Prediction of Coffee Rust Disease Using Bayesian Networks

Cora B. Pérez-Ariza
Universidad de Granada, Spain (cora@decsai.ugr.es)

Ann E. Nicholson
Monash University, Australia (ann.nicholson@monash.edu)

M. Julia Flores
University of Castilla-La Mancha (julia.flores@uclm.es)

Abstract

In this paper we present an agricultural case study for learning Bayesian networks (BNs), namely prediction of coffee rust. Wide-spread in all major production areas, coffee rust causes premature defoliation, weakening the plant and reducing subsequent yield. It is typically controlled by use of chemical fungicides which must be applied before symptoms of infection are observed. Improved prediction would reduce the use of fungicides, producing healthier quality product and decreasing both economic costs and environmental impact.

We use a dataset obtained from an experimental farm in Brazil over 8 years. Our preliminary data analysis informed our pre-processing of the original dataset. We also identified a number of structural priors, which our BN learner, CaMML (Causal Minimum Message Length), incorporates into its scoring metric and hence into its structure learning. Previous research has applied other classification methods to this coffee rust dataset. We compare the results from a range of BNs learnt by CaMML with these previous approaches. The incorporation of structural priors in the BN learning yielded better models in terms of accuracy and interpretability. Although the BN's predictions were comparable with some of the other techniques, they were clearly worse than decision trees, which seem to be taking advantage of context sensitive cases; this suggests avenues for improving the quality of the BN models.

1 Introduction

In this work we present an application of Bayesian networks to an agricultural case study, namely the prediction of coffee rust. This is a devastating disease for coffee plantations, as it damages plants and reduces their yield through premature fall of infected leaves. Coffee rust has provoked intensive losses in many coffee-producing countries. It attacks all coffee species, being more severe in *Coffea arabica*. The disease first appeared in 1970 in the occidental hemisphere, in Brazil, and was rapidly extended to other major coffee producing countries in Central and South America, whose coffee trees are more sensitive to rust infection. Coffee rust is primarily caused by a fungus called *Hemileia vastatrix*. The symptoms of coffee rust infection include the appearance of spots on the upper leaf surface of plants. Its incidence can be measured by the percentage of leaves infected by the fungus.

In Brazil, damage from coffee rust leads to yield reduction of up to 35% in regions where climate conditions are propitious to the disease. The impact is thus considerable due to the economic importance of the coffee crop. The traditional way to prevent the disease is to apply agrochemical fungicides on fixed calendar dates. However, the fungicides both contaminate the environment and reduce the quality of the coffee. Moreover, as the intensity of the disease between seasons is subject to major variations, the use of agrochemicals is not always justified. The prevalence and impact of coffee rust have lead to many studies about it, including its prediction and how to decide whether to apply fungicides. Plant pathologists have tried to characterise this infection forecasting (Kushalappa, 1990), and more recently machine learning researchers in Brazil have applied decision trees (Meira et al., 2008), regression Support Vector Machines (SVM) (Luaces et al., 2010) and non-deterministic classifiers (Luaces et al., 2011) (described in more detail in Section 2), to the problem; but it has not yet been solved.

Bayesian networks have been applied to many real applications in environmental sciences, in fields such as farming, water resources, reforestation and ecological modelling; agricultural examples include fungicide for mildew in wheat (Jensen, 1995) and growing barley without pesticides (Kristensen and Rasmussen, 2002). Aguilera et al. (2011) have recently reviewed BN applications in the environmental sciences published between 1990 and 2010. This background encouraged us to apply Bayesian networks to this particular agricultural case study, the prediction of coffee rust.

BNs may be built either by eliciting expert knowledge or by automated causal discovery. With the availability of a dataset, used by the aforementioned Brazilian machine learning (ML) researchers, our approach here is focused on ML. The dataset comprises monthly accounts of the incidence of disease on an experimental farm in Brazil over 8 years. Additionally, the dataset registers the values of variables known to stimulate the growth of fungus (Avelino et al., 2006): weather conditions, fruit load of the plantation, and spacing between plants. We describe the dataset and the preprocessing undertaken (e.g. variable selection, discretization) in Section 4.

In this paper, we perform experiments with a set of BNs (Section 5). We first construct two simple handcrafted BNs (ignoring the weather variables) for use as baseline models. Then we learn the BN from data, using the CaMML (Causal Minimum Message Length) BN learner (Section 3), on both the full dataset and a reduced dataset after applying variable selection. We also run CaMML with different structural priors, which constrain its search across possible models. In Section 6, we compare the handcrafted and learned BNs against standard prediction methods available in Weka (Witten and Frank, 2011) (Naive Bayes and decision trees), and against the reported results from the Brazilian researchers.

2 Previous ML approaches

We now briefly describe the ML approaches previously applied to the Brazilian coffee rust dataset, as this informs the modelling choices we have taken with dataset pre-processing (described in Section 4), and provides an additional comparison for the results given in Section 5.

In (Meira et al., 2008), the authors develop a decision tree with the aim of aiding the understanding of coffee rust epidemics. The class variable was the change in infection rate, defined as the percentage of infected leafs, based on the monthly observations. They used three classes to classify the monthly change in the infection: reduction or stagnation ($\leq 0\%$); moderate growth (> 0 & $\leq 5\%$); and accelerated growth ($> 5\%$). They used 14 predictive variables: space, load, and weather measurements related to temperature, rain and relative humidity. The model correctly classified 73% using cross-validation, while the success rates were 88%, 57% and 79%, respectively, for these infection rate classes (i.e. the model did poorly on the moderate increase in incidence). The most important explanatory variables were found to be mean temperature during leaf wetness periods, expected yield, mean of maximum temperatures during the incubation period and relative air humidity.

More recently, these researchers presented a more sophisticated approach to the same problem (and dataset), using SVM regression (Luaces et al., 2010); rather than predicting the change in incidence, they consider actual incidence. They trained a SVM that provided a correlation of about 0.94 between predicted and actual incidences. However, when they try to devise an alarm system for predicting values above a given threshold, the number of false negatives turned to be too high (i.e. they were missing too many cases). So, they implemented what they call *nondeterministic* regressors, changing target points for intervals of fixed width. This allowed them to consider three possibilities: alarm, non-alarm and warnings (within a variable window around 4.5% incidence). This lowered the number of false negatives but, naturally, added false warnings.

Finally, (Luaces et al., 2011) is a further extension of this work to include additional nondeterministic classifiers (regressor SVM was adapted interpreting intervals as classes) defined by means of an optimization problem. To perform comparison between these alarm predictors, they present a framework where the costs of false negatives are higher than that of false positives, and both are higher than the cost of warning predictions. Under mild conditions they identified parameter settings, i.e. ranges of costs, where one of the nondeterministic classifiers outperforms the other learners.

In this paper we will analyse the behaviour of Bayesian networks for the coffee rust prediction problem. The advantages of BNs are (1) they are able to capture (in)dependencies existing among the variables involved, and the use of conditional probabilities make them capable of inherently dealing with uncertainty; (2) the graphical representation of relationships between variables facilitates the interpretation and formulation of conclusions about the domain of study; and (3) BNs can combine causal relationships with probabilistic logic, which helps to incorporate expert knowledge into the model. Next we describe the capabilities that the CaMML BN learner offers to do so.

3 The CaMML BN Learner

CaMML attempts to learn the best causal structure to account for the data, using a minimum message length (MML) metric (Wallace, 2005) with a two-phase search, simulated annealing followed by Markov Chain Monte Carlo search, over the model space. CaMML has been developed at Monash University over the past 16 years; see (Korb and Nicholson, 2010, Ch 9) for a full description. We used the version of most recent open-source version of CaMML,[1] which requires variables to be discretized. CaMML supports multiple ways of describing prior information about relationships between variables (O'Donnell et al., 2006), each of which can be accompanied by a confidence level. The types of priors are: full structure, direct causal connections, direct relation of unknonwn direction, causal dependency (ancestor relationship), correlation and temporal order (tiers) ($\{A_1, A_2, ...A_m\} \prec \{B_1, B_2, ...B_n\}$). In tiers, a partial ordering constrains arcs, based on the notion that tiers separate the variables on a timeline and that causality only occurs forward in time; this is standard in BN learning (e.g. Tetrad IV and K2). Thus $\prec$ means that the variables A_i occur before the B_j, which for CaMML becomes the structural constraint the A_i cannot be descendants of the variables B_j.

The expert may provide CaMML with a full structure or any combination of the prior types above, together with their confidence in each prior (expressed as a probability).[2] CaMML then combines these using the MML encoding of each kind of structural prior and the confidence probabilities, together with the default arc probability. For the experiments reported here, we provide CaMML with temporal tier priors, described in Sec. 5.2 below.

4 Description of the dataset

The data used for this work was obtained from an experimental farm[3] (Fundação Procafé, Varginha, Minas Gerais, Brazil) where the incidence of coffee rust was monitored from October 1998 to October 2006. The farm is organized into eight plots, where the rust is observed in different combinations of spacing between plants (dense or thin) and different fruit

[1] `https://github.com/rodneyodonnell/CaMML`

[2] Other BN learners also support the use of structural knowledge, but they have been limited to specifying one or two of these kinds of priors.

[3] This dataset, not yet publicly available, was provided to us by Luiz Henrique Rodrigues.

load (high or low). A meteorological station reports the temperature, humidity, wind speed, solar radiation and amount of rain registered every 30 minutes. Some metheorological information was missing in the original database due to failures in the station; we removed entries that contained missing data (as did (Meira et al., 2008), (Luaces et al., 2010) and (Luaces et al., 2011)). The rust observations were taken the first day of every month by picking 100 leaves from each plant of each plot and computing the average number of infected leaves.

We used the following data: (1) monthly information of the incidence of the coffee rust, and (2) a daily summary of all the weather observations taken that day.

Fig.1 shows the timeline for monthly observations, the possible timing of prediction (10, 15, 20, 25 or 30 days before next first day of the month), and the summary of daily weather conditions we used (the 45 days prior to the prediction day, as this is the potential period of infection). Note that for each target day, we know the incidence of coffee rust detected on the previous target day ($Incidence_{d0}$). After preprocessing, the final database comprises 1716 records of 34 variables. We ignore the time series aspect of the data and treat each record as independent and identically distributed (although we incorporate aspects of time through the month and year variables).

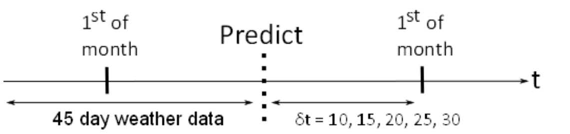

Figure 1: Sequence of prediction

First, we observed the incidence of coffee rust for all the registered months. Fig. 2 shows two examples of the time series (for high load plots, both thin and dense spacing); the periodicity of the coffee rust incidence is clear. The incidence of the rust over successive months (plotted in Fig. 3) is highly correlated (correlation coefficient $\rho_{I_{d0},I_{dt}} = 0.557157$). We use a threshold of 4.5% incidence to confirm an infection, the threshold typically used by farmers when deciding to chemically treat the plants, and used in (Luaces et al., 2010; Luaces et al., 2011). However, we discretized (by hand) the incidence variables into 5 states: [0], {0-4], {4-4.5], {4.5-5] and {5-100], in order to reflect different levels of infections (negative, warning and positive), around that threshold.

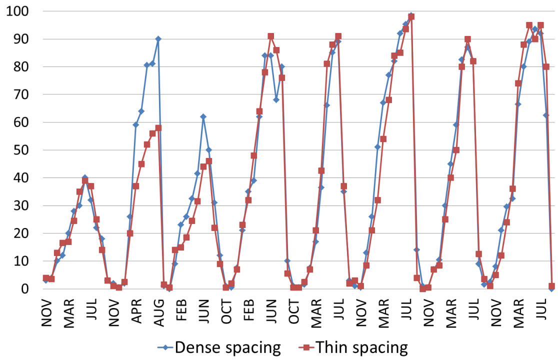

Figure 2: Monthly incidence of coffee rust for each month, in plots with highly loaded plants, for different spacings between plants

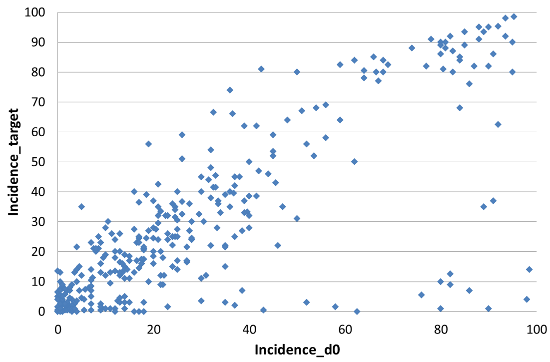

Figure 3: Correlation between the incidence of coffee rust and its previous presence

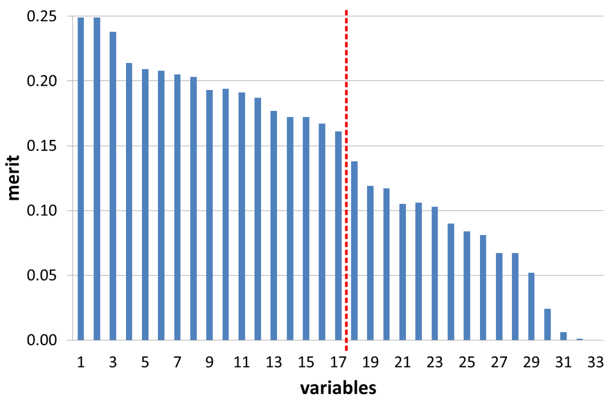

Figure 4: SU merit for the variables ranked.

Table 1: Variables used in the study. SU shows the attribution selection ranking, while $\star$ indicates the variable was used in experiments with dataset^r. The first block of variables corresponds to the incidence records, discretized by hand. The second block are all discrete variables, and the third block corresponds to the weather variables, all continuous and discretized using Weka.

Variable	Description	SU	($\star$)
Incidence_target	Percentage of leaves infected on target day (%) class variable		
Incidence_d0	Percentage of leaves infected on d0 (%)	4	$\star$
Year	year of measure	29	
Month	month of incidence	3	$\star$
Wet_season	YES if incubation period starts in Feb, March or May. NO otherwise	18	
Season	Summer(Dec-Feb), Autumn(Mar-May), Winter June-Aug), Spring(Sept-Nov)	1	$\star$
Load	fruit load of plants	30	
Spacing	spacing between plants	31	
Days_d0_now	days between previous 1st of month and prediction date	33	
Days_now_target	30, 25, 20, 15 & 10 days between prediction and next 1st of month	32	
Tmed_avg	average of temperatures during the period of infection (PINF) ($^\circ C$)	11	$\star$
Tmed_max	maximum of temperature during the PINF ($^\circ C$)	16	$\star$
Tmax_avg	average of maximum temperatures during the PINF ($^\circ C$)	25	
Tmax_max	maximum of maximum temperatures during the PINF ($^\circ C$)	20	
Tmin_avg	average of minimum temperatures furing the PINF ($^\circ C$)	12	$\star$
Tmin_min	Minimum of minimum temperatures furing the PINF ($^\circ C$)	10	$\star$
SolarRadiation_avg	Average of accumulated solar radiation during PINF (W/m^2?)	28	
SolarRadiation_acc	Accumulated of accumulated solar radiation during PINF (W/m^2?)	26	
NHoursSunlight_avg	Average of number of hours of sunlight during PINF	27	
WindSpeed_avg	Average of wind speed during PINF (Km/h)	22	
MaxWindSpeed_avg	Average of maximum wind speed reached during PINF (Km/h)	23	
MaxWindSpeed_max	Maximum of maximum wind speed reached during PINF (Km/h)	24	
Rain_avg	Average of accumulated rain during PINF (mm)	15	$\star$
Rain_acc	Accumulated of accumulated rain during PINF (mm)	13	$\star$
RelativeHumidity_avg	Average relative humidity during PINF (%)	7	$\star$
RelativeHumidity_max	Maximum relative humidity during PINF (%)	21	
NHRH95_avg	Average number of hours with rel. humidity > 95% during PINF	17	$\star$
NHRH95_max	Maximum number of hours with rel. humidity > 95% during PINF	19	
Tmed_HRH95_avg	Average temperature of hours with rel. humidity > 95% during PINF ($^\circ C$)	7	$\star$
Tmed_HRH95_max	Maximum temperature of hours with rel. humidity > 95% during PINF ($^\circ C$)	6	$\star$
NHNRH95_avg	Average of hours during night with rel. humidity > 95% during PINF	14	$\star$
NHNRH95_max	Maximum of hours during night with rel. humidity > 95% during PINF	9	$\star$
Tmed_NHNRH95_avg	Average temp. of hours during night with rel. humidity > 95% during PINF	8	$\star$
Tmed_NHNRH95_max	Maximum temp. of hours during night with rel. humidity > 95% during PINF	2	$\star$

The full set of variables in our dataset are shown in Table 1, with the plot attribute variables and time related variables above, followed by the weather variables summarising conditions over the 45 days period of infection. We discretized all the non-discrete variables (other than the incidence) automatically, using equal frequency discretization with 3 bins by Weka implementation. We also obtained a reduced dataset, dataset^r, applying Symmetrical Uncertainty (SU) with respect to the class[4] using cross-validation. Fig 4 shows the SU score for each variables, with our chosen cut point for the reduced dataset experiments (after the 17^{th} variable). The variables left out are weather variables related to wind conditions and solar radiation, the year, the variables related to the days before and after the prediction, and somewhat surprisingly, the variables related to the load and spacing conditions of the plot.

[4] $SU(C, A_i) = 2 \cdot \frac{H(C)-H(C|A_i)}{H(C)+H(A_i)}$, where C is the class, A_i an attribute and H() means entropy.

5 BNs for predicting coffee rust

5.1 Base cases: handcrafted BNs

We handcrafted two simple BN models (see Fig.5), to use for baseline comparison. The first contains only two variables, $Incidence_{d0} \rightarrow$

$Incidence_{target}$, while in the second, we add the load and spacing variables.

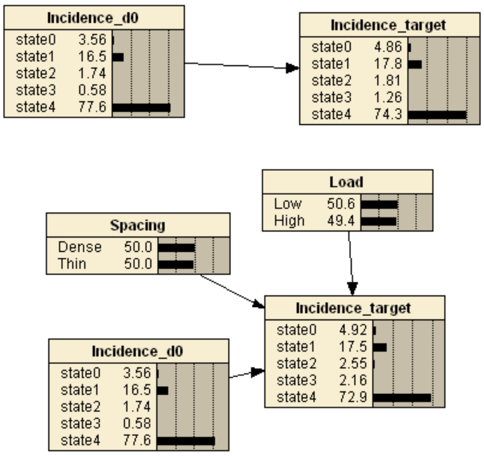

Figure 5: Simple BN handcrafted models

The parameters for both the handcrafted BNs were learnt (in Netica) with 80% of the available data. The other 20% was used for testing. The resultant confusion matrices are given in Table 2, showing error rates of 18.36% and 18.12%. Note that we combine the predictions into the following 3 categories (corresponding more closely to both the previous research and also the real world scenario):

- Negative: corresponds to States 0 and 1, 0-4% of leaves infected.
- Warning: corresponds to State 2, 4-4.5% of leaves infected.
- Positive: corresponds to States 3 and 4, $\geq$ 4.5% of leaves infected.

More interestingly, we also split the test data up into the cases where $Incidence_{d0}$ is Negative or Warning and Positive. We can see that there is a very large difference between the predictions when coffee rust was previously present (well predicted, in most cases still present) and when it was negative/warning (much less well predicted). However the negative/warning cases are really those that matter, given that, in the real world, given the presence of coffee rust, fungicides would have been applied.

The error rates for the simplest handcrafted model are 37.33% when the model has been learnt with the cases where the rust was not previously present, and 11.98% when $Incidence_{d0}$ is Positive. In the second model, the error rates are 34.66% and 11.98% respectively.

Table 2: Confusion matrices for the simple handcrafted BNs.

model		all data			neg. inc.			pos. inc.	
		$-$	$+$		$-$	$+$		$-$	$+$
First	$-$	55	25		45	25		0	0
	w	0	0		0	0		0	0
	$+$	38	225		3	2		32	235
		$-$	$+$		$-$	$+$		$-$	$+$
Second	$-$	56	25		49	26		0	0
	w	0	0		0	0		0	0
	$+$	37	225		0	0		32	235

5.2 BNs learnt by CaMML

We ran CaMML on both the full dataset (33 variables), and dataset^r (17 variables, after variable selection, as described above in Sec. 4), with no structural priors, with 10-fold cross validation. We used the CaMML default of $201n^3$ iterations (where n is the number of variables) in its search, and for each run tested with the "best" (lowest score) BN returned by CaMML. Fig. 6 shows one of the 10 BNs learned on reduced dataset^r. The lack of prior information in the models leads to some relations that are not present in the real world, e.g. weather variables influencing the month instead of vice versa. However, the class variable has as parents the month and a humidity related variable, which seems reasonable. Curiously, the variable $Incidence_{d0}$ is not directly linked to the target incidence.

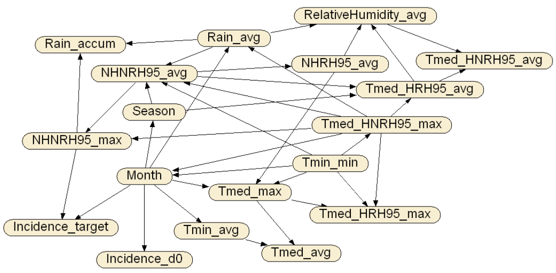

Figure 6: An example BN learned, without priors, from dataset^r

We also ran CaMML (both full and reduced dataset) with a range of structural priors; here (due to space) we describe only experiments on dataset^r, with the following simple tiers[5]:

1. $Incidence_{target}$ follows all other variables, using tiers ({All variables except Inc_{target}} $\prec$ { $Incidence_{target}$ }), confidence 1.
2. $Month$ and $Season$ (the only non-weather variables in $dataset^r$) come before all the weather variables, confidence 1.

Fig. 7 shows BNs learned on dataset^r, using the first tier (above) and with both tiers (below). These models are similar to that learned without priors, where CaMML had $Incidence_{target}$ as a leaf node. Note that now the parents of the class variable are the Season (itself a child of Month), a temperature related variable and $Incidence_{d0}$ (first tier only); these seem more logical given existing domain knowledge.[6]

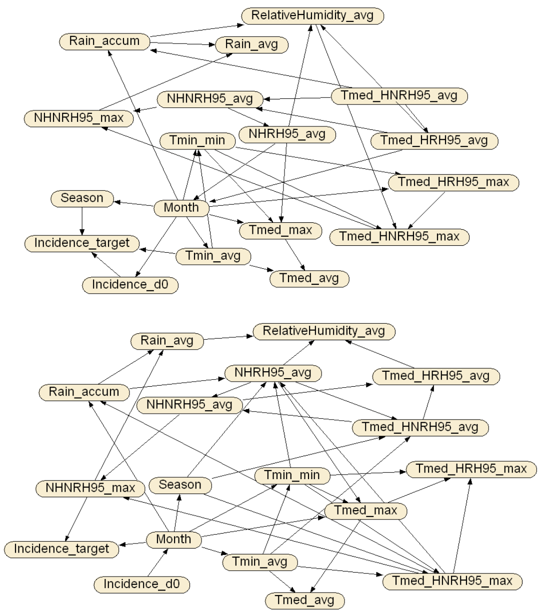

Figure 7: Example BNs learned from $dataset^r$ with first tier prior (above) and with both simple tiers (below).

Following the approach used in Sec. 5.1, Table 3 shows the confusion matrices obtained from the models learnt by CaMML with dataset^r. When learning without any prior information, the error rate is 9.84%. This rate decreases when learning with the first simple tier (9.32%) and with both tiers (8.97%).

Table 3: Confusion matrices for the models learned with dataset^r.

	no tiers			simple tier			two tiers	
	−	+		−	+		−	+
−	330	77		315	52		298	28
w	2	0		1	0		0	0
+	92	1215		108	1240		126	1264

6 Comparison and evaluation

The errors obtained in (Luaces et al., 2011) range from 1.47% to 8.82% depending on the parameterization, the regressors show error rates from 4.41% to 8.82%, and the nondeterministic classifiers range from the 2.94%-6.47% of the LibSVM to LibLINEAR's 1.47%-5.29%.

To compare our results to other standard ML classifiers , we learnt using Weka (and evaluating with 10-fold cross validation) a Naive Bayes (NB) model, a decision tree (C4.5), using Tree Augmented Naive Bayes (TAN) and using Averaged one-dependence estimators (AODE). The obtained results are shown in Table 4; the most notable is clearly the low error rate obtained by the decision tree on the full dataset. The analysis of this decision tree is illuminating, as it shows that for the different states of the variable $Incidence_{d0}$, the variables that influence the decision change. The presence of these context-specific independencies (Boutilier et al., 1996) in the data partly explains why the BN learned models do not perform as well even with the presence of prior information.

Table 4: Error rates for the Weka classifiers.

	NB	C4.5	TAN	AODE
dataset	12.82	0.64	5.65	5.36
dataset^r	13.9	8.3	7.63	7.16

Overall, our results using learned BNs are not that far from the previous studies, taking into account that those were tuned in their parametrizations and that this is the first study where BNs have been applied to this dataset.

[5] On the full dataset we used other types of priors, such as excluding possible connections between the background variables $Load$, $Spacing$, $Month$ and $Year$.

[6] Although the arc from $Incidence_{d0} \rightarrow Month$ with both tiers is undesirable; in our next iteration we'll add priors to make $Month$ a root node.

The learned BN are better predictors than both the simple BNs (as expected) and NB, and comparable to AODE, TAN and C4.5 using dataset^r. Finally, our results show how the incorporation of priors improves the performance of the model somewhat while producing more plausible causal structures.

7 Main conclusions and further work

We have described the novel application of probabilistic graphical models to coffee rust prediction, an agricultural case study. This process required a significant effort in analyzing the data and preprocessing it, as understanding the nature of the problem is the first step towards solving it. We applied causal learning (including with structural priors) to obtain a Bayesian network that predicts reasonably well (including outperforming NB), although it does not outperform the previous approaches yet. We have discovered through the learning of a decision tree how the influence of certain variables change for different contexts of $Incidence_{d0}$, suggesting further analysis of this is warranted.

An extended investigation of different methods for variable selection and discretization, as well as trying different combinations of priors, may improve prediction results. We also plan to apply pre-processing techniques for imbalanced datasets, and look at some of the recent work on characterising classification problem complexity (which is known to affect various methods differently). In the end, improved performance may be not be possible if there just isn't enough data to learn such a complex graphical model, especially for the case with incidence $< 4.5\%$ the previous month.

Apart from doing further ML experimentation, the future work includes obtaining feedback from domain experts about the structures learned. By combining BNs with cost matrices, we aim to produce a decision network that will allow farmers to explore the tradeoffs when deciding whether to apply fungicides.

Acknowledgments

This work has been partially funded by FEDER funds and the Spanish Government (MICINN) through projects TIN2010-20900-C04-01, 03 and the FPI scholarship programme (BES-2008-002049). We thank Luiz Henrique Rodrigues for providing the coffee rust dataset and answering our queries about it, and Steven Mascaro for assisting with preliminary data analysis.

References

P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón. 2011. Bayesian networks in environmental modelling. *Environmental Modelling Software*, 26(12):1376–1388.

J. Avelino, H. Zelaya, A. Merlo, A. Pineda, M. Ordo nez, and S. Savary. 2006. The intensity of a coffee rust epidemic is dependent on production situations. *Ecological Modelling*, 197(34):431 – 447.

C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. 1996. Context-specific independence in bayesian networks. In *UAI*, pages 115–123.

A. L. Jensen. 1995. *A probabilistic model based decision support system for mildew management in winter wheat.* Ph.D. thesis, Aalborg University.

K. B. Korb and A. E. Nicholson. 2010. *Bayesian Artificial Intelligence.* Chapman & Hall/CRC, Boca Raton, second edition.

K. Kristensen and I. A. Rasmussen. 2002. The use of a Bayesian network in the design of a decision support system for growing malting barley without use of pesticides. *Computers and Electronics in Agriculture*, 33(3):197–217.

A. C. Kushalappa. 1990. Development of forecasts: Timing fungicide applications to manage coffee rust and carrot blight. *Canadian Journal of Plant Pathology*, 12(1):92–99.

O. Luaces, L. H. A. Rodrigues, C. A. A. Meira, J. R. Quevedo, and A. Bahamonde. 2010. Viability of an alarm predictor for coffee rust disease using interval regression. In *IEA/AIE (2)*, pages 337–346.

O. Luaces, L. H. A. Rodrigues, C. A. A. Meira, and A. Bahamonde. 2011. Using nondeterministic learners to alert on coffee rust disease. *Expert Syst. Appl.*, 38(11):14276–14283.

C. A. A. Meira, L. H. A. Rodrigues, and S. A. Moraes. 2008. Anàlise da epidemia da ferrugem do cafeeiro com árvore de decisão. *Tropical Plant Pathology*, 33:114–124.

R. O'Donnell, A. E. Nicholson, B. Han, K. B. Korb, M. J. Alam, and L. Hope. 2006. Causal discovery with prior information. In A. Sattar and B. H. Kang, editors, *AI 2006: Advances in Artificial Intelligence*, LNAI Series, pages 1162–1167. Springer-Verlag, Germany.

C. S. Wallace. 2005. *Statistical and Inductive Inference by Minimum Message Length.* Springer, Berlin, Germany.

I. H. Witten and E. Frank. 2011. *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, third edition.