	-1.60	-0.91	0.72	1.58	-0.29
KNN(N=1)	-1.29	-0.61	-1.29	0.04	0.28	0.28	-0.15	-0.23	0.91	1.14	-0.13	-0.28	0.23
KNN(N=4)	-0.81	-0.88	-0.88	0.15	0.72	0.72	-0.43	-1.14	-0.91	-2.50	-1.16	-0.72	0.01
SMO	0.06	0.06	0.00	-0.14	-0.14	-0.14	0.00	0.22	-0.23	-0.46	0.00	-0.04	0.54



EXPERIMENTS

- Second Stage
 - Real data contains dependencies?
 - 15 datasets (without synth attributes)
 - Comparing FBL vs
 - Exhaustive Search
 - PCA (65, 75, 85, 95 % var.)
 - IG, Chi2 metrics based wrapper
 - Backward and Forward greedy searches (proposed by [13])



EXPERIMENTS

DataSet	NB	Best	O/B	FBL	IG	Chi ²	PCA	LF	LB
Abalone	0.240	0.266	9/2	0.266	0.266	0.266	0.245	0.265	0.265
Adult	0.827	0.835	15/13	0.835	0.752	0.752	ND	0.791	0.835
Cmc	0.508	0.554	10/4	0.554	0.508	0.508	0.453	0.553	0.537
Glass	0.495	0.603	10/3	0.579	0.556	0.556	0.542	0.598	0.598
Ionosphere	0.826	≥ 0.984	36/NA	0.915	0.863	0.872	0.920	0.984	0.900
Iris	0.960	0.967	5/2	0.967	0.967	0.967	0.933	0.960	0.953
Nursery	0.903	0.903	9/9	0.903	0.876	0.903	0.883	0.903	0.903
OpDigits	0.913	≥ 0.939	64/NA	0.913	0.865	0.881	0.939	0.925	0.918
PenDigits	0.857	0.865	16/13	0.865	0.615	0.836	0.886	0.864	0.864
Spam	0.793	≥ 0.908	57/NA	0.908	0.793	0.793	0.737	0.837	0.902
TAE	0.503	0.510	6/5	0.510	0.470	0.510	0.470	0.503	0.509
TicTacToe	0.696	0.724	9/5	0.718	0.699	0.699	0.742	0.718	0.728
WdbCancer	0.929	≥ 0.959	31/NA	0.956	0.924	0.923	0.937	0.959	0.945
Wine	0.966	0.989	13/10	0.977	0.792	0.977	0.983	0.989	0.977
Yeast	0.577	0.577	7/7	0.577	0.577	0.577	0.558	0.577	0.577



EXPERIMENTS

Algorithm	N.B.	FBL	PCA	LF	LB	IG	Chi ²
Avg. Accuracy	0.7331	0.7629	0.7327	0.7619	0.7610	0.7016	0.7346
Trim	0.00%	11.17%	-0.15%	10.78%	10.45%	-11.79%	0.57%
Times Best	ND	8	3	6	4	3	3



CONCLUSIONS

- FBL performs, at least, as well as the forward greedy search
- FBL is a guided search (less calls to the classifier)
- In 13/15 FBL improves NB
- FBL can be seen as the inverse approach of BN
 - FBL results in a simpler model
- FBL is robust. Final performance is, at least, equal to NB



CURRENT WORK

- We are comparing FBL with the state-of-the-art of wrappers (most of them included in weka) deeply
 - Wrappers
 - Rank Search (InfoGain, GainRatio, Chi2, ReliefF,...)
 - GeneticSearch
 - Greedy Stepwise (Forward, Backward)
 - BestFirst (Forward, Backward, Bidirectional)
 - Accuracies
 - Execution times
- More datasets (and bigger) are needed
 - 24 datasets
 - 4-65 attributes
 - 132-45222 examples



THANK YOU !