



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Incremental Learning de Redes Bayesianas e o UnBBayes

Pedro da C. Abreu

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Orientador

Prof. Dr. Marcelo Ladeira

Coorientador

Prof. Dr. Shou Matsumoto

Brasília
2015



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Incremental Learning de Redes Bayesianas e o UnBBayes

Pedro da C. Abreu

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Prof. Dr. Marcelo Ladeira (Orientador)
CIC/UnB

Prof. Dr. Donald Knuth Dr. Leslie Lamport
Stanford University Microsoft Research

Prof.^a Dr.^a Ada Lovelace
Coordenadora do Bacharelado em Ciência da Computação

Brasília, 07 de dezembro de 2015

Dedicatória

Na *dedicatória* o autor presta homenagem a alguma pessoa (ou grupo de pessoas) que têm significado especial na vida pessoal ou profissional. Por exemplo (e citando o poeta):
Eu dedico essa música a primeira garota que tá sentada ali na fila. Brigado!

Agradecimentos

Nos *agradecimentos*, o autor se dirige a pessoas ou instituições que contribuíram para elaboração do trabalho apresentado. Por exemplo: *Agradeço aos gigantes cujos ombros me permitiram enxergar mais longe. E a Google e Wikipédia.*

Resumo

Com o crescente interesse em aprendizado de máquina, frameworks robustos e implementações do estado da arte são uma necessidade do mundo moderno. Pensando nisso levantamos um estudo sobre o estado da arte de aprendizado no contexto de Rede Bayesiana (BN), já que BNs provêm um excelente modelo para lidar com relações causais e probabilidade bayesiana. Para tal utilizaremos o framework UnBBayes e desenvolveremos plugins para este.

Palavras-chave: LaTeX, metodologia científica, trabalho de conclusão de curso

Abstract

With the growing interest in machine learning, good frameworks and state of the art implementation are a need in modern world. With this in mind we provide a study in the state of the art of BN, since BNs provide a great model to deal with causal relations and bayesian probability. For such we shall use UnBBayes framework and implement plugins for it.

Keywords: LaTeX, scientific method, thesis

Sumário

1	Introdução	1
1.1	Some Theory here	2
1.2	Probabilidade física versus probabilidade Bayesiana	2
2	Redes Bayesianas	3
2.1	Definição Formal	4
2.2	D-separation	4
2.3	M-separation	4
2.4	Aprendizado	4
3	Aprendizado de Redes Bayesianas	5
3.1	Aprendizado de Parametros	5
3.2	Aprendizado de Estrutura	5
3.3	Aprendizado Incremental	5
4	UnBBayes	6
4.1	Plugins	6
4.1.1	Como desenvolver plugins	6
5	Design Patterns	7
5.1	Padrões de Comportamento	7
5.2	Padrões de Estrutura	7
5.3	Padrões de Construção	7
6	Arquitetura do UnBBayes	8
7	Resumo das Atividades	9
8	Resumo das atividades	10
8.1	O que foi feito	10
8.2	O que será feito	10

9 Conclusão	11
Referências	12

Lista de Figuras

2.1	extraída do artigo do Laecio	3
-----	--	---

Lista de Abreviaturas e Siglas

BN Bayesian Network.

CPT Conditional Probability Tables.

DAG Directed Acyclic Graph.

Capítulo 1

Introdução

BN é um modelo gráfico para relações probabilísticas dado conjunto de variáveis. Nas últimas décadas, redes Bayesianas se tornaram representações populares para codificar conhecimento especialista incerto para sistemas especialistas [2]. Mais recentemente, pesquisadores desenvolveram métodos de aprendizagem de redes Bayesianas a partir de dados. As técnicas desenvolvidas são novas e ainda em evolução, mas eles têm se mostrado muito eficientes para alguns problemas de análise de dados.

Existem diversas representações possíveis para análise de dados, entre elas, rule bases, decision trees, e redes neurais artificiais; e outras tantas como estimação de densidade, classificação, regressão e clusterings. Portanto o que métodos de BNs têm a oferecer? Segundo Heckerman [2] podemos oferecer pelo menos quatro respostas, sendo elas:

1. BNs lidam com um conjunto incompleto de dados de maneira natural.
2. BNs permitem aprender sobre as relações causais. Aprender sobre tais relações são importantes por pelo menos duas razões: O processo é útil quando se está tentando entender sobre um dado problema de domínio, como por exemplo, durante uma análise de dados exploratória. E mais, conhecimento de relações causais nos permitem fazer previsões na presença de intervenções. Por exemplo, um analista de mercado pode querer saber se é lucrativo aumentar o investimento em determinada propaganda para aumentar as vendas de seu produto. Para responder esta pergunta o analista pode determinar se esta propaganda é a causa para o aumento de suas vendas, e em caso afirmativo, quanto. O uso de BNs nos ajuda a responder tal pergunta até mesmo quando não há experimentos nos efeitos de tal propaganda.
3. BNs em conjunto com técnicas estatísticas bayesianas facilitam a combinação de conhecimento de domínio e dados. Qualquer um que tenha feito uma análise do mundo real sabe a importância de conhecimento prévio ou de domínio, em especial quando os dados são poucos ou caros. Pelo fato de alguns sistemas comerciais (i.e.,

sistemas especialistas) podem ser construídos a partir de conhecimentos prévios. BNs possuem uma semântica causal que permitem conhecimentos prévios serem representados de uma forma muito simples e natural. Além disto, BNs encapsulam tais relações causais com suas probabilidades. Consequentemente, conhecimento prévio e dados podem ser combinados com técnicas bem estudadas da estatística Bayesiana.

4. Métodos Bayesianos em conjunto com BNs e outros tipos de modelos oferecem uma forma eficiente para evitar over fitting dos dados. Como veremos, não há necessidade de excluir parte dos dados do treinamento do aprendizado da rede. Usando técnicas Bayesianas, modelos podem ser "suavizados" de tal forma que todo dado disponível pode ser usado para o treinamento.

1.1 Some Theory here

1.2 Probabilidade física versus probabilidade Bayesiana

Para entender BNs e as técnicas de aprendizado associadas, é importante entender a diferença entre a Probabilidade e Estatística padrão e a Bayesiana.

Resumidamente

Capítulo 2

Redes Bayesianas

Uma BN provê uma representação compacta de distribuições de probabilidades grandes demais para lidar usando especificações tradicionais e provê um método sistemático e localizado para incorporar informação probabilística sobre uma dada situação.

Uma BN é um Directed Acyclic Graph (DAG) que representa uma função de distribuição de probabilidades conjunta de variáveis que modelam certo domínio de conhecimento. Ela é constituída de uma DAG, de variáveis aleatórias (também chamadas de nós da rede), arcos direcionados da variável pai para a variável filha e uma Conditional Probability Tables (CPT) associada a cada variável.

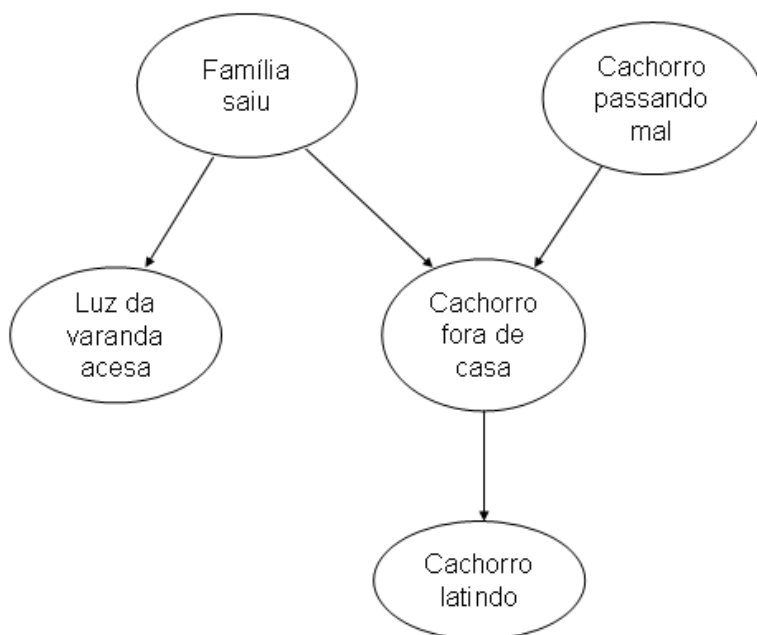


Figura 2.1: Exemplo family-out

Nesse exemplo, suponhamos que se queira determinar se a família está em casa ou se ela saiu. Pelo grafo, podemos perceber que o fato de a luz da varanda estar acesa e de o cachorro estar fora de casa são indícios de que a família tenha saído.

2.1 Definição Formal

Uma BN consiste em uma fatoração de uma distribuição de probabilidade e um DAG correspondente. Tais assertivas de independências condicionais podem ser inferidas diretamente da fatoração correspondente às

2.2 D-separation

2.3 M-separation

2.4 Aprendizado

Capítulo 3

Aprendizado de Redes Bayesianas

3.1 Aprendizado de Parametros

3.2 Aprendizado de Estrutura

3.3 Aprendizado Incremental

Capítulo 4

UnBBayes

4.1 Plugins

4.1.1 Como desenvolver plugins

Capítulo 5

Design Patterns

5.1 Padrões de Comportamento

5.2 Padrões de Estrutura

5.3 Padrões de Construção

Capítulo 6

Arquitetura do UnBBayes

Capítulo 7

Resumo das Atividades

Capítulo 8

Resumo das atividades

8.1 O que foi feito

8.2 O que será feito

Capítulo 9

Conclusão

Referências

- [1] George D. Greenwade. The Comprehensive Tex Archive Network (CTAN). *TUGBoat*, 14(3):342–351, 1993.
- [2] David Heckerman. A tutorial on learning with bayesian networks. 156:33–82, 2008. 1