

基于贝叶斯网络的主效应因素分析

仝小敏, 李国栋, 刘娜, 吉祥

中国电子科学研究院, 北京
Email: tongxiaomin@cetc.com.cn

收稿日期: 2021年4月8日; 录用日期: 2021年5月3日; 发布日期: 2021年5月12日

摘要

主效应因素分析是故障定位、道路交通事故分析、航空事故分析、风险分析、致因分析等领域关注和研究的重要课题, 对于分析事故原因以及采取有效预防措施、降低事故发生率具有重要意义。现有基于贝叶斯网络的致因分析方法多数基于构建的贝叶斯网络结构对影响事故发生的因素进行单独的灵敏度分析, 忽略因素之间的交互效应, 从而得到片面的结论。本文提出一种基于贝叶斯网络的主效应因素分析方法, 在充分利用因素之间的关联关系进行贝叶斯网络结构构建之后, 分析各因素对结果因素的影响路径并进行联合影响度分析。实验证明, 本文方法克服了简单割裂地对各因素进行单独分析的缺点, 能够得到更加可靠全面的致因分析结论。

关键词

贝叶斯网络, 主效应因素, 影响路径, 联合影响度, 致因分析

Analysis of Main Effect Factors Based on Bayesian Network

Xiaomin Tong, Guodong Li, Na Liu, Xiang Ji

China Academy of Electronics and Information Technology, Beijing
Email: tongxiaomin@cetc.com.cn

Received: Apr. 8th, 2021; accepted: May 3rd, 2021; published: May 12th, 2021

Abstract

Main effect factor analysis is an important topic in fault location, road traffic accident analysis, aviation accident analysis, risk analysis, cause analysis and other fields. It is of great significance to analyze the causes of accidents, take effective preventive measures and reduce the incidence of

accidents. Most of the existing causal analysis methods based on Bayesian network are based on the constructed Bayesian network structure to conduct a separate sensitivity analysis of the factors that affect the occurrence of the accident, ignoring the interaction effect between the factors, so as to get a one-sided conclusion. In this paper, a main effect factor analysis method based on Bayesian network is proposed. After making full use of the correlation between factors to construct the Bayesian network structure, the influence path of each factor on the result factor is analyzed and the joint influence degree is analyzed. Experiments show that this method overcomes the shortcomings of simple split analysis of each factor, and can get more reliable and comprehensive causal analysis conclusions.

Keywords

Bayesian Network, Main Effect Factors, Impact Path, Joint Influence Degree, Cause Analysis

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

主效应因素分析广泛应用于故障诊断、事故分析、风险分析等领域[1][2],能够为事故原因定位、致因分析以及制定有效预防措施提供定性定量分析依据。贝叶斯网络作为一种结合数据挖掘和不确定性知识推理的人工智能技术,已经广泛应用于各个领域,其优势之处主要体现在能够通过有向无环图表达出节点之间的相互关系,通过节点之间的条件概率进行不确定性知识推理[3]。基于贝叶斯网络的致因分析方法,通常包括因素提取、贝叶斯网络结构构建、贝叶斯网络参数计算、直接影响因素定位、灵敏度分析与排序等步骤,最后将灵敏度最高的直接影响因素作为导致事故发生的主效应因素。

现有基于贝叶斯网络的主效应因素分析方法多数基于构建的贝叶斯网络结构对影响事故发生的因素进行单独的灵敏度分析,从而实现致因定位或者故障定位。这种方法虽然在贝叶斯网络结构构建时考虑了因素之间的关联关系,在因素灵敏度分析时却忽略了因素之间的交互效应,有可能导致片面的分析结论,从而影响致因分析的全面性和可信度。

如果一个因素不仅直接对结果节点产生影响,同时通过影响另一个因素进而对结果节点产生影响,那么我们认为这两个因素之间是存在交互效应的,在因素对结果影响分析的时候,除了要考虑因素本身对结果节点的直接影响,必须分析因素对交互效应的节点的影响,进而形成全面的分析结论。基于此,本文提出一种基于贝叶斯网络的主效应因素分析方法,在充分利用因素之间的关联关系进行贝叶斯网络结构构建之后,分析各因素的影响路径并进行灵敏度分析和致因分析。这种方法不是简单割裂地对各因素进行单独分析,而是充分考虑因素之间的交互效应,从而得到更加可靠全面的致因分析结论。

2. 相关工作

随着国内外学者对主效应分析理论的不断研究和发展,学术界提出了多种分析模型,包括因果连锁模型、轨迹交叉模型、瑟利模型、突变模型等[4]。主效应因素分析方法的发展趋势是多种因素协同分析,单一因素分析方法不能满足现阶段复杂过程的诊断分析需求。隶属于机器学习算法的贝叶斯网络作为一种可以融合多源信息的不确定性知识表达与推理模型,在数据学习的基础上能与专家知识等多种方法进行有效结合,协同诊断。其优势体现在[5]: 1) 基于图模型,能贴切并直观地描述变量间的因果关系和条

件相关性; 2) 具备基于数学概率理论的不确定性推理能力, 能在有限的、不完整的信息条件下进行学习、推理和更新; 3) 能有效融合和表达多源信息。因此, 基于贝叶斯网络的主效应分析方法适合于表达复杂过程中错综复杂的因果关系。其在信息的表达能力、不确定条件下的学习能力、与其他方法的结合能力、模型更新能力、多源信息融合处理能力以及解释性等方面的综合性能比神经网络、故障树等其他智能算法更适用于主效应因素分析问题[6] [7] [8]。

文献[9]将贝叶斯网络作为一种事故分析手段, 提出了一种基于危险因素 - 事故 - 事故危害的三层贝叶斯网络拓扑模型, 更加全面地分析事故的致因因素以及其可能导致的各种后果, 同时实现对事故类型和事故后果的预测。文献[10]基于贝叶斯网络模型, 综合运用诊断推理和支持推理形式, 分析致灾因素的因果关系, 揭示了人、机、环境与管理因素相互作用的内在规律, 对事故致因因素进行了更深入的挖掘, 有助于及时发现系统中潜在的事故隐患。文献[11]提出了一种基于贝叶斯网络模型理论的道路交通事故预测方法。在综合考虑交通事故成因的基础上利用领域专家知识构建网络模型, 在已有的事故数据的基础上提出基于贝叶斯法则的学习算法, 并通过计算变量间的条件概率来计算事故发生的可能性, 达到事故预测的目的。文献[12]通过建立基于事故样本数据的贝叶斯网络模型, 将贝叶斯网络应用于道路交通事故原因推理中。该模型特点是综合了驾驶员、道路状况、天气等多方面因素, 但同时也存在一定的不足, 即对这些因素难以进行有效的量化。文献[13]利用贝叶斯网络对铁路事故进行致因分析, 在融合领域知识和样本数据的基础上建立贝叶斯网络。文献[14]结合人为组织因素应用贝叶斯网络进行风险分析, 利用贝叶斯网络的推理能力, 对海上运输系统内的人为组织因素, 包括船公司、港口管理、规章制度等, 进行了风险分析。文献[15]基于前人所建的网络结构, 进一步确定了船舶碰撞的致因因素, 并确定出贝叶斯网络结构, 利用软件对船舶碰撞事故的发生概率进行了预测分析。

上述文献表明, 贝叶斯网络可以从定量的角度进行主效应因素分析, 其自身具有的推理功能可以成为主效应因素分析的有效工具。然而上述方法的缺陷在于缺乏对因素间交互效应的考虑, 影响了分析结论的客观可行程度。本文算法的核心思想在于交互效应分析。在构建贝叶斯网络模型后, 首先对利用网络结构连接关系, 定位对结果节点产生直接影响的因素作为直接影响因素, 然后考虑到这些因素之间可能存在着交互效应, 分析每个直接影响因素对结果节点的影响路径, 形成因素对结果的影响节点集合。再从影响节点集合中筛选受直接影响因素影响较大的节点作为交互效应节点, 最后利用直接影响因素和对应的交互效应节点联合分析其对结果节点的影响程度, 作为直接影响因素的影响度。影响度最大的直接影响因素就是所要寻找的主效应因素, 对待分析样本的发生起到至关重要的作用。

3. 主效应因素分析流程

本文算法流程如图 1 所示, 主要包括数据标准化、建立贝叶斯网络模型、直接影响因素定位、影响路径分析、独立影响度分析、联合影响度分析等六个关键步骤。其中灰色框中的步骤是训练分析阶段进行的, 输入为训练数据, 输出为影响路径分析结果。橙色框中的步骤是主效应因素分析阶段进行的, 输入为待分析数据, 输出为主效应因素分析结果。详述如下:

3.1. 训练数据标准化

训练数据标准化主要是将原始数据转换为用于主效应因素分析的离散化标准数据。记用于贝叶斯网络构建训练的原始数据集为 $D = \{X_1, X_2, \dots, X_N\}$, N 表示训练样本数量, 第 i 个样本 $X_i = \{x_{i1}, x_{i2}, \dots, x_{iM}, y_i\}$, x_{ij} 表示 X_i 的第 j 个特征, M 表示训练样本特征的维数, y_i 表示第 i 个样本对应的事故发生情况, 若发生为 1, 未发生为 0。对于离散型特征和连续型特征采取不同的标准化方法。

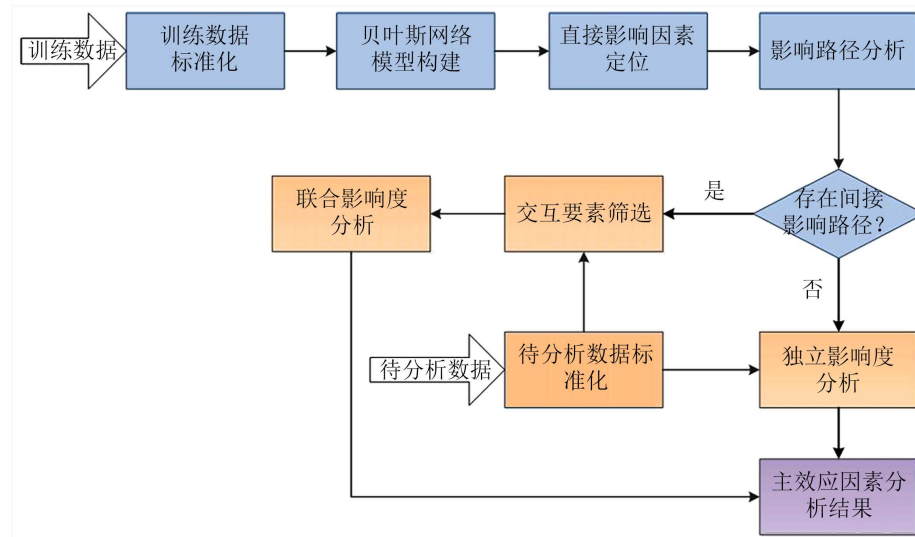


Figure 1. The main effect factor analysis process proposed in this paper

图 1. 本文提出的主效应因素分析方法流程

假设第 j 个特征为离散型特征，即该特征取值个数为确定的有限正整数 K ，为了简化计算，需将所有样本中该特征的取值规范化到 $\{0, 1, 2, \dots, K-1\}$ 这个集合范围内。首先对该特征的 K 个不同取值进行从小到大排序处理，得到排序后的指标取值集合。规范化后的特征 $x'_{ij} = k$ ， k 表示 x_{ij} 在排序后的指标取值集合中的序号。

假设第 j 个特征为连续型特征，则需要离散化处理。采用二值化离散方法，首先求该特征的均值为：

$$E_j = \frac{\sum_i x_{ij}}{N} \quad (1)$$

离散化后的特征为：

$$x'_{ij} = \begin{cases} 0 & x_{ij} < E_j \\ 1 & x_{ij} \geq E_j \end{cases} \quad (2)$$

至此，完成了训练数据标准化。

3.2. 建立贝叶斯网络模型

标准化后的训练数据集记为 $D' = \{X'_1, X'_2, \dots, X'_N\}$ 。基于标准化后的数据集 D' ，采用 BIC 评分 + 爬山法构建贝叶斯网络结构。在贝叶斯网络中，每个节点对应于样本的一维特征，所有样本特征构成一个包含 $M+1$ 个节点的网络结构。所有可能的网络结构构成了网络结构空间 G ，贝叶斯网络结构空间中的点 g 是一个 $(M+1) \times (M+1)$ 的二值矩阵， $g(i, j)$ 为 1 表示存在一条节点 i 指向节点 j 的有向边。BIC 评分的计算公式如下：

$$\text{BIC}(g | D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \lg \frac{m_{ijk}}{m_{ij}} - \sum_{i=1}^n \frac{q_i(r_i-1)}{2} \lg m \quad (3)$$

其中 g 表示网络结构， D 表示训练数据集， $\text{BIC}(g | D)$ 表示网络结构 g 的 BIC 评分， Node_i 表示第 i 个节点， r_i 表示 Node_i 可能的取值个数， q_i 表示 Node_i 的父节点取值的组合数。 m_{ijk} 表示 X_i 取第 k 个值， Node_i

的父节点取值为第 j 个组合的样本个数, m_j 表示 $Node_i$ 的父节点取值为第 j 个组合的样本个数。

建立贝叶斯网络, 首先随机生成网络空间 G 中的一个点 g_0 作为爬山搜索的起点, 并计算 g_0 的 BIC 评分 $BIC(g_0 | D')$ 。接着对 g_0 进行随机局部操作, 随机局部操作类型包括加边、减边、转边三种, 得到随机操作后的网络结构 g'_0 , 并计算 g'_0 的 BIC 评分 $BIC(g'_0 | D')$ 。如果 $BIC(g'_0 | D') > BIC(g_0 | D')$, 则认为 g'_0 比 g_0 更优, 令 $g_0 = g'_0$ 。重复进行局部随机操作与评分比较, 直到达到迭代次数。

至此, 确定了贝叶斯网络结构中节点之间的连接关系, 用矩阵 g_r 表示。接着利用数据集 D' 统计计算网络中各节点的条件概率表。例如计算节点 i 的条件概率, 首先要确定节点 i 的父节点结合 π_i , 然后统计 π_i 不同取值组合情况下节点 i 取不同值的次数, 计算 $p(i | \pi_i)$ 的值作为该节点的条件概率。

3.3. 直接影响因素定位

假设节点 $M+1$ 为表示事故发生情况的结果节点, 那么根据贝叶斯网络结构 g_r 可以确定结果节点的父节点集合 π_{M+1} 。在 π_{M+1} 中有一部分节点与结果节点之间存在直接的因果关系, 这部分节点的序号构成集合记为 π_s , 满足 $g_r(i_s, M+1) = 1, i_s \in \pi_s$, π_s 就是我们要确定的直接影响因素集合。 π_{M+1} 中还有一部分节点通过其他节点对结果节点产生影响, 成为间接影响因素节点。例如图 2 的贝叶斯网络结构模型中, 因素 1、因素 2、因素 3 都是与结果节点有直接连接关系的节点, 因此是结果的影响因素。

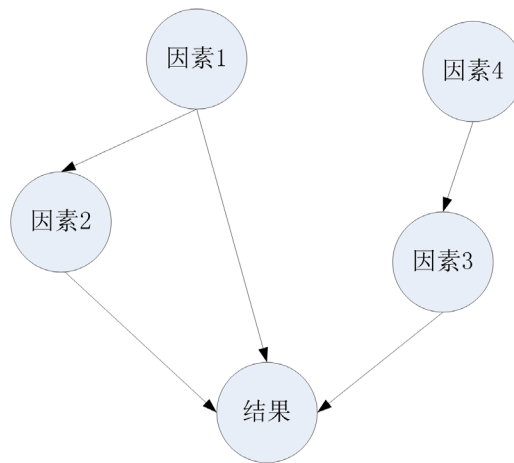


Figure 2. Examples of direct influencing factors positioning

图 2. 直接影响因素定位示例

3.4. 影响路径分析

直接影响因素集合 π_s 中的每个特征对应的特征节点与结果节点之间必然存在直接连接, 这种直接连接对应于因素对结果的影响路径。除此之外, 直接影响因素还有可能通过其他的因素对结果产生影响, 例如节点 i_s 可能通过若干个间接影响因素和另外一个直接影响因素 i'_s 对结果产生间接影响, 这些间接影响因素和另一个直接影响因素就构成了节点 i_s 到结果节点的间接影响路径。遍历矩阵 g_r , 寻找每个直接影响因素节点对结果节点的影响路径集合。记节点 i_s 的影响路径集合为 $R_{i_s} = \{r_{0i_s}, r_{1i_s}, \dots, r_{ni_s}\}$, 其中 $r_{0i_s} = \{i_s, M+1\}$ 表示节点 i_s 到结果节点的影响路径, $n \geq 0$ 表示间接影响路径的总数, $r_{1i_s}, \dots, r_{ni_s}$ 表示节点 i_s 到结果节点的 n 条间接影响路径, 假定第 j 条间接影响路径为 $r_{ji_s} = \{i_s, i_{j1}, i_{j2}, \dots, i_{jl}, M+1\}$, 包含 $l+2$ 个节点, 那么 i_{jl} 必然对结果节点有直接影响, 贝叶斯网络结构中有一条从节点 i_{jl} 到结果节点的影响路径。如果 R_{i_s} 中只包含一条直接影响路径 r_{0i_s} , 那么直接进行独立影响度分析, 得到诸因素分析结论。如

果存在间接影响路径, 那么进行联合影响度分析, 得到主效应因素分析结果。

例如图 2 的贝叶斯网络结构模型中, 三个直接影响因素中, 因素 2 对结果节点的影响路径只有一条直接影响路径, 只需要进行独立影响度分析即可, 而因素 1 对结果节点的影响路径除了直接影响路径之外, 还有“因素 1- > 因素 2- > 结果”这条间接影响路径。这就是说因素 1 对结果节点的影响还可能通过影响因素 2 而间接影响结果, 后续主效应因素分析需要考虑与因素 2 的交互效应, 进行联合影响度分析。

3.5. 待分析数据标准化

取待分析的数据样本按照 3.1 训练数据标准化的方法进行标准化, 标准化后的样本记为 X 。主效应因素分析主要是分析对样本 X 影响最大的特征因素。

3.6. 独立影响度分析

独立影响度分析即是分析直接影响因素集合 π_s 中每个因素对结果的影响程度。针对 π_s 中的任意一个因素 i_s , 构造样本 X' , 满足除结果因素 $M+1$ 以外的所有特征取值与 X 一致, 将 X' 输入贝叶斯网络模型, 计算结果因素取值与 X 一致的概率记为 p_{M+1} 。随机改变 X' 中因素 i_s 的取值, 重新输入贝叶斯网络模型, 计算结果因素取值与 X 一致的概率记为 p'_{M+1} 。那么, 因素 i_s 对结果节点的影响度为 $\Delta p_{i_s} = p_{M+1} - p'_{M+1}$ 。对 π_s 中所有因素分别计算其对结果节点的影响度, 并选择影响度最大的因素作为主效应因素分析结果。

3.7. 联合影响度分析

影响路径集合 $R_{i_s} = \{r_{0i_s}, r_{1i_s}, \dots, r_{ni_s}\}$ 中所有节点构成了一个影响路径的节点集合 $H_{i_s} = \{i_s, i_{s1}, i_{s2}, \dots, i_{sl}, M+1\}$, 包括直接影响因素节点 i_s 、间接影响节点 $i_{s1}, i_{s2}, \dots, i_{sl}$ 、结果节点 $M+1$, l 表示节点 i_s 和结果节点 $M+1$ 之间的间接影响节点的个数。构造样本 X' , 满足 R_{i_s} 集合以外的所有特征取值与 X 一致, 其余特征取值不定。

3.7.1. 交互因素筛选

首先通过计算节点 i_s 对节点 $i_{s1}, i_{s2}, \dots, i_{sl}$ 的影响, 筛选与节点 i_s 有交互效应的节点。设定 X' 的特征 i_s 取值为 X 中的特征 i_s 的值, 输入构建好的贝叶斯网络结构中, 计算节点 $i_{s1}, i_{s2}, \dots, i_{sl}$ 取值为 X 中特征 $i_{s1}, i_{s2}, \dots, i_{sl}$ 的值的概率, 记为 $p_{s1}, p_{s2}, \dots, p_{sl}$ 。随机改变 X' 的特征 i_s 的取值, 计算节点 $i_{s1}, i_{s2}, \dots, i_{sl}$ 取值为 X 中特征 $i_{s1}, i_{s2}, \dots, i_{sl}$ 的值的概率, 记为 $p'_{s1}, p'_{s2}, \dots, p'_{sl}$ 。若 $|p'_{si} - p_{si}| < \theta$, 则认为节点特征 i_s 的改变不会影响节点 i_{si} 的取值, 反之认为节点特征 i_s 的改变会影响节点 i_{si} 的取值, 因而会对结果节点的取值产生交互效应。 $\theta \in [0, 1]$ 为交互效应判定阈值, θ 取值越大, 交互因素筛选要求越高, 则要求因素之间具有很强的交互相关性, 才作为交互效应因素, 进行联合影响度分析, 极端情况下取值为 1, 则趋同于独立影响度分析。 θ 取值越小, 交互因素筛选要求越低, 这时候考虑交互因素的比较全面, 但是也会形成冗余的结论。因此本文实验中取值 0.3, 兼顾独立性和交互性要求。

3.7.2. 联合影响度分析

选择 $i_{s1}, i_{s2}, \dots, i_{sl}$ 中与 i_s 有没有交互效应的节点, 将 X' 中相应节点特征取值为 X 中相应的特征值, 当前 X' 中取值待定的特征包括节点 i_s 、与节点 i_s 有交互效应的节点、结果节点 $M+1$ 。设定 X' 的特征 i_s 取值为 X 中的特征 i_s 的值, 输入构建好的贝叶斯网络结构中, 计算结果节点 $M+1$ 取值为 X 中结果节点取值的概率记为 p_{M+1} 。随机改变 X' 的特征 i_s 的取值, 重新输入构建好的贝叶斯网络结构中, 计算结果节点 $M+1$ 取值为 X 中结果节点取值的概率记为 p'_{M+1} 。记特征 i_s 对结果节点的影响度为 $\Delta p_{i_s} = p_{M+1} - p'_{M+1}$ 。

对直接影响因素集合 π_s 中的每个因素计算其对结果节点的影响度, 并选择影响度最大的节点作为影响待分析样本的主效应因素。

4. 实验验证

4.1. 数据描述

威斯康辛州乳腺癌(BreastCancer)数据集是 sklearn.datasets 的内置数据集, 包含了威斯康辛州记录的 569 个病人的乳腺癌恶性/良性(1/0)类别型数据(训练目标), 以及与之对应的 30 个维度的生理指标数据。

4.2. 实验流程

在 window10 系统上使用 Python3.7 实现本文所提出的算法, 并使用 BreastCancer 数据集对本文算法进行实验验证。利用原始数据集中包含的[569 × 30]维生理指标数据和对应的 569 个类别型数据构成 569 个 31 维的训练样本数据。其中除类别型数据外, 其他生理指标数据均为连续型特征, 需要利用公式(1)计算各自特征离散化阈值并安装公式(2)进行特征离散化, 得到标准化后的训练样本数据。之后利用爬山算法构建贝叶斯网络结构模型, 该模型中类别型数据对应的节点作为结果节点, 定位直接影响因素、分析影响路径。输入待测样本进行影响度分析。

4.3. 实验结果

构建的贝叶斯网络模型如图 3 所示。

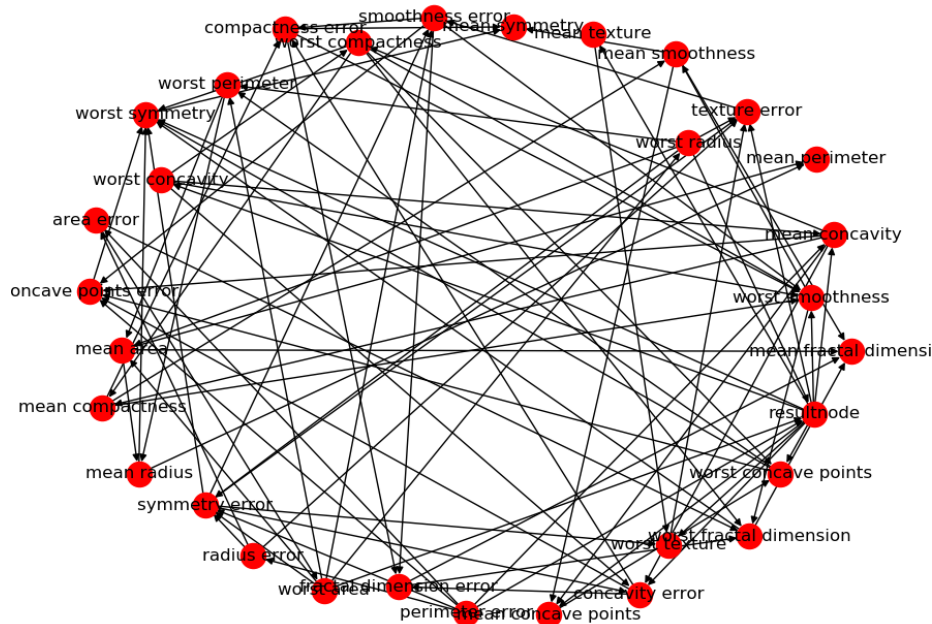


Figure 3. Bayesian network model
图 3. 贝叶斯网络模型

在该模型中结果节点“resultnode”的直接影响要素包括“worst area”、“worst radius”、“perimeter error”, “worst area”的影响路径包括直接影响路径“worst area”→“resultnode”和间接影响路径“worst area”→“worst radius”→“resultnode”, “worst radius”的影响路径只有直接影响路径“worst radius”→“resultnode”, “perimeter error”的影响路径包括直接影响路径“perimeter error”→“resultnode”和

4 条间接影响路径, 分别是 “perimeter error” → “radius error” → “area error” → “worst area” → “worst radius” → “resultnode”, “perimeter error” → “radius error” → “area error” → “worst area” → “resultnode”, “perimeter error” → “area error” → “worst area” → “worst radius” → “resultnode”, “perimeter error” → “area error” → “worst area” → “resultnode”。

取第二个样本作为待测试样本。该样本结果节点取值为 0, 表示乳腺癌良性, 计算结果表示 “worst area” 节点存在交互效应节点 “worst radius”, 说明 “worst area” 在对 “resultnode” 产生直接影响的同时, 通过影响 “worst radius” 对 “resultnode” 产生间接影响, 考虑因素之间的交互效应, 首先预测当前样本结果节点取值为 0 的概率, 然后改变样本中 “worst area” 节点的值, 同时将交互效应节点 “worst radius” 的值不设定, 预测结果节点取值为 0 的概率。将这两个概率做差, 得到 “worst area” 对结果节点的影响度为 0.9935。“worst radius” 节点不存在交互效应节点, 只对 “resultnode” 产生直接影响, 首先预测当前样本结果节点取值为 0 的概率, 然后改变 “worst radius” 节点取值, 预测结果节点取值为 0 的概率。两个概率做差, 计算 “worst radius” 节点对结果节点的影响度为 0.9935。“perimeter error” 节点存在交互效应节点 “radius error”、“area error”, 与 “worst area” 影响度的计算类似, 计算结果为 “perimeter error” 节点影响度 0.0347。可见 “worst area” 和 “worst radius” 是影响第二个样本 “resultnode” 判定为良性的主效应因素。

取第 20 个样本作为待测样本。该样本结果节点取值为 1, 表示结果为乳腺癌恶性。计算结果表示 “worst area”、“worst radius”、“perimeter error” 节点都不存在交互效应节点, 只对 “resultnode” 产生直接影响, 只需计算当前考虑节点取值变化前后对预测概率的影响即可, 计算结果中 “worst area”、“worst radius”、“perimeter error” 节点影响度分别为 0.035、0、0.001, 因此判定 “worst area” 为影响第 20 个样本结果判定为恶性的主效应因素。

4.4. 对比分析

实验结果中, 对于第二个样本, 如果不考虑因素之间的交互效应, “worst area” 对结果节点的影响度只有 0.2085, 得到的结论是: “worst radius” 是影响第二个样本 “resultnode” 判定为良性的主效应因素。这与考虑交互效应得到的分析结论是不一致的。这是因为 “worst area” 取值变化直接影响了 “worst radius” 的取值, 如果忽略了 “worst area” 和 “worst radius” 之间的关联关系, 就可能导致片面的结论。

5. 结论

本文提出的主效应因素分析方法在利用贝叶斯网络模型进行主效应因素分析的过程中, 利用网络结构模型筛选存在交互效应的因素, 同时考虑因素之间的交互效应以及这些因素共同对结果产生的影响, 从而得到主效应分析结论。实验证明, 与现有方法相比, 本文方法克服了单因素分析方法将因素孤立化的问题, 在充分分析因素之间的联动交互影响的基础上提取主效应因素, 提高了主效应因素分析结论的合理可信性。

致 谢

对本文的创作给予指导和帮助的同事朋友, 在此一并感谢。

基金项目

本文受到装备预研项目 “云化服务 XX 验证技术” (NO: 315105307)、国防科技项目 “XX 跨域体系集成与试验验证技术” (NO: 2019-JCJQ-ZD-193-00) 等资助。

参考文献

- [1] Long, Y.G., Wang, L.M., Duan, Z.Y. and Sun, M.H. (2019) Robust Structure Learning of Bayesian Network by Identifying Significant Dependencies. *IEEE Access*, **7**, 116661-116675. <https://doi.org/10.1109/ACCESS.2019.2936399>
- [2] 王辉, 张娟, 赵雅, 刘琨, 冯文峰. 一种新型贝叶斯模型的网络风险评估方法[J]. 小型微型计算机系统, 2020, 41(9): 1898-1904.
- [3] 陈景文, 褚恩亮, 李英春, 杨亚莉. 基于小波包-贝叶斯的小电流接地故障区域定位方法[J]. 实验技术与管理, 2019, 36(8): 148-152.
- [4] Heinrich, H.W. (2011) Industrial Accident Prevention: A Scientific Approach. *Industrial & Labor Relations Review*, **4**, 609.
- [5] 李俭川. 贝叶斯网络故障诊断与维修决策方法及应用研究[D]: [博士学位论文]. 长沙: 国防科学技术大学, 2002.
- [6] 王波. 煤矿安全生产事故析因分析研究[D]: [硕士学位论文]. 徐州: 中国矿业大学, 2015.
- [7] 张东, 聂百胜, 王龙康. 我国煤矿安全生产事故的致灾因素分析[J]. 中国安全生产科学技术, 2013, 9(5): 136-140.
- [8] Foster, P. (2013) The Safety Journey: Using a Safety Maturity Model for Safety Planning and Assurance in the UK Coal Mining Industry. *Minerals*, **3**, 59-72. <https://doi.org/10.3390/min3010059>
- [9] 俞娉婷, 刘振元, 陈学广. 基于贝叶斯网络的一种事故分析模型[J]. 中国安全生产科学技术, 2006, 2(4): 45-50.
- [10] 罗帆, 陈小佳, 顾必冲. 基于贝叶斯网络的航空灾害成因机理探析[J]. 北京航空航天大学学报, 2005, 31(8): 934-938.
- [11] 秦小虎, 刘利, 张颖. 一种基于贝叶斯网络模型的交通事故预测方法[J]. 计算机仿真, 2005, 22(11): 230-232.
- [12] 王发智. 基于贝叶斯网络的交通突发事件态势评估技术[D]: [硕士学位论文]. 大连: 大连理工大学, 2006.
- [13] Marsh, W. and Bearfield, G. (2004) Using Bayesian Networks to Model Accident Causation in the UK Railway Industry. In: Spitzer, C., Schmocker, U. and Dang, V.N., Eds., *Probabilistic Safety Assessment and Management*, Springer, London, 3597-3602. https://doi.org/10.1007/978-0-85729-410-4_575
- [14] Trucco, P., Cagno, E., Ruggeri, F. and Grade, O. (2008) A Bayesian Belief Network Modeling of Organizational Factors in Risk Analysis: A Case Study in Maritime Transportation. *Engineering and System Safety*, **93**, 845-856. <https://doi.org/10.1016/j.res.2007.03.035>
- [15] Friis-Hansen, P. and Simonsen, B.C. (2002) GRACAT: Software for Grounding and Collision Risk Analysis. *Marine Structures*, **15**, 383-401. [https://doi.org/10.1016/S0951-8339\(02\)00009-6](https://doi.org/10.1016/S0951-8339(02)00009-6)