Water Saturation Prediction in the Reservoir Zone of a Gas Field using SVR Method

Introduction

Estimation of reservoir parameters like porosity, water saturation and permeability play an important role in evaluating reservoir. Therefore, it is essential to estimate the correct amount of water saturation as an important parameter for evaluating hydrocarbon reservoirs. Without knowing water saturation, estimation of reservoir and fluid distribution evaluation in reservoir is impossible and as a consequence making an appropriate plan for reservoir production and drilling has high uncertainty (Moradzadeh et al 2013). Typically in hydrocarbon reservoir water saturation is calculated by well log data and applying water saturation equations (Moradzadeh et al 2001), which has a different parameters dependant on formation structure of reservoir and can cause some errors (Archie 1942) if there is small amount of shale in it (Tiab et al 2004). There is not only water, but also oil and gas are in pores of rocks. Every error in estimation of water saturation can result in loosing beneficial zones of oil and gas and economical failure. Different methods in determining water saturation have already been used, including determining water saturation by core (Luthi 2001), using seismic data (Moradzadeh 2013) and also using seismic attributes (Moradi 2011 unpubl. results).

Method

Theory of support vector regression

Support vector regression is an application of machine learning, which has solved the curse of dimensionality. It is a phenomena in which situation during data analysis some problems can occur and the volume of the data increase rapidly (Bellman 1957 & 1961).

Assume a learning data set such \{(x_1, y_1),..., (x_l, y_l)\} as subset of \(x \times \mathbb{R}\) so that \(x\) refers to input sampling space. \(x=\mathbb{R}^d\). In support vector regression the aim is obtaining \(f(x)\) function in such a way that this function for each learning point has the most value of \(\varepsilon\) which is the difference from real amount of \(y_i\), call as loss function, and simultanously be flat as much as possible. In other words, the errors lower than \(\varepsilon\) are not considered in this circumstance, but the errors more than \(\varepsilon\) are not accepted. \(f(x)\) function as standard linear form is as below:

\[
f(x) = \langle w, x_i \rangle + b; \quad w \in \mathbb{R}, \quad b \in \mathbb{R}
\]

In Eq1 \(<.>\) refers to inner product in \(x\) space (input space). Being flat means the small amount of \(w\). To achieve this goal, the Euclidean norm should be minimized and this problem changes to convex optimization problem. In this case there is a function like \(f\) which approximate all the points with errors lower than \(\varepsilon\), otherwise there should be a tolerance for errors by slack varaibles which make us face with \(\varepsilon\)-insensitive function (Figure 1). This problem can be solved by standard dualization method. This is a method to expand support vector machines to unlinear functions and because of this standard dualization method with Lagrange multiplier can help to solve the problem. At the end, it is specified that the complexibilty of \(f\) function which is defined by support vectors are independent of the dimensions of input space of \(x\) and it only depends on the number of support vectors. It is possible to evaluate \(b\) by Karush-Kuhn-Tucker conditions. These conditions explain a point which is specified as the solution of the problem and in this point the multiplication of dualization varaibles and constraint should be zero. Making unlinear algritur is the next step. This happens by preproccesing of training points \(x_i\) by \(\mathbb{F}:X \rightarrow f\) to a feature space \(\mathbb{F}\) and then using support vector algritur. This procces can not be applied point by point in \(\mathbb{F}\) practically so a more optimized method should be found. This means that instead of applying \(\mathbb{F}\) for every point in order to take them into feature space and then obtaining inner product of them, this procces can be done in input space in format of Kernel equation. The most usefull Kernels in this method are linear and Gausian (Scholkopf 2002).
Data

The South Pars/north dome field is a natural gas condensate field, which has 105 km distance from south coast of Iran, and it is the largest independent gas field all over the world that is used commonly by Iran and Qatar (Khanmohamadi et al 2009). The gas field’s area is 9700 ($km^2$) which one third of it is for Iran and two third is for Qatar (Kahkesh et al 2003). It’s structure is the continuation of an enormous dome structure which is in the north east of Qatar and is known as north dome of Qatar. This enormous structure has north east- south west direction and gentle ridges. The main reservoir sequence of this field is constructed by shallow marine carbonate sequences of Kangan and Dalan formations in the age of middle Permian period until lower Triassic period (Najmabadi 1993). For more accurate investigations on the reach hydrocarbon formations subsurface operations applied (Iliat et al 2015), which is done by well logging. The logs which are used in this investigations are Gamma ray, Sonic, Neutron, Density, Photo electric factor, Specific resistance, MSFL, LLS and LLD logs.

Results and Discussion

The data has been modified and prepared by Petrel sofware to use in matlab. Then by using soppurt vector regression the function of water saturation is estimated. In the end, a blind test is done in such a way that 1 out of 4 wells is neglected and we train and estimate the water saturation by 3 wells. In these wells 75% of the data for training and 25% for primary examination is selected randomly. In the support vector regression method by Gaussian kernel in the primary examination the correlation between real and estimated data is obtained 74.28% and in the blind test the correlation between real and estimated data is obtained 81.07%. In the support vector regression method by Linear kernel kernel in the primary examination the correlation between real and estimated data is obtained 62.27% and in the blind test the correlation between real and estimated data is obtained 78.1%. In figure 2, two scatter plots of real and estimated water saturation by Gaussian and Linear kernels are shown. In this figure because of the bigger correlation between estimated and real data in Gausian kernel with respect to Linear, in the plot belongs to Gaussian kernel data are more closer to the line with the slope equals to one and there are less errors. In figure 3 the depth plot of water saturation between the depth of 2850 (m) to 3250 (m) for real and estimated data is shown which can be realized by which the graph belongs to the Gaussian kernel has more overlap between red and blue line.
Figure 2 scatter plot of real water saturation vs estimated water saturation, right one belongs to Gaussian kernel and left one belongs to Linear kernel.

Figure 3 Depth plot of real (red) and estimated (blue) water saturation, the right one belongs to Gaussian kernel and the left one is for Linear kernel.
It is noticeable that the loss function’s value in support vector regression method with Linear kernel is obtained 0.0517 and with Gaussian kernel is obtained 0.044.

**Conclusion**

We can say that due to porosity (if it is high) in the zone with low water saturation the probability of having reservoir is high. As a consequence in the depth between 3000 to 3200 (m) we can do coring. After modifying and preparing the data and applying SVR in Matlab and obtaining appropriate output we can conclude that Gaussian kernel shows more reliability with respect to Linear kernel in estimation of water saturation.

**References**


Kahkesh, Chehrazi, M, Roshanaeizade, Gh. [2002] Drilling reports in well number 2 of south pars gas field in NISOC.


Najmabadi, S. [1993]. Final geological report of well number 1 of south pars gas field of Persian gulf from internal report of Pars oil and gas company.


