

引用: 郭秋麟, 任洪佳, 于京都, 等. 基于贝叶斯网络的油气勘探风险预测方法: 以准噶尔盆地腹部侏罗系三工河组为例[J]. 中国石油勘探, 2023,28(1):108-119.
Guo Qiulin, Ren Hongjia, Yu Jingdu, et al. Prediction method of petroleum exploration risks based on Bayesian network: a case study of the Jurassic Sangonghe Formation in the hinterland of Junggar Basin[J]. China Petroleum Exploration, 2023,28(1):108-119.

基于贝叶斯网络的油气勘探风险预测方法 ——以准噶尔盆地腹部侏罗系三工河组为例

郭秋麟¹ 任洪佳² 于京都¹ 刘继丰³ 陈宁生¹

(1 中国石油勘探开发研究院; 2 燕山大学信息科学与工程学院; 3 北京天腾网格技术开发有限公司)

摘要: 有效预测油气勘探风险对优化油气勘探部署、提高钻探成功率及经济效益具有至关重要的意义。在分析油气勘探风险预测方法发展状况的基础上, 提出基于贝叶斯网络的油气勘探风险预测方法, 论述了石油地质问题向概率预测模型的转化过程, 构建了 AODE 模型的算法和预测步骤。以准噶尔盆地腹部侏罗系三工河组为例, 开展了油气成藏地质条件的定量评价, 确定了供烃、储层、圈闭、盖层与保存 4 项主控地质参数, 建立了由 203 口探井参数组成的数据集。五折交叉验证结果揭示: (1) 4 种贝叶斯网络 (朴素贝叶斯分类器、树增强贝叶斯分类器、平均一阶依赖估计器和 k 阶依赖贝叶斯分类器) 的训练集判别正确率均大于 85%, 说明训练集参数分类有效; (2) 测试集判别正确率均大于 82%, 说明预测成功率高。平均一阶依赖估计器模型的预测效果最好, 准确率达到 85.22%, 因此采用该模型预测研究区油气勘探风险。结果揭示: 平均一阶依赖估计器模型预测结果不仅在储量区内与勘探结果吻合度较高, 而且在储量区外预测了 3 类油气资源分布有利区。

关键词: 油气勘探风险; 风险预测; 贝叶斯网络; 三工河组; 准噶尔盆地

中图分类号: TE19 **文献标识码:** A **DOI:** 10.3969/j.issn.1672-7703.2023.01.010

Prediction method of petroleum exploration risks based on Bayesian network: a case study of the Jurassic Sangonghe Formation in the hinterland of Junggar Basin

Guo Qiulin¹, Ren Hongjia², Yu Jingdu¹, Liu Jifeng³, Chen Ningsheng¹

(1 PetroChina Research Institute of Petroleum Exploration & Development; 2 School of Information Science and Engineering, Yanshan University; 3 Beijing Tinten Grid Technology Co, Ltd.)

Abstract: Effective prediction of petroleum exploration risks is of great significance for optimizing exploration deployment and improving drilling success rate and economic benefits. Based on the analysis of the progress of risk prediction technology, a new prediction method of petroleum exploration risks is proposed based on Bayesian network. The transformation process from petroleum geological problems to probability prediction model is discussed, and the algorithm and prediction steps of average one-dependence estimators (AODE) training model

基金项目: 中国石油天然气股份有限公司科学研究与技术开发项目“剩余油气资源空间分布技术研究”(2021DJ07), “页岩油勘探开发技术研究”(2021DJ18); 中国石油天然气股份有限公司重大科技专项“陆相中高成熟度页岩油勘探开发关键技术研究与应用”(2019E-2601)。

第一作者简介: 郭秋麟(1963-), 男, 福建漳州人, 博士, 2008年毕业于中国科学院地质与地球物理研究所, 教授级高级工程师, 现主要从事油气资源评价及油气勘探方面的研究工作。地址: 北京市海淀区学院路20号中国石油勘探开发研究院, 邮政编码: 100083。E-mail: qlguo@petrochina.com.cn

通信作者简介: 任洪佳(1991-), 男, 黑龙江大庆人, 博士, 2022年毕业于吉林大学, 现主要从事机器学习、贝叶斯网络及人工智能应用于石油地质等方面的研究工作。地址: 河北省秦皇岛市河北大街西段438号, 邮政编码: 066000。E-mail: rhongjia@163.com

收稿日期: 2022-01-29; 修改日期: 2022-12-07

are constructed. The quantitative evaluation of geological conditions of hydrocarbon accumulation in the Jurassic Sangonghe Formation in the hinterland of Junggar Basin is conducted, four main geological parameters of hydrocarbon supply, reservoir, trap, cap rock and preservation are determined, and a data set composed of parameters from 203 exploration wells is established. The results of five-fold cross validation show that: (1) The discrimination accuracy of the training set of four Bayesian networks (Naive Bayesian classifier, tree-augmented Bayesian classifier, AODE and k-dependence Bayesian classifier) is greater than 85%, indicating that the classification of parameters in the training set is effective; (2) The discrimination accuracy of the test set is greater than 82%, indicating a high success rate of prediction. AODE has the best prediction results, with an accuracy of 85.22%, therefore, it is used to predict the risk of petroleum exploration in the study area. The prediction results of AODE model not only agreed well with the exploration results within the reserve area, but also supported to predict favorable areas of three types of oil and gas resources outside the reserve area.

Key words: petroleum exploration risk, risk prediction, Bayesian network, Sangonghe Formation, Junggar Basin

0 引言

油气勘探风险分析实质上是对油气勘探成功率的预测, 预测结果不仅可以为油气风险勘探目标优选提供科学依据, 而且对提高油气勘探成效具有重要的实践意义和技术价值。因此近几十年来, 许多国家的大型石油公司以及不同学者逐渐重视油气勘探风险预测问题, 并围绕该问题开展广泛研究。2010年之前, 油气勘探风险预测方法主要分为3类: 基于不同地质条件的叠图法、信息融合法和基于随机过程及多元统计学的数据分析法。叠图法相对研究较早, 其核心是将勘探要素(如烃源岩、储层、保存等)分为不同的图层进行叠加, 从而凸显勘探风险信息, 该类方法技术的优点是工作流程简单易懂且计算难度低^[1-2]。信息融合法改进了叠图法的缺点, 基于多元统计理论, 对多元地质数据进行处理, 通过设计函数、分配权重、标准化等技术手段, 实现对多元地质数据的降维, 并进一步结合地理信息系统制图工具, 实现地质风险的空间可视化^[3]。数据分析法着重体现对油气勘探风险概率在空间分布上的预测, 其特点是利用已有探井、油气地质条件、储量区块分布等资料, 从分析预测的角度表征油气资源在勘探工作中的空间分布特征; 该典型方法包括马氏距离判别法^[4]、谱分解法^[5]、序贯高斯随机模拟法^[6]等。上述3类方法虽然取得了一定的效果, 但是在应用过程中仍存在一些不足:

(1) 叠图法存在结果图中内容过于繁冗, 且不同地质要素间主次关系不明确等缺点。(2) 叠图法对专家经验依赖程度较强, 其评价过程多以定性、半定性为主, 且在分析过程中对探井约束的重视程度有所不足, 因此评价结果主观程度较高, 不同的专家对同一地区的预测结果可能存在较大的偏差。(3) 信息融合法和数据分析法, 在数据量较大时, 存在数据挖掘不够, 以及无法充分而有效地利用数据等缺点。2010年以来,

为克服上述方法的不足, 根据量化地质风险的不确定性, 油气勘探风险评价中引入了三维马尔可夫链模型^[7]、地质模型的随机模拟^[8]和贝叶斯概率^[9-10]分析方法。综合应用经验, 将地质、钻井、工程等不同领域的信息进行融合, 量化表征地质风险, 实现了地质风险概率预测模型的构建和勘探成功率预测^[11-15]。此外, 贝叶斯网络作为不确定知识表示与推理的有效工具, 近年来也被用来解决油气勘探问题, 例如, 基于贝叶斯网络编码油气系统在烃源岩层、储层和圈闭层面上的相关性, 解决勘探远景分析问题^[16-18]; 基于贝叶斯网络整合地质因素预测油气空间分布^[19-20]。然而, 上述文献中的贝叶斯网络方法只适用于某个特定的研究区, 其通用性以及各个方法之间的优缺点和适用范围还有待进一步研究。

本文将贝叶斯网络技术引入到油气勘探风险预测中, 分别构建了朴素贝叶斯分类器、树增强贝叶斯分类器、平均一阶依赖估计器和 k 阶依赖贝叶斯分类器4种油气勘探风险预测模型, 总结了各种模型的优缺点及适用范围, 并以准噶尔盆地腹部侏罗系三工河组203口探井及相关地质资料为数据基础, 对各模型开展五折交叉验证。通过模型对比, 优选出最优模型并对研究区油气勘探风险进行预测, 预测结果可为该区未来油气勘探目标优选及探井部署提供定量的决策依据。

1 预测方法

1.1 石油地质问题转化概率预测模型

油气勘探风险预测, 即油气资源空间分布预测, 是根据地球科学资料、探井资料、盆地模拟资料并结合专家知识来定量预测潜在目的层的含油气空间变化规律。换句话说, 就是在钻井之前有效地预测地下某个位置含有油气资源的概率。该地质问题可以转化为在多维地质因素空间中求解具有不确定性的二分类问

题。假设某含油气区带中已经钻探 m 口探井，探井结果分为油气井和非油气井两类，相对应的区域为生产区和非生产区。实际勘探结果表明：(1) 决定地下特定位置是否含有油气是多个地质因素共同作用的结果；(2) 同一地质因素在生产区和非生产区具有不同的特征。因此分类的基本假设是，同一地质因素在不同组(类)间有显著差异，但在同一组(类)中没有显著差异。通过研究探井结果与影响油气发生的地质因素(假设有 n 个)之间的关系，可知地质因素 $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ 与探井结果 C 有如下映射关系：

$$f: \mathbf{X} \rightarrow C \quad (1)$$

其中， $C=0$ 和 $C=1$ 分别表示探井结果为非油气井和油气井。上述映射关系表明，多维地质因素 \mathbf{X} 与探井结果 C 之间存在函数关系，即如果确定了 \mathbf{X} 和 C 之间的函数关系，在钻井之前则可以利用未开发地区的多维地质信息计算地下含油气的概率，从而减少勘探风险及优化钻井策略。通过前人研究发现，这种计算可以用条件概率的形式表示。为了便于阅读和理解，本文约定大写字母(如 X, Y)表示变量名称，小写字母(如 x, y)表示对应变量的特定值，大写字母粗体(如 \mathbf{X}, \mathbf{Y})表示变量集合，小写字母粗体(如 \mathbf{x}, \mathbf{y})表示对应变量特定取值集合。假设已知某地下未勘探地点的地质信息为 $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ ，则该位置的含油气后验概率可以表示为 $P(C=1|\mathbf{x})$ ，根据贝叶斯定理可进行如下转化：

$$P(C=1|\mathbf{x}) = \frac{P(C=1, \mathbf{x})}{P(\mathbf{x})} \quad (2)$$

为了减少计算公式(2)的复杂度，本文研究采用贝叶斯网络整合相关地质变量，进而计算油气的发生概率。

1.2 贝叶斯网络分类基本原理

贝叶斯网络是一种基于概率理论与图论进行大数据分析、推理和预测的数学模型。贝叶斯网络用来解决分类问题时称为贝叶斯网络分类器，其分类原理为：基于某一对象的先验概率，使用贝叶斯定理计算出其后验概率，然后使用最大后验(MAP)估计选出具有最大后验概率的类标签作为该事件的预测类标签^[21-22]。假设已知贝叶斯网络 B ，对于待分类的测试实例 $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ (n 表示属性个数)，预测类标签 c^* 的 MAP 计算过程如下：

$$\begin{aligned} c^* &= \arg \max_{c \in \Omega_C} P_B(c|\mathbf{x}) = \arg \max_{c \in \Omega_C} \frac{P_B(\mathbf{x}, c)}{P_B(\mathbf{x})} \\ &= \arg \max_{c \in \Omega_C} \frac{P_B(\mathbf{x}, c)}{\sum_{c' \in \Omega_C} P_B(c', \mathbf{x})} \propto \arg \max_{c \in \Omega_C} P_B(c, \mathbf{x}) \quad (3) \\ &= \arg \max_{c \in \Omega_C} P_B(c) \prod_{i=1}^n P_B(x_i | Pa(x_i)) \end{aligned}$$

其中， P_B 表示根据贝叶斯网络 B 的概率分布。 Ω_C 表示变量 C 的所有可能取值的集合，即公式(3)中 $\Omega_C = \{0, 1\}$ 。 $Pa(X_i)$ 表示变量 X_i 的父节点集合， $Pa(x_i)$ 表示 $Pa(X_i)$ 的具体取值，即 $Pa(x_i) \in \Omega_{Pa(X_i)}$ 。

1.3 贝叶斯网络分类器基本模型

本文主要介绍 4 种不同的贝叶斯网络分类器，分别为朴素贝叶斯分类器(NB)、树增强贝叶斯分类器(TAN)^[23]、 k 阶依赖贝叶斯分类器(KDB)^[24] 和平均一阶依赖估计器(AODE)^[25]。NB 是最简单的贝叶斯网络分类器，它假设给定类标签 C 时属性彼此独立，即

$$P_B(\mathbf{x}|c) = \prod_{i=1}^n P_B(x_i|c) \quad (4)$$

由于 NB 具有有效性和简洁性的特点，已广泛应用于分类学习。NB 分类决策公式如下：

$$c^* = \arg \max_{c \in \Omega_C} P_B(c) \prod_{i=1}^n P_B(x_i|c) \quad (5)$$

TAN 放松了 NB 中属性之间独立性的假设，将 NB 从零阶依赖扩展到一阶依赖，即 $|Pa(X_i)| \leq 2$ 。TAN 使用条件互信息 $CMI(X_i; X_j|C)$ 度量任意不相等的属性对 X_i 和 X_j 的依赖程度， CMI 定义如下：

$$CMI(X_i; X_j|C) = \sum_{x_i, x_j, c} P(x_i, x_j|c) \log_2 \frac{P(x_i, x_j|c)}{P(x_i|c)P(x_j|c)} \quad (6)$$

一般情况下，TAN 比 NB 具有更好的分类性能，通常被认为是最佳的单依赖分类器。TAN 分类决策公式如下：

$$c^* = \arg \max_{c \in \Omega_C} P_B(c) P_B(x_r|c) \prod_{i=1, i \neq r}^n P_B(x_i|c, x_{j(i)}) \quad (7)$$

其中， x_r 代表根节点 X_r 的具体取值， $X_{j(i)}$ 表示 X_i 父节点集 $Pa(X_i)$ 中除了 C 以外的节点， $x_{j(i)}$ 为 $X_{j(i)}$ 的具体取值。

KDB 允许每个属性节点最多有 k 个父节点。在构造 KDB 结构的过程中，首先对属性排序，即根据属性变量 X_i 与类变量 C 之间的互信息值对所有属性进行降序排列。然后为每个属性节点找到 k 个父节点。假设当前待选节点为 X_i ，依次为其找到 k 个具有最大 $CMI(X_i; X_j | C)$ 值的属性 X_j 作为父节点，且 $X_j \in \{X_1, X_2, \dots, X_{i-1}\}$ 。KDB 分类决策公式如下：

$$c^* = \arg \max_{c \in \Omega_C} P_B(c) \prod_{i=1}^n P_B(x_i | c, x_{i1}, \dots, x_{ip}) \quad (8)$$

其中， $\{x_{i1}, \dots, x_{ip}\}$ 是变量 X_i 的父变量集中除了 C 以外的节点的具体取值， $p = \min(i-1, k)$ 。

AODE 基于集成的思想来提高 NB 的分类性能。在 AODE 模型构建过程中，依次选择属性节点 X_i 作为超父节点，并使用类变量 C 作为其他属性的共同父节点来构建子分类器，即子分类器的数量等于属性节点的数量。最后，所有子分类器结果的平均值就是 AODE 的分类结果。AODE 分类决策公式如下：

$$c^* = \arg \max_{c \in \Omega_C} \frac{1}{n} \sum_{1 \leq i \leq n} \left(P_B(c, x_i) \prod_{j=1, j \neq i}^n P_B(x_j | c, x_i) \right) \quad (9)$$

假设 C 表示探井类别， $X = \{X_1, X_2, \dots, X_n\}$ 表示 n 个与油气相关的地质因素， $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ 表示某个待计算位置的地质信息（在本文中，节点、变量和地质因素的含义是等价的，他们分别对应图模型、概率论及石油地质上的概念，即属性节点、属性变量和地质属性表示相同的含义）。本文以 AODE 模型为例说明基于贝叶斯分类器的油气空间分布模型构建与预测过程，主要步骤如下：

(1) 如图 1 所示，依次选择每个地质属性作为其他属性的超父节点，建立 n 个子分类器的拓扑结构

($AODE_{s_1}, AODE_{s_2}, \dots, AODE_{s_n}$)。

(2) 根据每一个拓扑结构 $AODE_{s_i}$ 计算其条件概率分布 θ_{s_i} ，得到 n 个子分类器 ($AODE_1, AODE_2, \dots, AODE_n$)。

(3) 对于每一个类标签 $c_p (p=0,1)$ 重复以下步骤：

① 对于每一个 $AODE_o (o=1,2,\dots,n)$ ，使用 $AODE_o$ 计算 \mathbf{x} 属于 c_p 的概率 $P(c_p | \mathbf{x})_o$ 。

② 平均所有概率 $P(c_p | \mathbf{x}) = \frac{1}{n} \sum_{o=1}^n P(c_p | \mathbf{x})_o$ 。

(4) 返回所有类标签概率 $P(c_o | \mathbf{x})$ ， $P(c_1 | \mathbf{x})$ 。

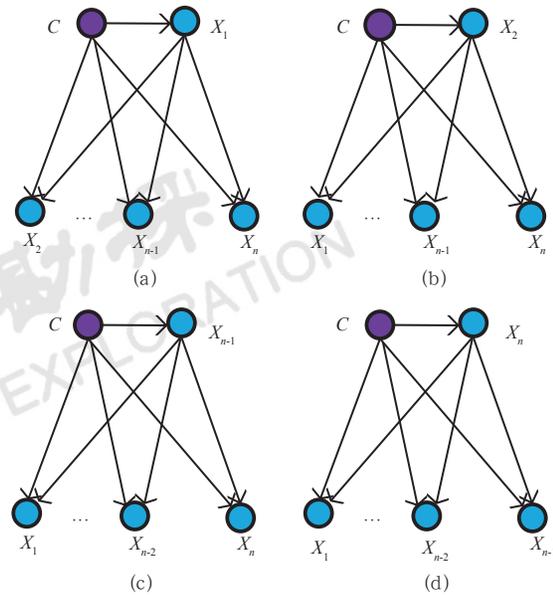


图 1 AODE 网络拓扑结构示意图

Fig.1 Schematic diagram of network topology of AODE model

1.4 4 种不同贝叶斯网络分类器的优缺点及适用范围

通过对各个方法进行深入研究分析，表 1 总结了 4 种不同贝叶斯网络分类器的优缺点及适用范围。

表 1 4 种不同贝叶斯网络分类器的优缺点及适用范围统计表

Table 1 Advantages, disadvantages and applicable scope of four types of Bayesian network classifiers

算法	优点	缺点	适用范围
朴素贝叶斯分类器 (NB)	(1) 简单、简洁高效，易于理解； (2) 不需要结构学习	假设地质属性相互独立	地质属性之间相互独立
树增强贝叶斯分类器 (TAN)	最优一阶贝叶斯网络	根节点选择为随机选择	地质属性之间具有低复杂度的相关性
平均一阶依赖估计器 (AODE)	(1) NB 的集成，准确度明显好于 NB； (2) 不需要结构学习	每个子模型的权重相同	地质属性之间具有低复杂度的相关性
k 阶依赖贝叶斯分类器 (KDB)	随 k 值增加模型准确度逐渐提高	随 k 值增加模型计算复杂度逐渐提高	地质属性之间具复杂的相关性

2 应用实例

2.1 研究区地质背景

研究区位于准噶尔盆地腹部(图2),东西宽160km,南北长170km,面积约为 $2.7 \times 10^4 \text{km}^2$,包括盆1井西凹陷、莫索湾凸起和石东凸起等^[26]。目的层为侏罗系三工河组(J_3s),烃源岩为二叠系下

乌尔禾组(P_2w),盖层为侏罗系内部泥岩(图3)。研究区南低北高,岩相南细北粗,油气自下而上、由南向北运移,形成断鼻型、断块型和岩性地层型油气藏。截至2019年底,钻遇三工河组并完成试油的探井和评价井为203口,已发现三级石油、天然气地质储量分别为 $1.128 \times 10^8 \text{t}$ 和 $459 \times 10^8 \text{m}^3$,主要分布在莫索湾凸起、莫北凸起、石西油田、石南油田和夏盐鼻凸(图2)。

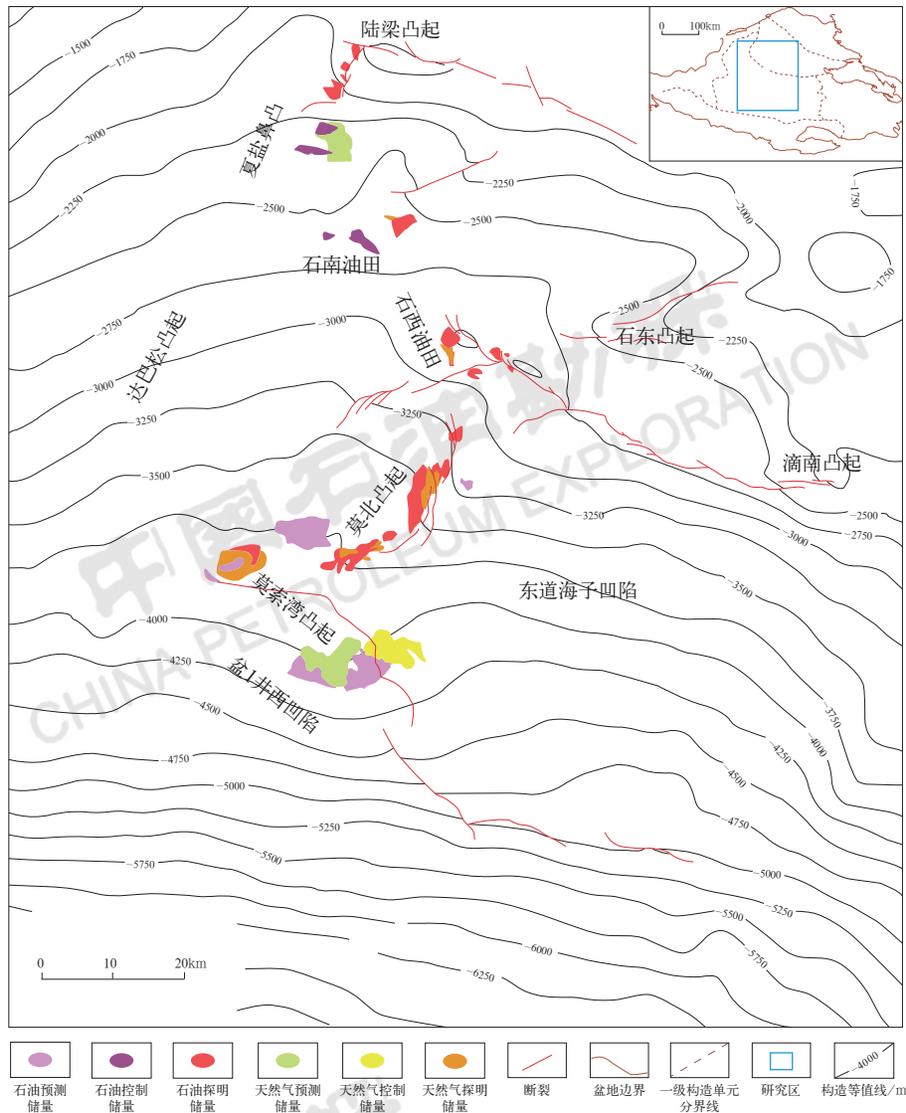


图2 研究区构造位置及油气分布图

Fig.2 Structural location of the study area and oil and gas distribution

2.2 地质条件定量评价

从供烃条件(油气来源与充注条件)、储层条件、圈闭条件、盖层与保存条件四大方面定量评价油气成藏地质条件,成藏地质条件评价价值高的地区油气资源分布概率高。

2.2.1 供烃条件

研究区三工河组油气主要来自下伏二叠系下乌尔禾组烃源岩,垂向上主要通过烃源断裂沟通,侧向运移主要受砂体和构造脊控制。研究区南部离油源较近,断层起到重要沟通作用,供烃条件整体相对有利;北部离油源远,多达上百千米,断层不能沟通烃源岩,主

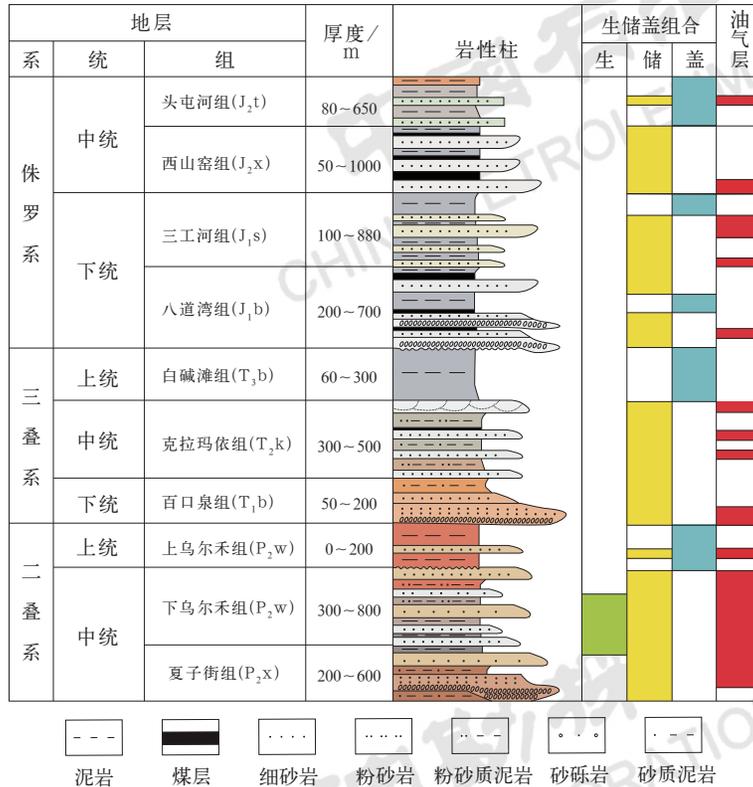


图 3 准噶尔盆地腹部地层序列及生储盖组合综合柱状图

Fig.3 Stratigraphic sequence and comprehensive stratigraphic column of source rock, reservoir, and cap rock assemblages in the hinterland of Junggar Basin

要依靠长距离侧向运移提供油气，供烃条件整体相对不利。基于以上认识，在完成断裂分布研究和油气运移路径模拟的基础上（图 4、图 5），将供烃条件划分

为 5 个等级，即：沟通油源断裂 + 主路径、主路径、次要路径、南部近源区和北部远源区（表 2）。同样，采用随机抽样方法，绘制出供烃条件定量评价图（图 6）。

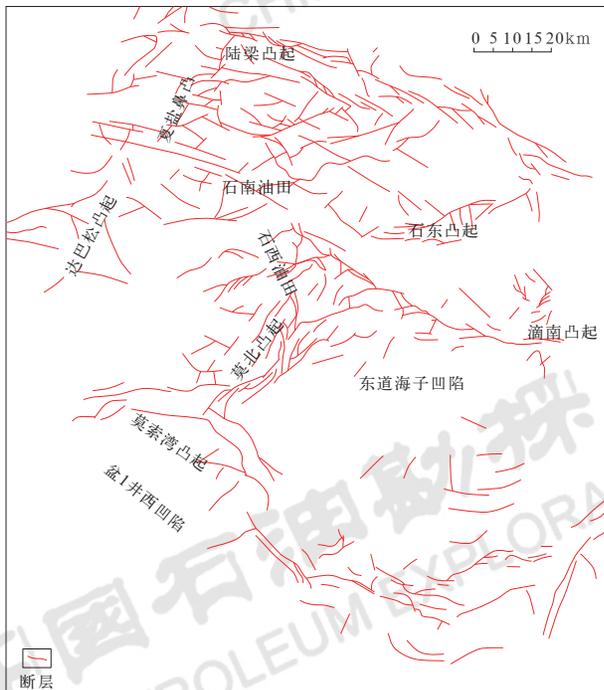


图 4 准噶尔盆地腹部三工河组断裂分布图

Fig.4 Fault distribution of Sangonghe Formation in the hinterland of Junggar Basin



图 5 准噶尔盆地腹部三工河组油气运移路径图

Fig.5 Hydrocarbon migration pathway of Sangonghe Formation in the hinterland of Junggar Basin

表2 准噶尔盆地腹部三工河组供烃条件评价价值表

Table 2 Evaluation results of hydrocarbon supply conditions of Sangonghe Formation in the hinterland of Junggar Basin

分级	供烃条件类型	最小值	最大值
I	沟通油源断裂 + 主路径	0.7	0.8
II	主路径	0.4	0.6
III	次要路径	0.3	0.5
IV	南部近源区	0.15	0.25
V	北部远源区	0.1	0.2

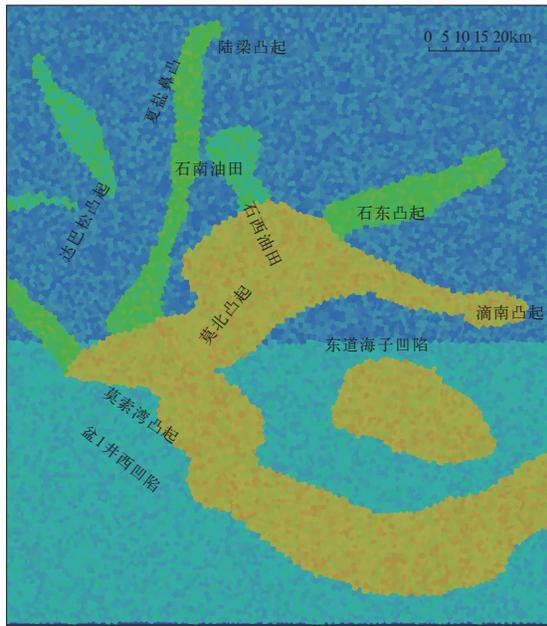


图6 准噶尔盆地腹部三工河组供烃条件定量评价图

Fig.6 Quantitative evaluation map of hydrocarbon supply conditions of Sangonghe Formation in the hinterland of Junggar Basin

2.2.2 储层条件

根据研究区 38 个三工河组油藏评价数据, 可知油藏储层平均有效厚度为 8.6m, 平均有效孔隙度为 13.87%, 平均含油饱和度为 58.7%, 平均原油密度为 0.857t/m³ (表 3); 通过分析研究区 16 个三工河组天然气藏评价数据, 可知气藏储层平均有效厚度为 7.6m, 平均有效孔隙度为 12.64%, 平均含气饱和度为 60.4%, 平均体积换算系数为 285.7 (表 3)。总体看, 研究区储层有效厚度较薄、有效孔隙度较大, 属于常规储层。

沉积相图 (图 7) 揭示, 研究区主要发育水下分流河道、分流间湾、砂质碎屑流、席状砂、滩坝等三角洲相和滨浅湖相。由统计分析可知: 位于水下分流河道和席状砂的井中, 油气井占 66.2%; 位于滩坝、砂质碎屑流和分流间湾的井中, 油气井占 37.8%; 位于滨浅湖的井中, 油气井占 21.2%。以沉积相为基础, 以统计数据为依据, 结合 38 个油藏和 16 个气藏的物性数据 (表 3), 对储层条件进行定量评价, 按优劣顺序依次为: 水下分流河道、席状砂、滩坝及砂质碎屑流、分流间湾和滨浅湖, 相应的评价值见表 4。在同一个沉积相中, 每一个点的评价值为所属沉积相最小值到最大值之间的随机抽样结果。以水下分流河道为例, 抽样范围为 0.7 到 0.9, 随机抽样结果为 0.7 到 0.9 之间的任意实数, 包括 0.7 和 0.9。根据以上方法, 将沉积相图转化为储层条件定量评价图 (图 8), 实现储层定量评价。

表3 准噶尔盆地腹部三工河组油气藏储层主要参数表

Table 3 Main reservoir parameters of Sangonghe Formation in the hinterland of Junggar Basin

油气藏类型	参数	最小值	平均值	最大值
油藏	有效厚度 /m	2.6	8.6	21.9
	有效孔隙度 /%	11.30	13.87	17.20
	含油饱和度 /%	49.3	58.7	69
	原油密度 /(t·m ⁻³)	0.824	0.857	0.899
气藏	有效厚度 /m	3.4	7.6	15.9
	有效孔隙度 /%	11.00	12.64	14.00
	含气饱和度 /%	55.0	60.4	70.0
	体积换算系数	270.0	285.7	300.3

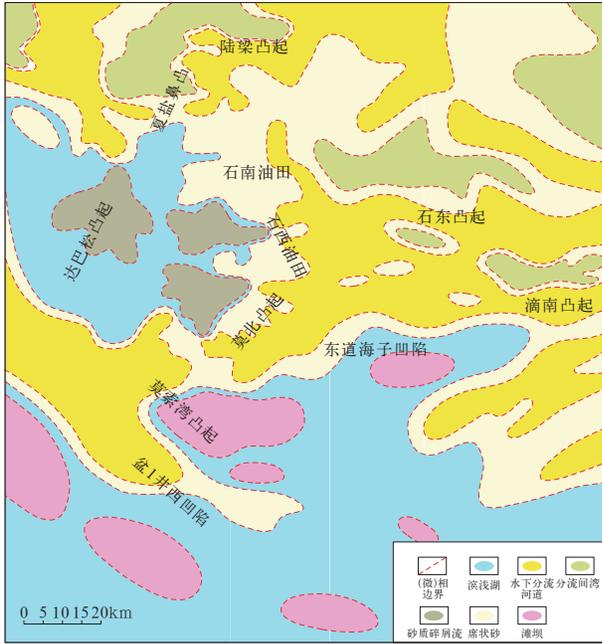


图 7 准噶尔盆地腹部三工河组沉积相图

Fig.7 Sedimentary facies map of Sangonghe Formation in the hinterland of Junggar Basin

表 4 准噶尔盆地腹部三工河组储层条件评价表

Table 4 Reservoir evaluation results of Sangonghe Formation in the hinterland of Junggar Basin

分级	沉积相	最小值	最大值
I	水下分流河道	0.7	0.9
II	席状砂	0.6	0.8
III	滩坝及砂质碎屑流	0.5	0.7
IV	分流间湾	0.3	0.5
V	滨浅湖	0.2	0.4

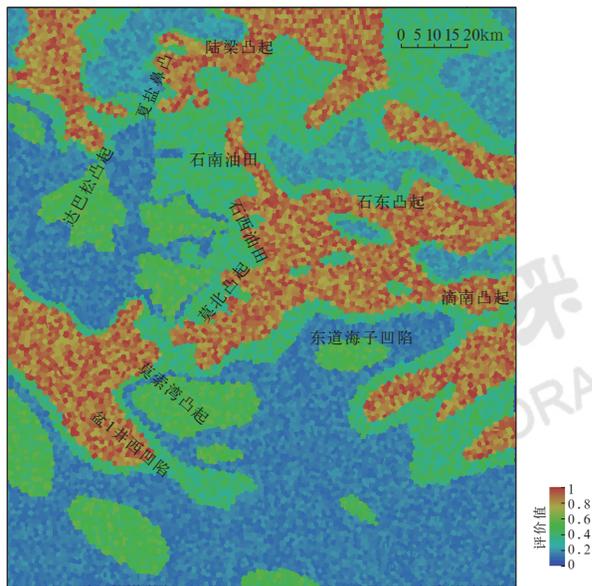


图 8 准噶尔盆地腹部三工河组储层条件定量评价图

Fig.8 Quantitative reservoir evaluation map of Sangonghe Formation in the hinterland of Junggar Basin

2.2.3 圈闭条件

根据研究区三工河组顶界构造图及圈闭解释成果资料，将构造型圈闭划分为落实圈闭和待落实圈闭两级（表 5）。根据沉积相图将岩性圈闭划分为岩性透镜体（滩坝等）圈闭和岩性遮挡（分流间湾等）圈闭两级。另外，已发现油气藏为确定性圈闭。同样采用随机抽样方法，绘制出圈闭条件定量评价图（图 9）。

表 5 准噶尔盆地腹部三工河组圈闭条件评价表

Table 5 Trap evaluation results of Sangonghe Formation in the hinterland of Junggar Basin

分级	圈闭类型	最小值	最大值
I	确定性圈闭（已发现油气藏）	1.0	1.0
II	落实圈闭（构造型）	0.7	0.8
III	待落实圈闭（构造型）	0.6	0.7
IV	岩性透镜体圈闭（滩坝等）	0.5	0.6
V	岩性遮挡圈闭（分流间湾等）	0.3	0.5
VI	其他	0	0.3

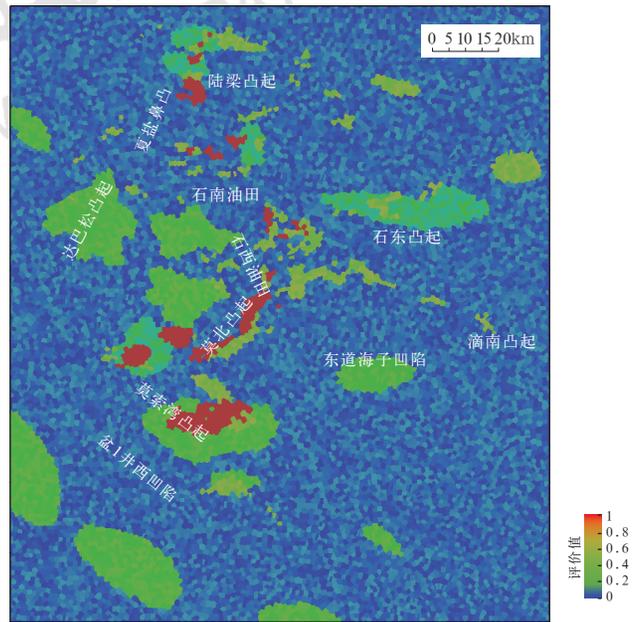


图 9 准噶尔盆地腹部三工河组圈闭条件定量评价图

Fig.9 Quantitative trap evaluation map of Sangonghe Formation in the hinterland of Junggar Basin

2.2.4 盖层与保存条件

盖层与保存条件主要考虑上覆地层断裂发育情况和不整合风化带分布。上覆地层断裂起到破坏作用，评价值较低，取 0.1~0.2；不整合风化带起到较好的遮挡作用，评价值较高，取 0.7~0.9；其他区域泥岩相对发育，具备局部盖层作用，评价值取 0.4~0.6。同样，通过随机抽样的方法获得盖层与保存条件定量评价图（图 10）。



图10 准噶尔盆地腹部三工河组盖层与保存条件定量评价图
Fig.10 Quantitative evaluation map of cap rocks and preservation conditions of Sangonghe Formation in the hinterland of Junggar Basin

(图6)、储层条件定量评价(图8)、圈闭条件定量评价(图9)、盖层与保存条件定量评价(图10), 得到203口井的地质参数表(表6), 即三工河组探井数据集。

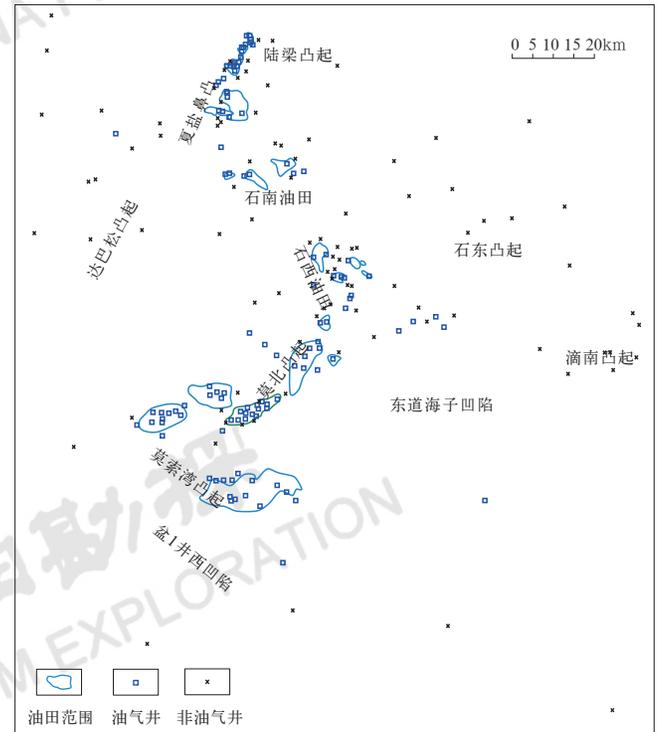


图11 准噶尔盆地腹部三工河组油气井与非油气井分布图
Fig.11 Distribution of oil and gas wells and non-oil gas wells of Sangonghe Formation in the hinterland of Junggar Basin

2.3 训练集地质参数构建

2.3.1 油气井与非油气井分布

研究区钻遇三工河组试油井203口, 其中油气井109口, 非油气井94口(图11)。将这203口井作为贝叶斯网络分类器的训练集。

2.3.2 数据集

依次采集研究区三工河组供烃条件定量评价

表6 研究区三工河组203口试油井的地质参数及井类别统计表

Table 6 Geological parameters and well types of 203 tested wells of Sangonghe Formation in the study area

序号	井名	地质参数定量评价				井类别
		供烃条件	储层条件	圈闭条件	盖层与保存条件	
1	MB13	0.13	0.36	0.08	0.44	非油气井
2	MB18	0.13	0.16	0.20	0.57	非油气井
3	MB2	0.79	0.77	1.00	0.46	油气井
4	MB9	0.78	0.83	1.00	0.75	油气井
5	MO-006	0.78	0.78	1.00	0.73	油气井
6	MO-123	0.11	0.39	0.14	0.48	非油气井
7	MO-124	0.11	0.43	0.04	0.41	非油气井
8	MO-125	0.12	0.38	0.14	0.14	非油气井
9	MO-17	0.79	0.81	1.00	0.58	油气井
10	MO-21	0.78	0.83	1.00	0.75	油气井

续表

序号	井名	地质参数定量评价				井类别
		供烃条件	储层条件	圈闭条件	盖层与保存条件	
11	MO-7	0.79	0.83	0.65	0.78	油气井
12	Pen7	0.79	0.80	1.00	0.72	油气井
13	QT1	0.13	0.38	0.15	0.18	非油气井
14	Shi001	0.13	0.31	0.07	0.53	非油气井
15	Shi003	0.13	0.42	0.12	0.58	非油气井
16	Shi020	0.79	0.76	1.00	0.60	油气井
17	SN28	0.14	0.44	0.17	0.70	非油气井
18	SN3	0.15	0.36	0.01	0.11	非油气井
19	SX2	0.79	0.80	1.00	0.75	油气井
...
203	MO-22	0.80	0.83	1.00	0.72	油气井

2.4 五折交叉验证与方法选择

2.4.1 数据分组

本文采用五折交叉验证方法测试 4 种贝叶斯网络的分类表现。图 12 展示了五折交叉验证原理示意图。具体而言，首先将 203 口探井数据集随机近似等分成 5 组，其中 4 组作为训练集，剩余的 1 组作为测试集；然后通过 5 次不同的组合形成 5 套训练集和测试集；最后将 5 次测试结果的平均值作为评估方法的最终输出值。表 7 详细地描绘了不同的分组情况及对应的实验结果。

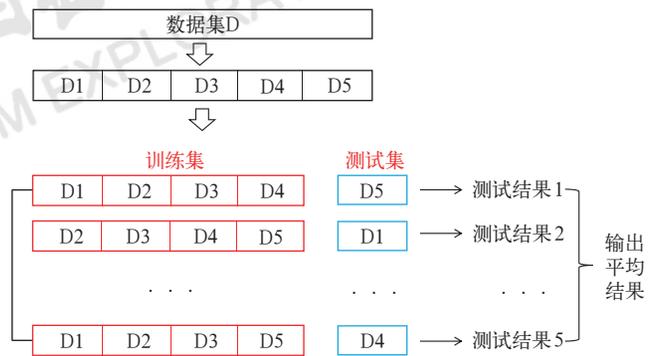


图 12 五折交叉验证原理示意图

Fig.12 Schematic diagram of five-fold cross-validation method

表 7 数据分组及五折交叉验证结果表

Table 7 Data grouping and five-fold cross-validation results

折次	训练井数 / 口	测试井数 / 口	NB		TAN		AODE		KDB	
			训练结果符合井数 / 口	测试结果符合井数 / 口						
1	162	41	137	34	140	35	140	36	143	34
2	162	41	141	33	139	33	142	34	144	33
3	162	41	139	36	140	36	141	37	144	36
4	163	40	139	33	141	33	141	33	145	33
5	163	40	140	32	144	32	145	33	146	32
正确率 / %			85.71	82.76	86.70	83.25	87.32	85.22	88.92	82.76

2.4.2 结果分析

从表7可知, NB、TAN、AODE和KDB在训练集上判别正确率分别为85.71%、86.70%、87.32%和88.92%,均大于85%,说明训练集参数分类有效;测试集判别正确率分别为82.76%、83.25%、85.22%和82.76%,均大于82%,说明预测成功率高。

2.4.3 方法选择

从训练集准确度结果来看, KDB和AODE高于TAN和NB,表明KDB和AODE对数据的拟合程度更好;但是从测试集准确度来看, AODE和TAN高于KDB和NB,说明AODE和TAN的泛化能力(预测能力)更好。综合来看, KDB过拟合数据, NB欠拟合数据,而AODE无论是对数据的拟合还是泛化能力均好于TAN,因此选择AODE作为研究区油气勘探风险预测方法。

2.5 油气勘探风险预测结果及有利区优选

本文采用二维PEBI(Perpendicular Bisection)网格技术构建了15951个评价单元。利用研究区203口井组建的训练集进行AODE模型训练,然后采用AODE模型对所有评价单元进行含油气概率预测,并绘制出含油气概率分布图(图13),概率大于50%的单元为可能的含油气单元。

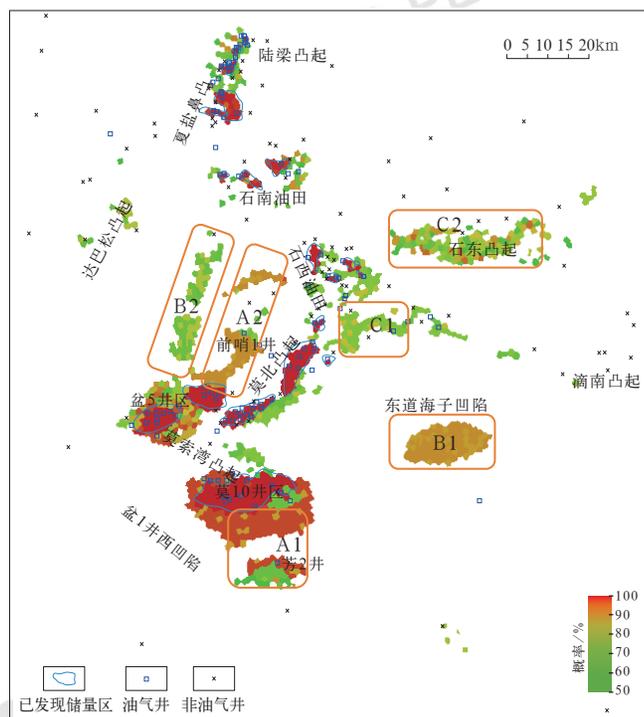


图13 准噶尔盆地腹地三工河组油气概率预测结果及有利区分布图

Fig.13 Hydrocarbon distribution probability and favorable areas prediction of Sangonghe Formation in the hinterland of Junggar Basin

从图13中可以发现,已发现储量区(三级储量)预测的油气概率最大,说明预测结果与勘探结果吻合度高。在储量区外,预测出一些油气概率大于50%的区域,主要包括3类不同类型的有利区:

(1) A类有利区。A类有利区位于储量区周边。根据预测结果优选两块(图13):①莫索湾凸起东南侧(A1),包括芳2井附近;②莫北凸起西北侧(A2),包括前哨1井附近。

(2) B类有利区。B类有利区目前暂时还未布置探井,根据预测结果优选两块(图13):①东道海子凹陷附近(B1);②莫北凸起与达巴松凸起之间(B2)。

(3) C类有利区。C类有利区属于复杂区,已有一些钻探和发现,但未能提交储量。根据预测结果优选两块(图13):①莫北凸起东北侧(C1);②石东凸起附近(C2)。

A类有利区,把握程度较高,是下一步油田扩边和滚动勘探的重点区;B类有利区,不确定性相对较高,是下一步风险勘探和甩开勘探的重点区;C类有利区,地质条件复杂,需要进一步开展研究,取得新认识后,才能制定钻探部署,提交储量。

3 结论

地质风险预测既是基于多维地质因素空间中具有不确定性的二分类问题,也是推算目的层是否存在油气资源的技术性问题。本文采用4种贝叶斯网络分类器对准噶尔盆地腹地三工河组地质风险进行预测,得出以下结论:

(1) NB、TAN、AODE和KDB 4种方法在训练集和测试集上的准确度分别均高于85%和82%,说明4种方法的有效性。通过系统分析4种方法表明AODE综合性能最好,说明选择AODE模型作为研究区油气勘探风险预测方法的合理性。

(2) 模型预测符合率达到80%以上,还存在一定的不确定性,根据模型预测数据,结合地质认识,在储量区外,预测出了一些概率大于50%的区域,主要包括三类有利区,其中,A类有利区位于莫索湾凸起东南部 and 莫北凸起西北部,该类有利区属于扩边类型,把握程度较高,也是下一步油田扩边和滚动勘探的重点区;B类有利区位于莫北凸起和达巴松凸起之间以及东道海子凹陷附近,该类有利区属于前沿未钻探新区,且不确定性相对较高,是下一步风险勘探和甩开勘探的重点区;C类有利区位于石东凸起附近和莫北凸起东北侧。

参考文献

- [1] White D A. Oil and gas play maps in exploration and assessment[J]. AAPG Bulletin, 1988,72(8):944-949.
- [2] White D A. Geologic risking guide for prospects and plays[J]. AAPG Bulletin, 1993,77(12):2048-2061.
- [3] 付小东, 张金区, 王云鹏, 等. 基于网格与GIS的油气资源空间分布与定量评价系统及其应用[J]. 地质科技情报, 2006,25(5):69-74.
Fu Xiaodong, Zhang Jinqu, Wang Yunpeng, *et al.* Oil and gas resources spatial distribution and quantitative evaluation system based on the grids and GIS and its application[J]. Geological Science and Technology Information, 2006,25(5):69-74.
- [4] 胡素云, 郭秋麟, 谯卓恒, 等. 油气空间分布预测方法[J]. 石油勘探与开发, 2007,34(1):113-117.
Hu Suyun, Guo Qiulin, Chen Zhuoheng, *et al.* A method of predicting petroleum resource spatial distribution and its application[J]. Petroleum Exploration and Development, 2007, 34(1):113-117.
- [5] 郭秋麟, 谢红兵, 米石云, 等. 油气资源分布的分形特征及应用[J]. 石油学报, 2009,30(3):379-385.
Guo Qiulin, Xie Hongbing, Mi Shiyun, *et al.* Fractal model for petroleum resource distribution and its application[J]. Acta Petrolei Sinica, 2009,30(3):379-385.
- [6] Olea R A, Cook T A, Coleman J L. A methodology for the assessment of unconventional (continuous) resources with an application to the greater natural buttes gas field, Utah[J]. Natural Resources Research, 2010,19(4):237-251.
- [7] 李军, 杨晓娟, 张晓龙, 等. 基于三维马尔可夫链模型的岩性随机模拟[J]. 石油学报, 2012,33(5):846-853.
Li Jun, Yang Xiaojuan, Zhang Xiaolong, *et al.* Lithologic stochastic simulation based on the three-dimensional Markov chain model[J]. Acta Petrolei Sinica, 2012,33(5):846-853.
- [8] Chen Z H, Osadetz K G. An assessment of tight oil resource potential in Upper Cretaceous Cardium Formation, western Canada sedimentary basin[J]. Petroleum Exploration and Development, 2013,40(3):320-328.
- [9] 庞敏, 吴欣桐, 梁琳. 基于贝叶斯概率的页岩气开发投资决策研究[J]. 统计与决策, 2014(19):63-65.
Pang Min, Wu Xintong, Liang Lin. Research on investment decision-making in shale gas development based on Bayesian probability[J]. Statistics and Decision-Making, 2014(19):63-65.
- [10] 王泽龙, 刘先贵, 李志勇, 等. 基于贝叶斯理论的四维地震与生产动态数据的油藏历史拟合新方法[J]. 石油科技论坛, 2022,41(2):9-22.
Wang Zelong, Liu Xiangui, Li Zhiyong, *et al.* A novel Bayesian-based reservoir history matching method for 4D seismic and production data[J]. Petroleum Science and Technology Forum, 2022,41(2):9-22.
- [11] 王宜林, 赵文智, 杜金虎, 等. 中国石油风险勘探工程决策管理与实施成效[J]. 中国石油勘探, 2021,26(1):1-16.
Wang Yilin, Zhao Wenzhi, Du Jinhui, *et al.* CNPC: decision management and implementation results of venture exploration projects[J]. China Petroleum Exploration, 2021,26(1):1-16.
- [12] 杨丽娜, 解国军. 概率论和数理统计在油气资源评价中的应用[J]. 内蒙古石油化工, 2016,42(4):121-124.
Yang Lina, Xie Guojun. Application of probability theory and mathematical statistics in oil and gas resource assessment[J]. Inner Mongolia Petrochemical Industry, 2016,42(4):121-124.
- [13] Peel F J, Brooks J R V. A practical guide to the use of success versus failure statistics in the estimation of prospect risk[J]. AAPG Bulletin, 2016,100(2):137-150.
- [14] Fanailoo P, Clemens J. Application of Bayesian networks for estimating risk in shale operations[C]//Unconventional Resources Technology Conference, Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2016:1776-1782.
- [15] 盛秀杰, 金之钧, 肖晔. 区带勘探中的油气资源评价方法[J]. 石油与天然气地质, 2017,38(5):983-992.
Sheng Xiujie, Jin Zhijun, Xiao Ye. Petroleum resources assessment methodology in play exploration stages[J]. Oil & Gas Geology, 2017,38(5):983-992.
- [16] Lilleborge M, Hauge R, Eidsvik J. Information gathering in Bayesian networks applied to petroleum prospecting[J]. Mathematical Geosciences, 2016,48(3):233-257.
- [17] Martinelli G, Eidsvik J, Hauge R, *et al.* Bayesian networks for prospect analysis in the north Sea[J]. AAPG Bulletin, 2011, 95(8):1423-1442.
- [18] Van Wees J D, Mijnlief H, Lutgert J, *et al.* A Bayesian belief network approach for assessing the impact of exploration prospect interdependency: an application to predict gas discoveries in the Netherlands[J]. AAPG Bulletin, 2008,92(10): 1315-1336.
- [19] de Lucena B R D, Lustosa L J. Bayesian approach to the assessment of geological risk in oil and gas exploration[J]. International Journal for Innovation Education and Research, 2020,8(7):203-210.
- [20] Ren H J, Wang X C, Guo Q L, *et al.* Spatial prediction of oil and gas distribution using tree augmented Bayesian network[J]. Computers & Geosciences, 2020,142:104518.
- [21] Bielza C, Larranaga P. Discrete Bayesian network classifiers: a survey[J]. ACM Computing Surveys (CSUR), 2014,47(1):1-43.
- [22] Wang X, Ren H, Guo X. A novel discrete firefly algorithm for Bayesian network structure learning[J]. Knowledge-Based Systems, 2022,242:108426.
- [23] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers[J]. Machine Learning, 1997,29 (2):131-163.
- [24] Sahami M. Learning limited dependence Bayesian classifiers[C]. Menlo Park: AAAI Press, 1996:335-338.
- [25] Webb G I, Boughton J R, Wang Z. Not so naive Bayes: aggregating one-dependence estimators[J]. Machine Learning, 2005,58(1):5-24.
- [26] 王小军, 宋永, 郑孟林, 等. 准噶尔盆地复合含油气系统与复式聚集成藏[J]. 中国石油勘探, 2021,26(4):29-43.
Wang Xiaojun, Song Yong, Zheng Menglin, *et al.* Composite petroleum system and multi-stage hydrocarbon accumulation in Junggar Basin[J]. China Petroleum Exploration, 2021,26(4): 29-43.