Doctoral Thesis Project: Theoretical Studies and New Approaches to Bayesian Network Classifiers

Gherardo Varando

Computational Intelligence Group, Departamento de Inteligencia Artificial of
Universidad Politécnica de Madrid,
Avd. Montepríncipe s/n Boadilla del Monte 28660, Madrid, Spain
gherardo.varando@upm.es
http://cig.fi.upm.es/index.php/members/156

Informations

- Provisional title of the thesis: *Theoretical Studies and New Approaches to Bayesian Network Classifiers*.
- Thesis directors: Prof. Pedro Larrañaga and Prof. Concha Bielza.
- Research group: Computational Intelligence Group (CIG).
- Department: Departamento de Inteligencia Artificial.
- University: Universidad Politécnica de Madrid.
- Starting date: 01 November 2013 (officially enrolled in the PhD program in the 2014/2015 academic year).

Project Summary

Problem Description

Bayesian network (BN) classifier [1], [4] are classifications models that have shown to be competitive in many cases. The most famous variants of BN classifier are naive Bayes (NB) and tree-augmented naive Bayes (TAN). Despite their large use as classification algorithms NB, TAN and more generally Bayesian augmented naive Bayes classifier (BAN) has received not so much attention from a theoretical point of view, especially on what concerns their expressibility. At this respect, some crucial paper are, in chronological order: Minsky [8], in which the NB is shown to induce a linear decision function over binary variables, Peot [10], Domingos and Pazzani [3], Jaeger [5], Ling and Zhang [7]. Especially Peot [10] and Jaeger [5] deal with expressibility issue, that is, the problem of determine how many and which type of decision functions is possible to represent with a BN classifier over some predictor variables. Those results use polynomial to express the decision functions but have some limitations, first of all the variables considered are mainly binary, or the structure of the Bayesian networks considered are restricted to NB and TAN.

Somehow related to the expressibility problem is the fact, observed for example by Friedman et al. [4], that usually the BN structure search is not related with the classification task but instead with *modelling* jointly the random variables. The same can be said for the probability estimation, given the structure. That is, in general, the BN classifier is learned via maximum likelihood principle (or Bayesian approach) and not with the explicit aim of performing the best classification (as *support vector machines* are able to). Some attempts have been made to improve the classification performance of a NB or TAN classifier. The most *simple* approach is to perform variable selection. More elaborate ways are the so called weighted naive Bayes classifier [18] or the discriminative learning of NB [6], [9]. To our knowledge a direct, simple and theoretically solid method to discriminatively train a NB classifier (or a BAN in general) has not been developed in the literature.

First results on expressive power

We focus on Bayesian augmented naive Bayes classifier, that includes NB and TAN as special cases. We assume the predictor variables to be categorical random variables and the class variables to be a binary variable.

First of all we extended [16] the results of Minsky [8], Peot [10] and Jaeger [5] about the decision functions induced by Bayesian network classifier to BAN classifier with categorical predictor variables. In particular we proved that the decision functions induced by a BAN classifier could be represented by polynomials within a specific family that depends on the topological structure of the Bayesian network considered. Moreover, if the BAN structure does not admit V-structures (when two parents shared the same child but are not connected), the given polynomial representation provides a complete characterisation. In fact for every polynomial of the given family it is possible to find an equivalent BAN classifier, that is a classifier that induces the same decisions. Using the given polynomial representation we were able to prove upper bounds on the number of decision functions representable by BAN classifiers with a given structure. Those bounds are not sharp.

We have then extended the expressive analysis of BAN classifier to the multi-label setting [12], [17], comparing binary relevance and chain classifiers. We proved for the first time that chain classifiers using BAN as base models are theoretically more expressive than binary relevance method. In addition we computed a polynomial form of the multi-label decision function bounding also in this case the expressive power.

Future works

We expect now to be able to apply the polynomial representation framework to the learning phase of BAN classifier, mainly in two aspects:

- In model selection algorithm, developing algorithm to select different type of BAN structure.

– In the conditional probability tables estimations, replacing the usual MLE or Bayesian estimation with an estimation of parameters that aims to obtain the best performance in classification.

We have obtained some results on how to perform parameter estimation in NB and we hope to be able to implement the method and compare it with state-of-the-art approaches for parameters estimation.

Related work

Multi-output regression. Inspired by the work on the expressive power of chain classifiers and binary relevance we have also tackle the same problem in the multi-output regression setting. In particular we proved that chain regression and binary relevance method are equivalent if linear models with ordinary least squares estimation or ridge estimation are used as base models [2].

B-spline density estimation When the predictor of the classifier are continuous, usually Gaussian Bayesian network are used. We extended previous research, developing conditional density estimators, based on B-splines [11]. We would like now to use the B-spline estimation techniques to develop Bayesian network classifier where conditional densities are modelled by those type of conditional densities.

Relevance

In general this thesis has the objective to contribute in some theoretical aspects of machine learning, in particular on Bayesian network classifier. The results could be useful both as a theoretical framework for future research, both as foundations on which build solid algorithms to improve the performance of BN classifier. Since machine learning techniques are now commonly used in a lot of practical application such as text categorization, costumers classification, automatic detections and in general in a lot of web services, we think that our contribution could be useful and applicable to a lot of different problems in various domains. More in the specific, we are actually collaborating with colleagues in neuroscience, applying those techniques. We are successfully using B-spline densities estimations building NB classifier with continuous variables with the goal of automatically classify neurons starting from morphological and electrophisiological variables. We are also working in the multi-output regression problem between the set of electrophisiological and morphological variables. All the algorithms will be made available, for example two first R packages are currently available:

– Rbmop package [14].
– Rbnet package [15].

Moreover for making the application easier to non-experts we are implementing a simple graphical interface [13].

Project plan summary

- Expressibility analysis: (2014-2015)
  - Naive Bayes (Varando et al. 2015 [16]).
  - BAN in absence of V-structures (Varando et al. 2015 [16]).
  - General BAN (partly in Varando et al. 2015 [16]).
  - Sharp bounds on expressive power.
- Multi-label BN classifier (2014-2015)
  - Problem adaptation method: Binary Relevance and Chain classifier. (Varando et al. 2014 [12], Varando et al. 2015 [17].)
- Alternative learning algorithms (2015-2017)
  - NB (work in progress, expected preprint in 2016).
  - NB with continuous variables using B-spline (work in progress with application to neuroscience).
  - TAN.
  - General BAN.
  - General BAN with continuous and categorical variables.
  - Model selection.
- Applications and implementations as open source packages (2015-2017):
  - Rbmop: B-spline estimations [14].
  - Rbnet: Bayesian networks using B-spline densities [15].
  - BnetApp an Rbnet GUI, preview available [13].

Acknowledgments. The thesis and the project described here has been partially supported by the Spanish Ministry of Economy and Competitiveness through the Cajal Blue Brain (C080020-09) and TIN2013-41592-P projects and by the Regional Government of Madrid through the S2013/ICE-2845-CASI-CAM-CM project.

References

1. Concha Bielza and Pedro Larrañaga. Discrete Bayesian network classifier: A survey. *ACM Computing Surveys*, 47(1), 2014. Article 5.
2. Hanen Borchani, Gherardo Varando, Concha Bielza, and Pedro Larrañaga. A survey on multi-output regression. *WIREs Data Mining and Knowledge Discovery*, 2015, (in press).
3. Pedro Domingos and Michael Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3):103–130, 1997.
4. Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.
5. Manfred Jaeger. Probabilistic classifiers and the concepts they recognize. In Tom Fawcett and Nina Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning (ICML-03)*, pages 266–273. AAAI Press, 2003.
6. Yushi Jing, Vladimir Pavlović, and James M. Rehg. Efficient discriminative learning of bayesian network classifier via boosted augmented naive bayes. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 369–376. ACM, 2005.

7. Charles X. Ling and Huajie Zhang. The representational power of discrete Bayesian networks. *Journal of Machine Learning Research*, 3:709–721, 2003.
8. Marvin Minsky. Steps toward artificial intelligence. In *Computers and Thought*, pages 406–450. McGraw-Hill, 1961.
9. T. Mouratis and S. Kotsiantis. Increasing the accuracy of discriminative of multinomial bayesian classifier in text classification. In *Computer Sciences and Convergence Information Technology, 2009. ICCIT '09. Fourth International Conference on*, pages 1246–1251, 2009.
10. Mark A. Peot. Geometric implications of the naive Bayes assumption. In Horvitz Eric and Jensen Finn, editors, *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, pages 414–419. Morgan Kaufmann Publishers Inc., 1996.
11. G. Varando, P.L. Lopez-Cruz, T. Nielsen, P. Larrañaga, and C. Bielza. Conditional density approximations with mixtures of polynomials. *International Journal of Intelligent Systems*, 30(3):236–264, 2015.
12. Gherardo Varando, Concha Bielza, and Pedro Larrañaga. Expressive power of binary relevance and chain classifiers based on Bayesian networks for multi-label classification. In Linda C. van der Gaag and Ad J. Feelders, editors, *Probabilistic Graphical Models*, volume 8754 of *Lecture Notes in Computer Science*, pages 519–534. Springer, 2014.
13. Gherardo Varando, Concha Bielza, and Pedro Larrañaga. Bnetapp an rbnet graphical interface, 2015. https://gherardo.shinyapps.io/BnetApp.
14. Gherardo Varando, Concha Bielza, and Pedro Larrañaga. Rbmop: B-spline density estimations, 2015. (beta) github.com/gherardovarando/Rbmop.
15. Gherardo Varando, Concha Bielza, and Pedro Larrañaga. Rbnet: Bayesian networks with B-spline, 2015. (beta) github.com/gherardovarando/Rbnet.
16. Gherardo Varando, Concha Bielza, and Pedro Larrañaga. Decision boundary for discrete Bayesian network classifiers. *Journal of Machine Learning Research*, 2015, (in press).
17. Gherardo Varando, Concha Bielza, and Pedro Larrañaga. Decision functions for chain classifiers based on Bayesian networks for multi-label classification. *International Journal of Approximate Reasoning*, 2015, (in press).
18. Nayyar A. Zaidi, Jesus Cerquides, Mark J. Carman, and Geoffrey I. Webb. Alleviating naive Bayes attribute independence assumption by attribute weighting. *Journal of Machine Learning Research*, 14:1947–1988, 2013.