

基于 HMM 的缝洞型油藏产量预测算法

崔书岳¹, 黄晓辉², 陈云亮², 郑松青³

1. 中国石油化工股份有限公司 石油勘探开发研究院 西北地区勘探开发研究中心, 北京 100083;

2. 中国地质大学 计算机学院, 武汉 430074;

3. 中国石油化工股份有限公司 石油勘探开发研究院 油田开发所, 北京 100083

摘要: 缝洞型油藏具有非均质性强, 开采难度大的特点. 预测油藏产量变化规律对后期开发方案的制定、调整以及实际的生产优化至关重要. 本文以塔河油田实际生产数据为基础, 建立了基于 HMM(隐马尔可夫模型)的缝洞型油藏产量变化规律预测模型. 该模型可预测未来一周油藏生产并产油趋势, 为油藏生产开发提供指导. 实验结果显示, 本文提出的基于 HMM 的产量预测算法预测精度最大可达 85.71%.

关键词: 油藏工程; 缝洞型油藏; 塔河油田; 产油量预测; HMM

中图分类号: TE344

文献标志码: A

文章编号: 1673-9868(2020)02-0137-08

目前, 塔里木盆地塔河油田是中国已探明石油地质储量最大的缝洞型油藏, 其埋藏深度大, 岩溶储层复杂, 缝洞发育程度高, 储集空间结构多变, 油水流动关系复杂, 分布受沉积、构造、古地貌及多期岩溶作用控制. 溶洞是该类油藏的主要储集空间, 也是区别于其他类型油藏的最主要特点. 以塔河油田为代表的缝洞型油藏是一种以大型的溶洞和宽裂缝为主要储集空间的特殊类型油藏, 其有效储集空间多是由岩溶和断裂的作用发育而来, 具有储集体分布复杂, 非均质性强, 主体缝洞内的流体流动不符合达西渗流规律等特征^[1]. 在实际生产过程中, 预测缝洞型油藏的产量变化规律十分重要, 良好的产量预测方法在开发方案的优化中起到关键的作用. 现有的油藏产量预测方法中, 基于传统的油藏渗流理论在描述这类缝洞型油藏时存在困难; 基于数值模拟的方法主要通过经验公式, 因此主观性较大, 在实际生产中应用较为困难; 基于数据挖掘的方法虽然能够取得不错的性能, 但无法反映油藏生成过程中产油量的变化状况, 并且存在不确定性. 因此本文基于塔河油田实际生产数据, 使用隐马尔可夫模型(Hidden Markov Model, HMM)^[2-3]来构建油藏产量预测模型, 以对油藏产量趋势进行预测.

1 隐马尔可夫模型

HMM 描述了含有未知参数的马尔可夫过程^[4-5], 该模型被广泛应用在态势评估、语音识别等领域^[6-8].

假设现有一观测序列 $x = \{x_1, x_2, \dots, x_m\}$, 其中 $x_m \in \{O_1, O_2, \dots, O_p\}$, p 为观测值集合中元素的个数; 一隐含状态序列 $y = \{y_1, y_2, \dots, y_n\}$, 其中 $y_n \in \{S_1, S_2, \dots, S_q\}$, q 为隐含状态集合中元素的个数^[9-10]. 用 a_{ij} 表示从状态 S_i 转移到状态 S_j 的概率, 即: $a_{ij} = P(S_j | S_i)$, 其中 $1 \leq i, j \leq q$. 由 a_{ij} 组成的

收稿日期: 2019-02-26

基金项目: 国家科技重大专项资助项目(2016ZX05014); 国家自然科学基金项目(51674285).

作者简介: 崔书岳(1979-), 男, 工程师, 硕士, 主要从事油藏数值模拟等方面的研究.

通信作者: 黄晓辉, 博士研究生, 主要从事地学数据管理方面的研究.

矩阵叫做状态转移概率矩阵, 记作 $\mathbf{A}$, 矩阵 $\mathbf{A}$ 描述了不同状态之间进行转移的概率^[11-13].

b_{jk} 用来表示状态值为 S_j 时产生的观测值为 O_k 的概率, 即:

$$b_{jk} = P(O_k | S_j)$$

其中 O_k 是状态 S_j 下的观测值, $1 \leq k \leq p$, $1 \leq j \leq q$. 由 b_{jk} 组成的矩阵叫做观测值概率分布矩阵, 记作 $\mathbf{B}$, 矩阵 $\mathbf{B}$ 描述了在不同的隐含状态下生成不同观测值的概率. HMM 可由一个五元组 $(P, Q, \mathbf{A}, \mathbf{B}, \alpha)$ 表示, 其中 P 为观测值的有限集合, Q 为状态值的有限集合, $\mathbf{A}$ 为状态转移概率矩阵, $\mathbf{B}$ 为观测值概率分布矩阵, α 为初始状态概率分布^[14].

HMM 能解决 3 类问题^[15-16]:

- 1) 解码问题: 已知 HMM 和观测序列, 求解产生该观测序列的最佳隐藏状态序列;
- 2) 学习问题: 已知观测序列, 求解最优的 HMM 参数, 使得产生该观测序列的可能性最大;
- 3) 估计问题: 已知 HMM 和观测序列, 求解在所有可能的隐藏状态序列下, 产生给定观测序列的概率.

2 基于 HMM 的产油趋势预测

本文结合 HMM 和油藏背景, 通过塔河油田区域真实的生产井历史产油数值和产气数值, 研发基于 HMM 的油藏生产井产油趋势预测方法.

图 1 是基于 HMM 的油藏生产井产油趋势预测方法整体流程, 整个预测模型的输入数据是塔河区域油藏生产井的历史产油数据和产气数据, 模型首先对输入数据进行预处理, 以确定最终 HMM 的输入信息, 预处理过程包括确定状态值集合、确定观测值集合、生成观测序列和生成状态序列 4 个步骤. 经过预处理产生的观测序列集合和状态序列集合作为 HMM 的输入数据, 该输入数据被分为训练数据和测试数据两部分, 其中训练数据用来训练 HMM, 生成状态转移概率矩阵和观测值概率矩阵, 而测试数据集用于测试训练出来的 HMM 的精度. 基于训练好的 HMM 可以用来对油藏生产未来一段时间的产油量趋势进行预测.

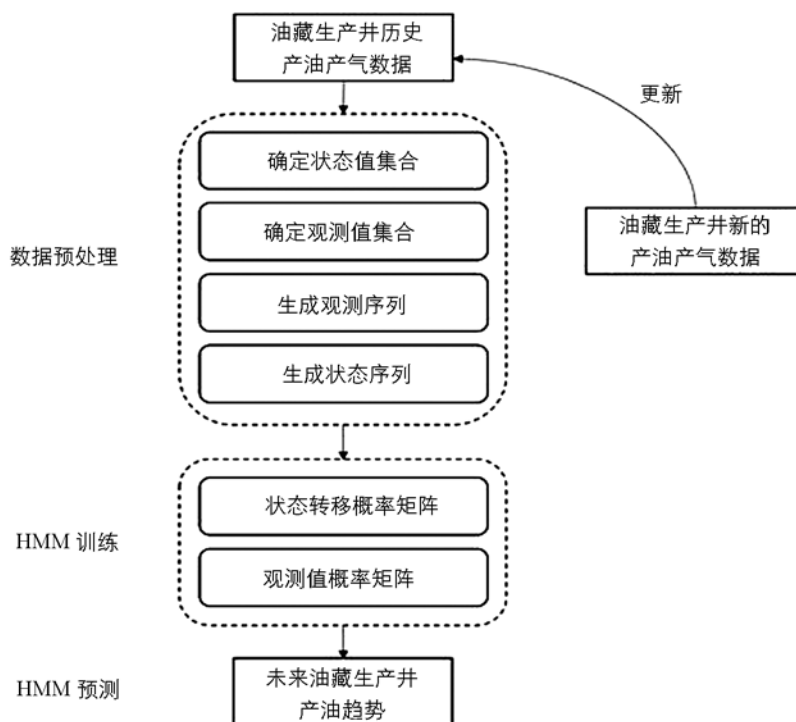


图 1 HMM 预测产油趋势流程图

由于油藏生产是一个持续的过程, 生产井每天都会得到新的产出数值, 将油藏生产井每日最新产生的产油值和产气值用来更新数据集, 重新训练 HMM, 可以保证 HMM 的预测精度.

2.1 生成观测值序列

为了对产油历史数据和产气历史数据进行预处理以生成 HMM 中的观测值序列, 本文引入生产指标的概念对历史产油数据和产气数据进行离散化处理. 生产指标指在油藏生产过程中, 每口油藏生产井在不同产出(油、汽、水)情况下预期的日产量. 产油生产指标指油藏生产井每日产油的预期值, 产气生产指标指油藏生产井每日产气的预期值, 产水生产指标指油藏生产井每日产水的预期值. 若油藏生产井某一日产油量大于或等于产油生产指标, 说明该油藏生产井当天产油量达标; 反之, 若油藏生产井某一日产油量小于产油生产指标, 说明该油藏生产井当天产油量未达标, 产出不足. 生产指标可以衡量油藏生产井每日的产出(油、气、水)效率. 一般情况下, 生产指标可由油藏开采部门结合每口油藏生产井的实际情况确定. 为将人工干预的影响降低到最小, 本文采用油藏生产井产出历史数据的平均值作为不同产出的生产指标.

通过引入产油生产指标和产气生产指标, 按照算法 1 的流程, 可以将油藏生产井历史产油数据和历史产气数据转换为历史产油标记序列和历史产气标记序列, 从而生成 HMM 中的观测值序列.

算法 1 标记序列产生算法

输入: 油藏生产井历史产油数据序列 Oil_D , 油藏生产井历史产气数据序列 Gas_D , 其中 D 为天数(即数据大小)

输出: 油藏生产井历史产油标记序列 $Label_Oil_D$ 和历史产气标记序列 $Label_Gas_D$

开始:

/ * 得到油藏生产井历史产油数据的平均值和历史产气数据的平均值 * /

$avgOil = \text{sum}(Oil_D) / D$

$avgGas = \text{sum}(Gas_D) / D$

FOR $i = 1 \rightarrow D$

IF $Oil_i \leq avgOil$

$Label_Oil_i = 0$

ELSE $Label_Oil_i = 1$

END

IF $Gas_i \leq avgGas$

$Label_Gas_i = 0$

ELSE $Label_Gas_i = 1$

END

END

通过算法 1 得到油藏生产井历史产油标记序列 $Label_Oil_D = \{L_{o1}, L_{o2}, \dots, L_{oi}, \dots, L_{oD}\}$, $1 \leq i \leq D$, $L_{oi} \in \{0, 1\}$ 和历史产气标记序列 $Label_Gas_D = \{L_{g1}, L_{g2}, \dots, L_{gi}, \dots, L_{gD}\}$, $1 \leq i \leq D$, $L_{gi} \in \{0, 1\}$. 油藏生产井每日产油标记和每日产气标记可形成一个二元组 $OG = \langle L_{oi}, L_{gi} \rangle$, $1 \leq i \leq D$, 将该二元组作为 HMM 中的一个观测点的值, 那么可以得到 HMM 中的有限观测值集合 $P = \{\langle 0, 0 \rangle, \langle 0, 1 \rangle, \langle 1, 0 \rangle, \langle 1, 1 \rangle\}$.

图 2 是基于油藏生产井历史产油数据和历史产气数据生成观测序列的一个示例.

2.2 生成状态值序列

本文引入 3 个状态值来描述油藏生产井在一段时间内(3 d, 5 d 等)的变化趋势, 分别为: 上升、下降和震荡. 图 3 以 1 d 作为一个观测点, 7 d 作为一个观测时间段, 描述了在观测时间段内油藏生产井产油的变

化趋势. 当油藏生产井的状态值为上升时, 表示油藏生产井在观测时间段内的产油变化趋势呈现上升趋势; 当油藏生产井的状态值为下降时, 表示油藏生产井在观测时间段内的产油变化趋势呈现下降趋势; 当油藏生产井的状态值为震荡时, 表示油藏生产井在观测时间段内的产油变化较平稳, 不明显.

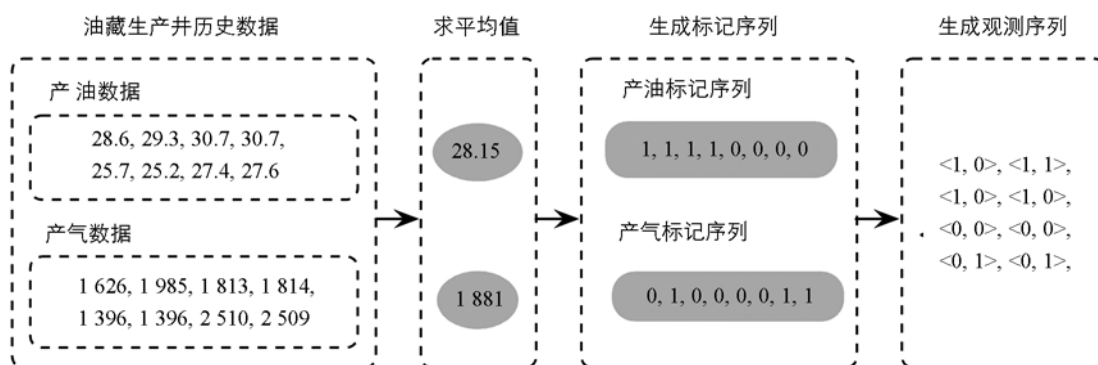


图 2 基于油藏生产井历史产油数据和历史产气数据生成的观测序列

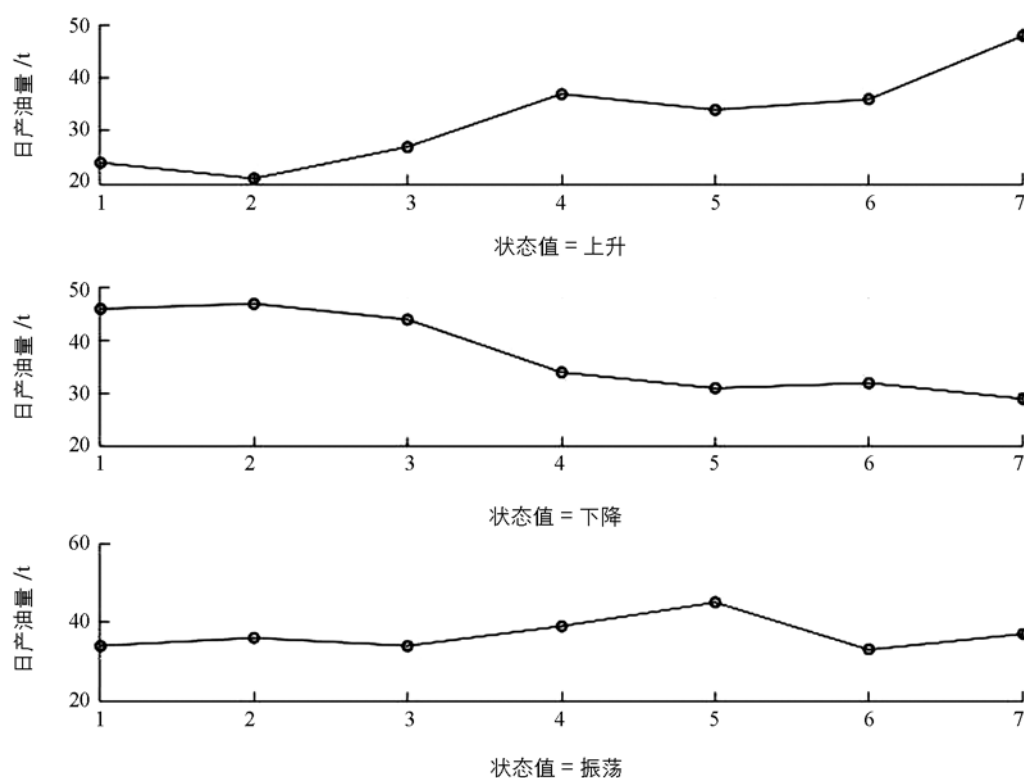


图 3 状态值示意图

在得到指定的一段观测时间内的油藏生产井产油数据后, 根据算法 2 中的算法流程可以得到部分观测点油藏生产井未来的产油趋势. 其中 ϵ 为设定的阈值, 其值为在观测时间段内最后一个观测点所观测到的油藏生产井产油数值与第一个观测点所观测到的油藏生产井产油数值的差.

从算法 2 可以看出, 在状态序列产生的过程中, 我们将一天作为一个观测点, 每一个观测点的状态值描述了该点后 (即该天之后) 油藏生产井的产油趋势, 该观测点的状态值是根据这一天后面 T 天的数据计算得出的, 因此, 对于一个给定的油藏生产井历史产油数据集, 其观测点的状态值无法计算, 所以油藏生产井历史产油数据集最后 T 个没有状态值. 在实验过程中, 为了消除该问题的影响, 可以选取全部油藏生产井历史产油数据集中的一部分作为实验数据集 (例如, 选取长度为 1 000 的总数据集集中的前 900 d 的数据作为实验数据).

算法2 状态序列产生算法

输入: 油藏生产井历史产油数据序列 Oil_D , D 为天数(即数据大小), 阈值为 ϵ , 观测时段为 T

输出: 油藏生产井历史状态 $State_Oil_{D-T}$

开始:

FOR $i=1 \rightarrow (D-T)$

$firstValue = Oil_{i+1}$

$lastValue = Oil_{i+2}$

IF $(lastValue - firstValue) \geq 0$

IF $(lastValue - firstValue) \geq \epsilon$

$State_Oil_i = \text{上升}$

ELSE $State_Oil_i = \text{震荡}$

END

ELSE IF $|lastValue - firstValue| \geq \epsilon$

$State_Oil_i = \text{下降}$

ELSE $State_Oil_i = \text{震荡}$

END

END

在本文提出的 HMM 中, 状态值有限集合 $Q = \{\text{上升}, \text{下降}, \text{震荡}\}$, 为方便实验, 在实验过程中, 可用 1, 2, 3 这 3 个数值分别表示上升、下降和震荡。

以图 2 中的油藏生产井历史产油数据为例, 描述如何基于算法 2 生成状态序列集。现有的油藏生产井历史产油数据集 $Oil_g = \{28.6, 29.3, 30.7, 30.7, 25.7, 25.2, 27.4, 27.6\}$, 假设观测时段 $T=3$, 阈值 $\epsilon=1$, 那么根据算法 2 的流程可得到该历史产油数据集产油趋势序列 $State_Oil = \{\text{上升}, \text{下降}, \text{下降}, \text{上升}, \text{上升}\}$, 用 1, 2, 3 这 3 个数值分别表示上升、下降和震荡, 那么得到的状态序列为 $State_Oil = \{1, 2, 2, 1, 1\}$ 。

3 实验及分析

3.1 实验环境

本实验硬件环境为: Inter(R) Core(TM) i7-6500U CPU @ 2.50 GHz, 8 GB 内存。

本实验软件环境为: Windows Ultimate 7 64 bit, Matlab R2015b, 代码使用 Matlab 编写。

本实验使用的数据集为塔河油田 5 口生产井 T705、T706、T801、T808 和 T702B 的历史生产数据, 数据集详细信息如图 4 所示。

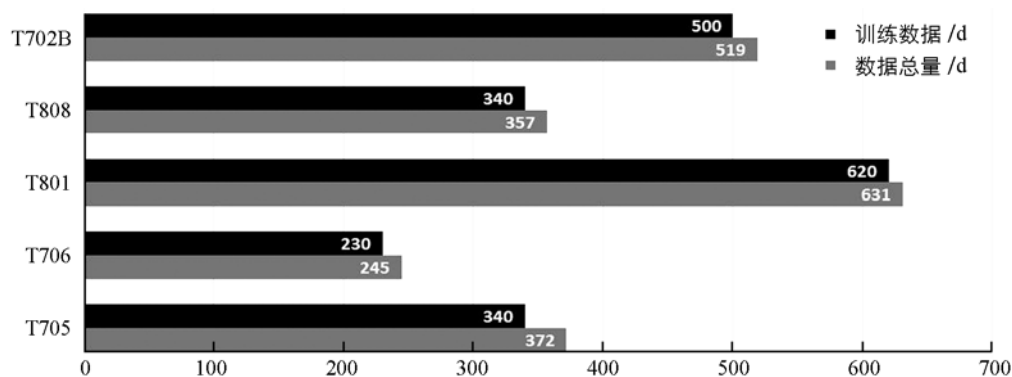


图 4 实验数据集

本实验中,观测时段设为 3 d,即每一个观测点的未来状态值由该观测点后续 2 个观测点确定;预测时段设为 7 d,即建立好的 HMM 可以预测未来 7 d 油藏生产井的产油趋势;阈值,即油藏生产井在观测时段前后的产油值差值。

3.2 实验结果

图 5 展示了本文提出的基于 HMM 的油藏生产井产油趋势预测方法在不同数据集上的结果。

对于不同油井的历史数据,将其分为训练数据和预测数据,基于每口井的训练数据集训练出来的 HMM 输入预测集进行预测,预测结果与真实结果进行比较其准确度也高达 71.42%(即 7 个观测点的状态值预测中,准确的有 5 个)以上。图 5 展示了每个数据集的训练数据总量和训练数据的天数均不同,但是所得到的预测精度却大致相同,由此可以看出 HMM 预测的准确度跟上一个状态有关系而跟状态的总数无关,能够较好地拟合油藏历史数据并进行预测。

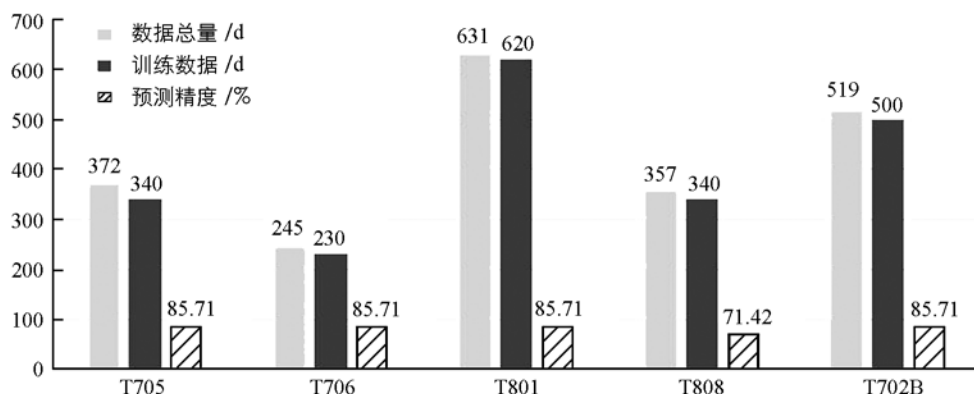


图 5 预测结果

图 6 的时间比较展示了 5 个不同的数据集训练数据以及预测数据所消耗的总时间,在同一台机器上进行了这 5 组实验,对于不同的数据集,平均运行时间是 42.17 ms。用时最长的是 T705 和 T706 数据集,用时 45.58 ms。用时最短的是 T702B 数据集,仅用 34.16 ms。HMM 模型拥有高效的训练效率的原因是模型是基于隐马尔可夫假设的,即当前时刻点的状态仅与上一个时刻点的状态有关。这一假设在减少了预测时计算量的同时,也减轻了 2 个概率矩阵的构建和计算难度。

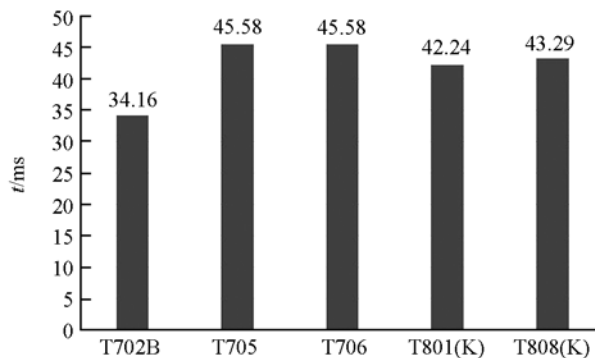


图 6 时间比较

由图 6 结合图 5 不同数据集的训练数据总量和训练数据天数比较可以看出,训练数据和预测数据的总时间与训练数据总量以及训练数据天数没有明显关系,因为训练数据和预测数据的总时间最少的数据集 T705 所使用的数据总量和训练数据天数并不是最少的,而且训练数据和预测数据总时间相同的 2 个数据集 T706 和 T801 所使用的数据总量和训练数据天数却有着近两倍以上差距。但总体来说,采用隐马尔可夫进行预测消耗的时间很短,在处理流数据时,可以进行实时处理并且保持平均较高的正确率。

图 7 展示了 T702B 训练天数与训练精度之间的关系,选取了间隔为 50 d 共计 20 组的训练数据,预测数据为固定的 7 d,由图 7 可以看出,在预测数据不变的情况下,训练精度随着训练天数增加的总体变化趋势是先增加再减少,由此可以看出针对 T702B 这口井, HMM 在 400~650 d 这个区间内训练效果相对较好,也就是拟合程度比较高,当训练天数低于该区间时,拟合程度不够;当训练天数高于该区间

时, 会导致训练的 HMM 预测模型过于贴近于训练集的变化规律, 导致模型过拟合, 使得将训练好的模型应用在测试数据集时出现预测精度下降的现象. 因此选取合适的训练天数是非常有必要的, 对训练精度有比较大的影响.

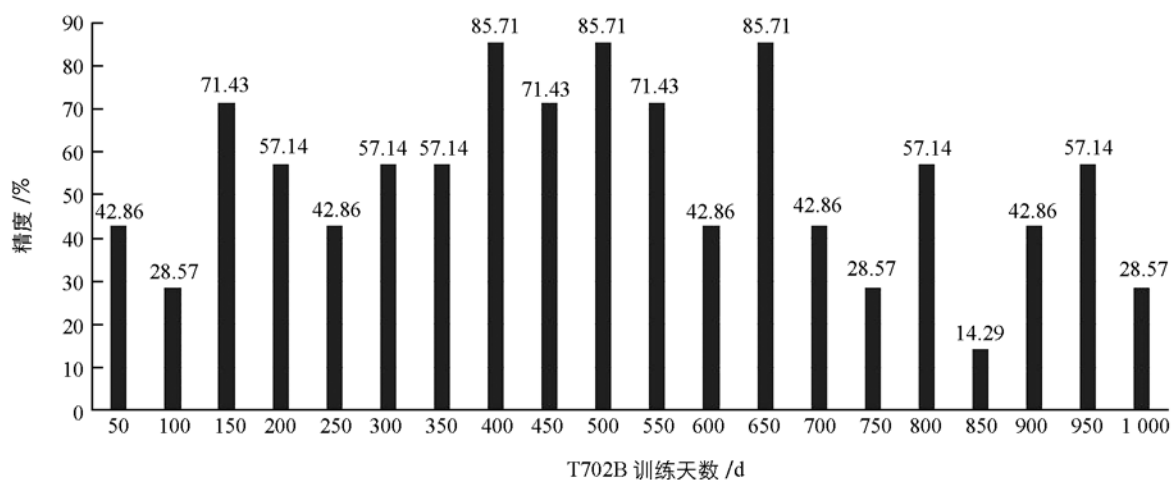


图 7 T702B 训练天数与训练精度关系

4 总 结

本文提出了一种基于 HMM 的油藏生产井产油趋势预测算法. 通过数据预处理、模型训练建立起基于隐马尔可夫的油藏生产预测模型, 用来预测油藏生产井未来一周的产油趋势, 为油藏生产开发提供指导. 实验结果表明, 本文提出的基于 HMM 的油藏生产井产油趋势预测算法, 其预测精度最高可达 85.71%, 并且预测消耗时间仅为 42.17 ms, 证明了本文所提出算法的有效性.

参考文献:

- [1] 康志宏, 陈琳, 鲁新便, 等. 塔河岩溶型碳酸盐岩缝洞系统流体动态连通性研究 [J]. 地学前缘, 2012, 19(2): 110-120.
- [2] JAIN A K, DUIN P W, MAO J. Statistical Pattern Recognition: A Review [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(1): 4-37.
- [3] 宋刚, 肖国强, 代毅, 等. 基于视频区域特征及 HMM 的体育视频分类研究 [J]. 西南师范大学学报(自然科学版), 2010, 35(2): 180-184.
- [4] 林已杰, 赖清, 周敏. 基于 BP 神经网络和马尔科夫模型的服务器软件老化预测方法 [J]. 西南师范大学学报(自然科学版), 2011, 36(4): 193-197.
- [5] 张一进, 李扬荣. 由 Markov 对偶分支 q -矩阵导出的算子的性质 [J]. 西南大学学报(自然科学版), 2013, 35(4): 88-90.
- [6] 吴建台, 刘光杰, 刘伟伟, 等. 一种基于关联分析和 HMM 的网络安全态势评估方法 [J]. 计算机与现代化, 2018(6): 30-36.
- [7] 鄢发齐, 王春明, 窦建中, 等. 基于隐马尔可夫模型的电力调度语音识别研究 [J]. 武汉大学学报(工学版), 2018, 51(10): 920-923.
- [8] 刘瑶, 余旭, 黄智兴. 基于 Kinect 传感器的动态手势实时识别 [J]. 西南大学学报(自然科学版), 2015, 37(8): 132-137.
- [9] CHAMPION C, HOUGHTON S M. Application of Continuous State Hidden Markov Models to a Classical Problem in Speech Recognition [J]. Computer Speech & Language, 2016, 36: 347-364.
- [10] 苏永莉, 阎建忠, 周洪. 重庆市的森林转型: 时空格局与动态模拟 [J]. 西南大学学报(自然科学版), 2016, 38(6):

82-91.

- [11] 赵文强. 正则 Markov 链的收敛与逼近(英文) [J]. 西南大学学报(自然科学版), 2010, 32(8): 148-151.
- [12] 赵文强, 商彦英. Markov 积分半群的非退化性及弱 * 可微性 [J]. 西南师范大学学报(自然科学版), 2006, 31(6): 24-27.
- [13] 赵文强, 李 嘉. Markov 积分半群的生成元 [J]. 西南师范大学学报(自然科学版), 2007, 32(5): 14-17.
- [14] OSTENDORF M, DIGALAKIS V V, KIMBALL O A. From HMM's to Segment Models: A Unified View of Stochastic Modeling for Speech Recognition [J]. IEEE Transactions on Speech and Audio Processing, 1996, 4(5): 360-378.
- [15] 朱 明, 郭春生. 隐马尔可夫模型及其最新应用与发展 [J]. 计算机系统应用, 2010, 19(7): 255-259.
- [16] 薛文鹏, 杨 芮, 杨 洪, 等. 基于马尔科夫链理论对贵州锦屏白背飞虱发生程度的预测 [J]. 西南大学学报(自然科学版), 2017, 39(8): 43-48.

An HMM-Based Method for Forecasting the Production Trend of Fractured-vuggy Reservoirs

CUI Shu-yue¹, HUANG Xiao-hui²,
CHEN Yun-liang², ZHENG Song-qing³

1. Northwest Research Center, Petroleum Exploration & Production Research Institute,
China Petroleum & Chemical Corporation, Beijing 100083, China;

2. School of Computer Science, China University of Geosciences, Wuhan 430074, China;

3. Oilfield Development Department, Petroleum Exploration & Production Research Institute,
China Petroleum & Chemical Corporation, Beijing 100083, China

Abstract: Fractured-vuggy reservoirs are characterized by high anisotropism and great difficulty in exploitation. Petroleum production trend prediction is an essential step for the design and adjustment of development plan and production optimization. Using the real production data of Tahe Oilfield, this paper proposes a prediction model for the petroleum production trend of fractured-vuggy reservoirs based on HMM (Hidden Markov Model). The proposed method can be utilized to predict the dynamic trend of oil production of the following week, thus providing guidance for the production and exploitation of fractured-vuggy reservoirs. In an experiment the proposed method gives a prediction accuracy as high as 85.71%.

Key words: petroleum reservoir engineering; fractured-vuggy reservoir; Tahe oilfield; prediction on oil production; HMM (Hidden Markov Model)

责任编辑 崔玉洁