

# 哈尔滨工业大学（深圳）

## 2017·FALL·ARTIFICIAL INTELLIGENCE

### PROJECT-2

## 任务概述

### 1. 目标

本次 Project 的目的是通过一个实际的应用场景对贝叶斯网络（Bayesian Networks）做更加深入的学习，从而对其在于真实案例中的概率表达有更加形象的理解。

### 2. 背景

医疗诊断（Medical Diagnosis）是本次 Project 的主题。

一些医学研究人员计划通过建立一个贝叶斯网络模型从而观察某些疾病及相应症状之间的相互关系，作为计算机领域的研究生，我们所需要做的工作是：根据病患的症状观察记录，学习并完善该贝叶斯网络模型中的概率参数。

不幸的是，在实际应用场景中病患的症状观察记录往往是不完整的，其在某些症状的观察结果上存在缺失，我们所需要做的就是病患记录存在一定量的缺失值的前提下尽力计算贝叶斯网络的概率参数使其能够应用于实际环境中的病情诊断。

### 3. 问题简述

本次 Project 将在医学研究人员已创建的贝叶斯网络上进行，该网络的结构如下图所示：

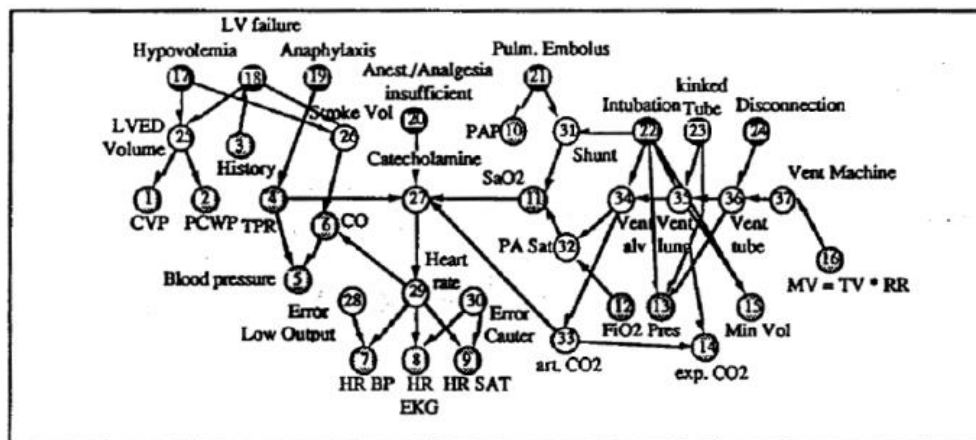


Fig. 1 The ALARM network representing causal relationships is shown with diagnostic (●), intermediate (○) and measurement (⊗) nodes. CO: cardiac output, CVP: central venous pressure, LVED volume: left ventricular end-diastolic volume, LV failure: left ventricular failure, MV: minute ventilation, PA Sat: pulmonary artery oxygen saturation, PAP: pulmonary artery pressure, PCWP: pulmonary capillary wedge pressure, Pres: breathing pressure, RR: respiratory rate, TPR: total peripheral resistance, TV: tidal volume

关于该贝叶斯网络的详细情况可参阅参阅：[\*The ALARM Monitoring System: A case Study with two Probabilistic Inference Techniques for Belief Networks.\*](#)

注意该贝叶斯网络，其中包含三类节点：

- (1) 诊断类节点（diagnostic），e.g. Hypovolemia;
- (2) 媒介类节点（intermediate），e.g. LVED;
- (3) 测量类节点（measurement），e.g. HR BP。

这样的贝叶斯网络可用多种格式表示，本次 Project 将采用的格式为 **.bif 格式（Bayesian Interchange Format）**，这是一种便于编程解析的格式，能够高效的表征贝叶斯网络。

本次 Project 的任务是通过病患的相关记录对上述的 ALARM 贝叶斯网络的概率参数进行学习。

## PROJECT 提供数据及代码情况介绍

为了便于各位同学快速开展自己的研究，本次 Project 提供部分代码及初始数据供大家使用，代码及数据的详细情况如下：

### 4. example.bif & alarm.bif

这两个 .bif 文件采取了**相同的格式编写**，前者用于代码测试及示范，后者则是本次 Project 主要的研究对象，以下将以 example.bif 文件内容为例进行介绍：

```
variable "B" { //2 values
    type discrete[2] { "True" "False" };
}
variable "M" { //2 values
    type discrete[2] { "True" "False" };
}
variable "I" { //2 values
    type discrete[2] { "True" "False" };
}
probability ( "B" ) { //1 variable(s) and 2 values
    table
        0.9    // p(True | evidence )
        0.1;   // p(False | evidence );
}
probability ( "M" ) { //1 variable(s) and 2 values
    table
        0.1    // p(True | evidence )
        0.9;   // p(False | evidence );
}
probability ( "I" "B" "M" ) { //3 variable(s) and 8 values
    table
        0.9 0.5 0.5 0.1 0.1 0.5 0.5 0.9;
}
```

上述为 example.bif 的**部分内容**，这部分阐述的内容为：

1. 该贝叶斯网络存在三个节点：B、M、I；
2. B、M、I三者均有两种可能的取值：True、False；
3. B、M二者没有父节点但这两个节点均为I的父节点；
4. B、M二者为True及False的概率分别为：0.9/0.1和0.1/0.9（如红色框内所示）；
5. 在以B/M为父节点的情况下，I的概率如图中绿色框所示；

6. I 的概率分为 8 个值，分别代表下面的情况：

$$P(I = \text{true} | B = \text{true}, M = \text{true}) = 0.9$$

$$P(I = \text{true} | B = \text{true}, M = \text{false}) = 0.5$$

$$P(I = \text{true} | B = \text{false}, M = \text{true}) = 0.5$$

$$P(I = \text{true} | B = \text{false}, M = \text{false}) = 0.1$$

$$P(I = \text{false} | B = \text{true}, M = \text{true}) = 0.1$$

$$P(I = \text{false} | B = \text{true}, M = \text{false}) = 0.5$$

$$P(I = \text{false} | B = \text{false}, M = \text{true}) = 0.5$$

$$P(I = \text{false} | B = \text{false}, M = \text{false}) = 0.9$$

简单概括 bif 文件格式的含义：**variable** 部分表示节点的名称及节点可能的取值类型，**probability** 部分则表示相应的概率情况，如：probability ( "B" )就表示  $P(B = \text{true}) = 0.9$  及  $P(B = \text{false}) = 0.1$  而 probability ( "I" "B" "M" ) 则表示  $P(I | B, M)$  的八种情况。

alarm.bif 与 example.bif 拥有相同的格式，其对于节点及概率的表达方式是一致的，但前者相较于后者在结构上更为复杂，节点数量更多，取值范围更大。

本次 Project 的目标是通过对病患样本的学习从而实现贝叶斯网络概率参数的计算，故所提供的 alarm.bif 文件中，probability 部分的所有概率均为“-1”，你们的任务是计算出正确的概率数值并用计算出的正确数值替换文件中对应位置的“-1”。这一任务需要注意两点：1) 所有的“-1”均是占位符，不表示任何概率；2) 需要计算出正确的概率，并替换正确位置上的“-1”。

## 5. records.dat

records.dat 文件为本次 Project 贝叶斯网络概率学习的样本文件，该文件含有超过 1 万条病患的观察记录，其格式如下所示：

```
"True" "Low" "False" "Low" "Low" "Normal" "False" "Normal" "?" "False" "Normal" "False" .....
"True" "Low" "False" "Low" "Low" "Low" "?" "Normal" "Normal" "False" "Normal" "False" .....
"False" "High" "False" "Normal" "Normal" "Normal" "False" "Normal" "Normal" "?" "Normal" .....
"False" "Normal" "False" "Normal" "Normal" "Normal" "False" "Normal" "Normal" "False" .....
.....
```

文件中的每一行代表了一个病人的观察记录，这些值以空格进行分割，分别代表对应节点的取值，此处的取值将按照 alarm.bif 中 variable 部分的顺序进行排列记录，例如：

alarm.bif 中 variable 部分的顺序为：

Hypovolemia、StrokeVolume、LVFailure、LVFDVolume、PCWP、CVP.....

records.dat 中某一行的记录为：

```
"True" "Low" "False" "Low" "Low" "Normal".....
```

分别对应着：

Hypovolemia = "True"

StrokeVolume = "Low"

LVFailure = "False"

LVFDVolume = "Low"

PCWP = "Low"

CVP = "Normal"

.....

正如前文所说，实际应用场景中，病患的观察记录往往存在缺失值，此处缺失值以“?”表示，若某一位置为“?”则表示对应的属性值缺失，但缺失值的情况是被控制在合理范围内的，每一行记录中最多存在一个“?”，即每个病人的观察记录中最多有一个缺失值。

## 6. 提供的代码框架

本次 Project 为同学们提供了一套用 **Java** 编写的代码用于帮助同学们快速开始贝叶斯网络的构建及学习，这部分的代码包括四个部分：

1. DJ.java;
2. InferenceGraph.java;
3. InferenceGraphNode.java;
4. **BayesNets.jar**。

其中，主要的代码为 BayesNets.jar，该代码为用于贝叶斯网络操作的工具包，能够帮助使用者快速的解析 bif 格式的贝叶斯网络文件从而更好的操作其中的节点和概率参数。

DJ.java 为该工具包的一个使用示范，能通过该代码简单的掌握 BayesNets.jar 工具包的使用。

InferenceGraph.java 及 InferenceGraphNode.java 则仅仅做参考所用，这两个文件是 BayesNets.jar 源码中的一部分，通过这两个文件可以了解到 BayesNets.jar 工具包的 API，由于仅做参考之用，所以在 Project 代码编写过程中同学们无需对这两个 java 代码文件进行修改或使用。

如果采用 java 作为本次 Project 的编程语言，则可以直接将 BayesNets.jar 工具包引入自己的工程中进行直接使用。但如果不采用 java 而采用 C++/Python 等其余编程语言，则需要自行编写程序用于贝叶斯网络的解析及操作。

本次 Project 不对编程语言进行限制，同学们可自行选择编程语言进行程序编写，但需要注意的是：请勿直接调用任何现有的工具库或工具包进行贝叶斯网络的概率推理，概率推理部分的代码需要大家自行完成，本次 Project 将对此重点检查，故需要大家务必遵守该要求。

## 7. 相关文件结构

本次 Project 提供给各个同学的将是一个名为 **“AI\_Project-2”** 的压缩包，解压后可得到以下四个文件夹，分别包含了本次 Project 所需的各类文件资料：

- 1) src 源代码文件夹，包含了 DJ.java、InferenceGraph.java、InferenceGraphNode.java；
- 2) lib 工具包文件夹，包含了 BayesNets.jar 工具包；
- 3) doc 文档说明文件夹，包含了本次 Project 的相关说明文件；
- 4) data 数据文件夹，包含了 alarm.bif、example.bif 及 records.dat。

## PROJECT 提交及相关要求

本次 Project 需要提交的文件分为**三部分**：

- 1) **包含计算出概率的 alarm.bif 文件**，将其命名为 alarm\_result.bif，此处需要注意的是保持 alarm.bif 文件的格式，仅对其中“-1”占位的概率用计算出的概率数值进行替换；
- 2) **一份报告**，以小组为单位，要求基本同 Project-1，需要注明每一名组员在本次 Project 完成过程中的工作及贡献，并对本小组在本次 Project 的研究工作进行阐述和分析；
- 3) **本小组对于本次 Project 求解过程中所有相关的工程代码**，在代码中需要简要注明代码的功能及相关使用方法以便于后续 Project 结果的核查。

本次 Project 的 **deadline** 为：**2017 年 11 月 1 日（周三）晚 11:30**  
希望大家按时提交 Project 结果。

所需的提交的所有文件打包为**压缩包**，并以 **“Project2-小组组长姓名”** 的形式进行命名，在 deadline 之前提交至 **TA 邮箱：icrc\_wsc@qq.com**。