Engineering Applications of Artificial Intelligence 65 (2017) 29–42

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# A survey of the applications of Bayesian networks in agriculture

Brett Drury [a,*], Jorge Valverde-Rebaza [a], Maria-Fernanda Moura [b], Alneu de Andrade Lopes [a]

[a] ICMC, University of São Paulo - Av. Trabalhador São-carlense, 400, Centro, São Carlos - SP, 13566-590, Brazil
[b] Embrapa Agriculture Informatics - Av. Dr. André Tosello, 209 Cidade Universitária Campinas, SP, Brazil

ARTICLE INFO

*Keywords:*
Bayesian networks
Agriculture
Computational agriculture
Bayesian inference

ABSTRACT

The application of machine learning to agriculture is currently experiencing a "surge of interest" from the academic community as well as practitioners from industry. This increased attention has produced a number of differing approaches that use varying machine learning frameworks. It is arguable that Bayesian Networks are particularly suited to agricultural research due to their ability to reason with incomplete information and incorporate new information. Bayesian Networks are currently underrepresented in the machine learning applied to agriculture research literature, and to date there are no survey papers that currently centralize the state of the art. The aim of this paper is rectify the lack of a survey paper in this area by providing a self-contained resource that will: centralize the current state of the art, document the historical progression of Bayesian Networks in agriculture and indicate possible future lines of research as well as providing an introduction to Bayesian Networks for researchers who are new to the area.



## 1. Introduction

Agriculture is an important part of the economy in many developed and emerging countries (OECD, 2015a, b). The advancements in the field of agricultural science in areas such as artificial fertilizers and automation of manual processes has increased the efficiency of farms which in turn has lead to higher yields and lower food prices (Agarwal, 1983). Increasing yields have allowed countries to support larger populations, and more recently, produce alternatives to gasoline such as ethanol. Increasing yields not only has produced lower priced food, but has the potential to save lives. For example, Norman Borlaug is credited with saving a billion lives with his development of "high-yield, disease-resistant wheat varieties". Increased crop yields from conventional sources is predicted to level off or even reduce by 2050 because of climate change and its associated effects (Jaggard et al., 2010). The limitation of traditional sources for innovation in agriculture may necessitate the use of alternative data-driven approaches that allow farmers/farm managers to make optimal decisions (McQueen et al., 1995; Kumar et al., 2015b, a).

A sub-field of the data-driven approach is machine learning. Machine learning techniques can be used in a number of possible scenarios to assist the farm manager. An obvious example is the automation of farm processes such as disease identification and treatment. Automated disease identification and treatment can reduce costs by limiting the affect of the disease and reducing the amount of costly human medical intervention.

Machine learning has a large number of sub-fields as well as possible strategies. One such sub-field is Bayesian Networks. Bayesian networks are rich frameworks, and have been to applied to many real-world problems that range from biomedical to petro-physics (Wiegerinck et al., 2013). It is a claim of this paper that Bayesian Networks are suitable for the Agriculture domain because they are able to: 1. represent inter-dependency between: a. causes or b. factors, 2. reason with partial or uncertain information, 3. incorporate new information and 4. make new inferences with new information.

The structure of the remainder of this paper will be the following: Section 2 presents a quantitative and qualitative analysis of the reviewed publications, Section 3 presents a brief introduction to Bayesian Networks, Section 4 discusses the main agricultural problem areas that Bayesian Networks have been applied to, Section 5 discusses general trends of graph construction, structure learning, inference and evaluation in agricultural Bayesian Networks, Section 6 presents possible future directions of Bayesian Networks in the agricultural domain, and Section 7 summarizes the conclusion and the main contributions of this work.

* Corresponding author.
*E-mail address:* brett.drury@gmail.com (B. Drury).

http://dx.doi.org/10.1016/j.engappai.2017.07.003
Received 25 January 2017; Received in revised form 17 June 2017; Accepted 3 July 2017
0952-1976/

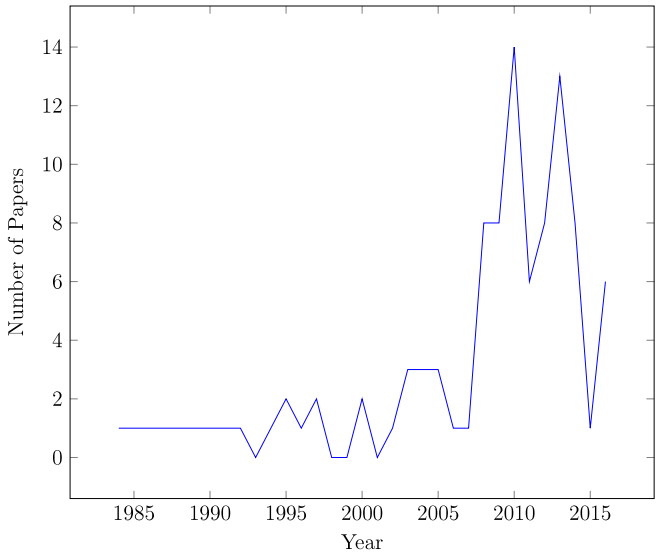


**Fig. 1.** Number of agricultural Bayesian network research papers published by year.

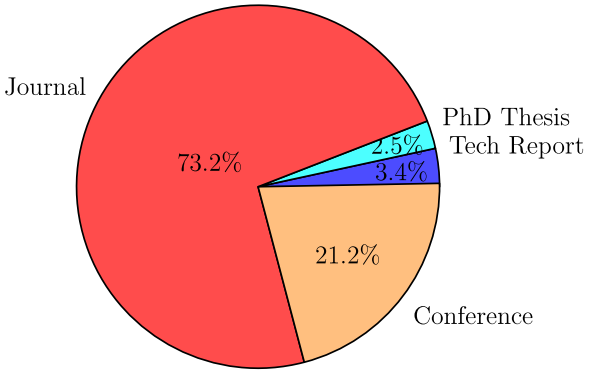


**Fig. 2.** Publication sources.

## 2. Analysis of the reviewed publications

The selection policy of this survey was to select articles that were published in peer reviewed locations such as: 1. academic journals, 2. academic conferences and 3. Ph.D. theses. The survey made an exception for high quality technical reports published in academic institutions.

The distribution of frequency of publication over time is demonstrated in Fig. 1. The time period covered by this review ranges from 1984 to 2016. Agricultural Bayesian Networks is a relatively mature field of study, but with a small number of papers. Although it should be noted that this may be due to the selection policy or the choice of publication location by agricultural researchers.

The frequency of the publication sources is demonstrated in Fig. 2. The majority of the papers identified for this survey were published in journals. Journal articles accounted for over 73%. The secondary source of publications were peer reviewed conferences. Ph.D. theses and technical reports accounted for relatively small number of papers, 2 and 3 papers respectively.

The articles identified for this survey were published in a number of different journals and conferences. The most frequent publication titles are in Fig. 3. The most frequent publication title was the *Journal of Preventive Veterinary Medicine*[1] with 18 articles. The most popular conference for the publication of articles about Bayesian Networks in agriculture was *Computer and Computing Technologies in Agriculture*[2] with 3 publications. The remaining publications published 3 or less articles in the field of Bayesian Networks in agriculture.

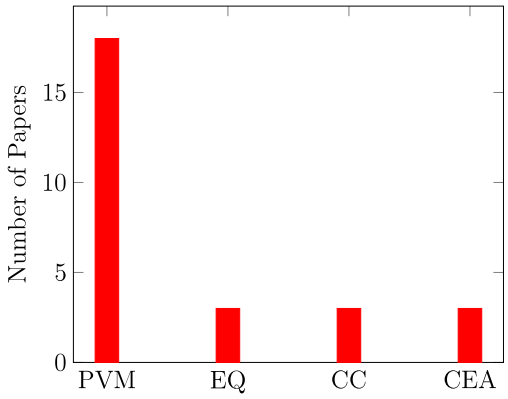


**Fig. 3.** Frequency of publication location, where PVM = Preventive Veterinary Medicine, EQ = Journal of Environment Quality, CC = Computer and Computing Technologies in Agriculture (Conference) and CEA = Computer Electronics in Agriculture (Minimum Publication Frequency is 3 publications).

In addition to the requirement that qualifying articles be published in peer reviewed locations, the selection policy required that the articles to be relevant. Relevancy for the purposes of this paper was that qualifying articles central theme must be an application to a problem in agriculture. Papers that addressed related areas such as: 1. environmental management, 2. aquaculture and 3. forestry, were excluded because they did not address an agricultural problem. The inclusion of these related papers would have alter the findings of this paper. Papers which used agricultural data sets to evaluate various Bayesian Network strategies such as structure and parameter learning, were also excluded because the central theme was not agriculture. This selection policy drastically reduced the number of qualifying papers.

The relatively small number of papers in this area is not a reflection of the applicability of Bayesian Networks to this domain. This paper, therefore, is intended to: 1. stimulate interest in the area of Bayesian Networks in agriculture, and 2. be a self-contained reference for researchers who are either new to the area of Bayesian Networks or to the application of Bayesian Networks in the agricultural domain.

## 3. Bayesian networks

The motivation of this section is to provide a brief introduction to Bayesian Networks. This section does not discuss Bayesian Networks in detail because it is beyond the scope of the article. Nevertheless, this section provides sufficient detail so that the reader: 1. will have a rudimentary grasp of Bayesian Network fundamentals and 2. will be able to locate further details. For these purposes this section discusses the following aspects of Bayesian Network fundamentals: 1. Bayesian Network definitions and assumptions, 2. Marginalization and inference, 3. Classification, 4. Structure learning, and 5. Strengths and Potential Limitations.

### 3.1. Fundamentals

Bayesian Networks have emerged from research into Artificial Intelligence as a formal means of analysing decision strategies under uncertain conditions. Bayesian Networks has Bayes' Theorem as a foundation. Bayes' theorem computes the posterior probability of the

[1] http://www.journals.elsevier.com/preventive-veterinary-medicine.

[2] http://www.iccta.cn/ch/index.aspx.