XIII Всероссийская
научно-практическая конференция
«Технологии Microsoft
в теории и практике программирования»

УДК 004

ТЕХНОЛОГИЯ СОЗДАНИЯ УНИВЕРСАЛЬНЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ДАННЫХ С НЕСКОЛЬКИХ НЕФТЯНЫХ СКВАЖИН

Вайтулевич А.В., Станкевич Ф.В.
Научный руководитель: Кудинов А.В., канд. техн. наук, доцент каф. ВТ

*Национальный Исследовательский Томский политехнический университет,
634050, Россия, г. Томск, пр. Ленина, 30
E-mail: vaanval@gmail.com*

Введение

Благодаря расширенному использованию информационных систем в сфере добычи нефти и газа предприятия стали обладать большим объёмом технологических данных, которые, в случае успешного применения интеллектуального анализа, могут быть использованы для улучшения процесса добычи.

This paper focuses on methods and instruments aimed to apply data mining to analyze the data generated by oil wells. We will analyze such data and use it to predict various parameters of oil wells and to prevent upcoming equipment failures. In order to implement data mining solutions CRISP-DM will be used, since it incorporates iterative approach needed for countering possible design flaws [1].

Анализ предметной области

Many oil production companies equip most of their oil with automatic technological parameters registering systems which monitor oil production and related processes. The state of an oil well can be described by a set of parameters associated with equipment sensors and sensors measuring physical parameters of underground oil layers. Analyzing an oil well state allows estimating its future production as well as predicting possible equipment failures. Predictions of similar parameters may vary depending on oil well type (production wells, exploration wells or injection wells) or on geological structures involved.

An average oil well penetrates several oil layers with varying physical properties like pressure or temperature. However, these properties can be same for different wells, given the same layer is penetrated. Therefore a single model based on similar physical properties can hypothetically be built which can then be reused for multiple oil wells. This will allow reducing the amount of models needed (compared to having a unique model for each oil well), cutting costs for their creation, storing and managing.

Создание моделей

Historical data from a real oil field was used for analysis. Oil wells with most data stored were determined by means of plain statistical analysis prioritizing the most recent data (less than 5 years old). A list of 45 oil wells of each of possible 3 types (exploration, injection and producing wells) was formed. The main goal for the next steps is to detect groups of highly-correlating technological parameters to build models upon. A model would then consist of a set of parameters that correlate well for most oil wells.

Microsoft Decision Tree Algorithm diagram was used to show correlation graphically (see fig. 1) and to get its numeric representation. For each of selected oil wells a model was created. It was then trained using the values of all parameters of the well. The values of each parameter were interpolated to match the starts of equal time periods, thus forming a matrix having timestamps as its lines and technological parameters as its columns, creating a uniform grid.
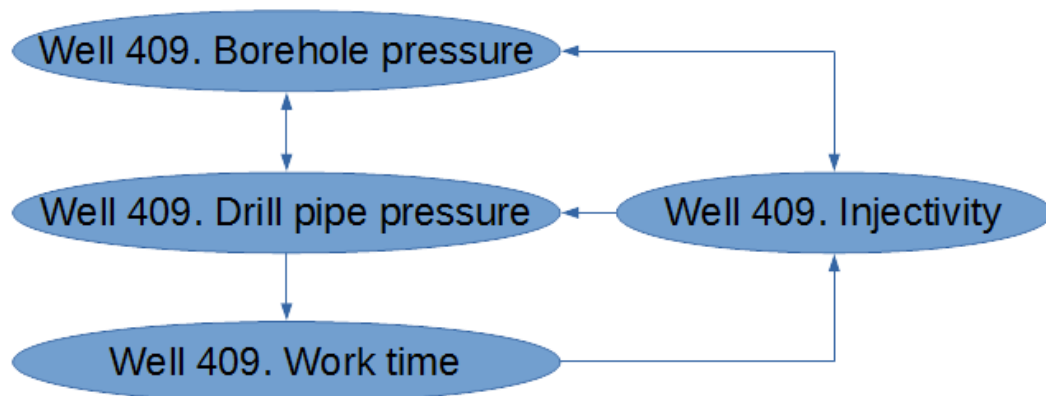

Fig. 3. Oil well parameters correlation graph

The resulting correlations contained information on correlations within separate oil wells only. To build models applicable for multiple wells at once data had to be further analyzed using a different algorithm. To do this the data was first aggregated into another matrix (essentially being combined adjacency matrix for all correlation graphs acquired in the previous step). Manual analysis of the resulting matrix, however, showed that some parameter pairs had high correlation values while being present only in minority of oil wells, which required removing them. This was done by using a classification model trained to detect whether parameter pair is present in majority of wells and not. The training set filtered using this model left only 32% of parameter pairs strongly linked for most oil wells. As a result clusters of parameters were obtained and could be used to train corresponding universal prediction models using Microsoft Time Series Algorithm.

Результаты

We have discussed and developed a technology that is able to identify groups of strongly correlating technological parameters which are be used to create prediction models suitable for most oil wells of a given oil field.

Several approaches to estimate model quality exist of which estimating mean absolute error (MAE) and mean absolute percentage errors (MAPE) were used [2]. As a result of model testing, single prediction proved to be most precise method with 3,15 % error for 969 test cases for one-day prediction. This, however, cannot be used in real-world scenario due to high amount of technological parameters each oil well has (350 on average). Suggested approach of using universal reusable parameter group models showed 8,71 % error which was considerably lower than when using linear approximation (23,1 % error). Additionally, it can be used to predict technological parameter values even when no data was present (in case of new oil wells) with error slightly lower (22,47 %) than using linear approximation for oil wells with historical data (23,1 %).

XIII Всероссийская
научно-практическая конференция
«Технологии Microsoft
в теории и практике программирования»

Заключение

The result of this work is an approach for building reusable prediction models of technological data of oil wells. Our approach consists of five steps:

Determining of oil wells that are used as training data sources

Training models for oil wells parameters and determining dependencies in technological data within separate oil wells

Classification of these dependencies into "reliable" and "unreliable". Removing unreliable dependencies

Clustering of dependencies to determine the groups of dependent technological data

Training reusable models based on these groups and estimating their quality

The prototype of intelligence data analysis system of technological data of oil production was created as the result of this work. The prediction accuracy for reusable group models was 91,29 % using mean absolute percentage error for estimation (error being 8,71 %) for 969 data samples for one day prediction compared to later obtained data.

Список литературы

1. Amorim A.L. de. J. KDD, SEMMA and CRISP-DM: A parallel overvie", Campus de Azurem, 2004.
2. Chuchueva. Mathematical Bureau. The main estimates for the accuracy of time series prediction. URL: http://www.mbureau.ru/blog/osnovnye-ocenki-tochnosti-prognozirovaniya-vremennyh-ryadov, access date: 01.06.2014