2015 IEEE 15th International Conference on Data Mining Workshops

Production Estimation for Shale Wells with Sentiment-based Features from Geology Reports

Bin Tong, Hiroaki Ozaki, Makoto Iwayama, Yoshiyuki Kobayashi
Research & Development Group, Hitachi, Ltd., Japan
{bin.tong.hh, hiroaki.ozaki.yu}@hitachi.com
{makoto.iwayama.nw, yoshiyuki.kobayashi.gp}@hitachi.com

Sahu Anshuman, Vennelakanti Ravigopal
Big Data Laboratory, Hitachi America, Ltd., USA
anshuman.sahu@hds.com
ravigopal.vennelakanti@hds.com

***Abstract*—Shale oil and gas have become very promising unconventional energies in recent years. To optimize operations in oil and gas production, a reservoir model is important for understanding the subsurface appropriately. Generally, sensor data, such as surface seismic data, are most popular data sources in modeling the reservoir with either a numerical simulation model or an Artificial Intelligence (AI)-based model. In this paper, to obtain data that describe the subsurface more exactly, information, including phrases that indicates possible bearing oil or gas and rock colors, is extracted from geology reports. Sentiments of the phrases is identified by sentiment analysis, and sentiment sequence over measured depths is then used to generate features. The rock-color similarities between wells are calculated as well, and integrated as distance metrics into a geology-based regression method. Extensive experiments on Bakken wells in the United States show the effectiveness of using the features extracted from geology reports and the rock colors in terms of estimating well production.**

Keywords*—geology report; sentiment-based feature; production estimation

I. Introduction

Since the shale revolution in the United States, shale gas has been rapidly emerging as a significant unconventional resource, which attracts a tremendous investment. In the shale oil and gas industry, operators and service providers offer solutions for exploration, drilling, and completion before well production. However, according to a survey [1], about 40% of shale wells underperform production expected by companies. The main reasons for this underachievement are due to inaccurate exploration and inappropriate understanding on the subsurface. It is so-called reservoir modeling, an essential process to be performed in well development, that helps to make the optimal operations for well development.

Traditional reservoir modeling simulates production from a field of multiple wells as a function of the characteristics of the reservoir in question. Its functional relationships follow specific physical laws, which are non-flexible. As a result, it might not be suitable for modeling different well fields or areas. Recently, Artificial Intelligence (AI)-based reservoir modeling with big data analytics [2][3] has been received a lot of attentions owing to its various advantages. The most significant advantage is that an AI-based model is flexible in regard to changes and diversity of data. Accordingly, in this work, we focus on the AI-based model.

When modeling the reservoir, Formation Evaluation (FE) [4] is used to interpret a combination of measurements, such as gamma radiation and resistivity, taken inside a wellbore to detect and quantify oil and gas reserves. However, these downhole measurements might fail to achieve reasonable accuracies due to temperature, pressure, and/or vibration in the wellbore. A geology report, which is issued by a geologist who analyzes physical properties of sample rocks along wellbore, provides a more accurate access to the subsurface. It is one of documents used in formation evaluation to complement surface seismic data and well-logging data.

In the present work, the geology report is studied to determine how it can contribute to the FE process. We mainly focus on extracting information from the geology report, in which the geologist gives an opinion on a number of specific properties of a sample rock. The opinion is identified as a sentiment by using sentiment analysis. sentiment sequence over depths is used to generate sentiment-based features. This work is related to aspect-based sentiment summarization [5]. Most of research studies focus on reviews or texts from the web, which are posted or commented on by individuals. These reviews can therefore be considered as being independent. However, the sentiment sequence over depths is essentially different from sentiments in the reviews, because neighboring sentiments might be related to each other. It can be explained by the fact that geology change is often not dramatic in a local area of the subsurface. To the best of our knowledge, the present work is the first to apply techniques of sentiment analysis to extract information from geology reports.

As an application, the features extracted from geology reports are utilized to estimate well production. Besides these features, the similarity of rock colors taken from different wellbores is calculated. The rock-color similarity is integrated as a distance measure into a popular location-based regression method to boost estimation performance. In other words, the geology information and well production can be more effectively correlated. These correlations help on-site engineers and geologists to improve the operations involved in well completion. It should be pointed out that a large amount of research studies estimate well production [6] after the well completion. These works mainly use techniques for analyzing time-series data, such as Neural Network and Fuzzy Logics [7] [8], to predict future production based on past production. In contrast, the present work estimates production before well completion, which is right after well drilling. This production estimation is important in formation evaluation, since the estimation results can be used to evaluate hydrocarbons reservoirs. It helps to optimize production operations, such as placing

978-1-4673-8493-3/15 $31.00 © 2015 IEEE
DOI 10.1109/ICDMW.2015.13

12551 – 12580	SANDSTONE: Light to medium tan, cream tan, off white, buff, fine to very finely grained, subrounded to subangular, moderately well sorted, dolomitic, moderately well cemented with dolomitic cement, good intergranular porosity, trace dolomite, trace calcite crystals, weak petroleum odor, light brown oil staining, light green oil breaking out in sample wash water, even dull yellow green fluorescence, fair cloudy cut.
12581 – 12610	SANDSTONE: Light to medium brown, tan, buff, off white, clear, very finely grained, fine to medium grained in part, subangular to subrounded, trace pyritic, argillaceous in part, moderately cemented with dolomitic cement, trace unconsolidated quartz sand grains, good intergranular porosity, strong petroleum odor, abundant argillaceous platy siltstone, light green to dark red oil breaking out in sample wash water, even yellow green fluorescence, instant milky, streaming yellow cut.

Fig. 1. A sample geology report.

fracturing stages in zone bearing oil or gas most.

The main contributions of this work are as follows. **1**, information is extracted from geology report and analyzed by sentiment analysis. **2**, the sentiments of a number of specific aspects in a depth series are summarized as features that reflect geological changes in the subsurface. **3**, the relationship between rock-color similarity and well production is experimentally determined. A popular geology-based regression method is extended by integrating the rock-color similarity into the geological distance space.

II. Preliminary on Shale Well and Geology Report

As an unconventional energy sources, shale oil and gas production are quite different from conventional ones, such as crude oil. The main difference is the use of techniques of horizontal drilling and hydraulic fracturing. Accordingly, a wellbore consists of two sections, namely a vertical section and a horizontal section which is drilled through the shale formation.

The geology report referred to in this paper is a log of rock samples taken from a wellbore. It records information about sample rocks at each depth along the wellbore. At drilling stage, the rock samples taken along wellbore are analyzed by a geologist. After the drilling is finished, a formation evaluation is conducted to evaluate the reservoir, and the evaluation results facilitate optimizing the operations of oil and gas production, such as setting optimal locations for fracturing. A sample geology report[1], which lists measured depth of a rock sample and a description of the rock sample in pairs, is shown in Figure 1. As shown in the figure, the left side is the depth range in feet, and the right side is the rock description in each depth range. The rock description is composed of a list of phrases that depict properties of the sample rocks, such as rock color, rock texture and physical or chemical properties. Some phrases express the opinions of the geologist with respect to the properties, and those opinions indicate whether nor not the depth range covers an oil or gas bearing zone. The aim of this work is to extract these opinions from the geology report, and utilize opinion information concerning rock samples along the wellbore to estimate oil and gas production.

III. Framework of Production Estimation

We propose a framework for estimating well production, which includes three main components, namely information extraction, feature extraction and production estimation. First, phrases of interest are extracted from a preprocessed geology report; Second, features are extracted on the basis of opinion series associated with the extracted phrases; Third, well production is estimated by integrating features extracted from the geology reports with other available data sources, such as formation tops data.

The fundamental ideas about estimating well production using information from geology reports are two-fold. First, from the viewpoint of geological feature, a geology report includes a list of phrases at each depth range along a wellbore. Some phrases show opinions of the geologist with respect to a number of specific properties, and these opinions indicate the possible oil or gas bearing sections in the wellbore. These opinions are summarized into features that help to estimate well production. Second, from the viewpoint of a production-estimation model, geology-based regression models often assume that near wells are more influential on a target well than distant wells. Besides geological distance, such as trajectory distance between two wellbores, similarity of rock colors along wellbore can be regarded as another distance metric for measuring the relationship between two wells. It is assumed that rock colors indicate the geological characteristics of the rocks to a certain degree.

IV. Information Extraction

Information Extraction includes two functionalities, which are phrase extraction and rock color extraction.

A. Phrase Extraction

We extract three categories of featured phrases. i.e., oil stain, porosity, and fluorescence cut. The main reason for extracting them is that they are key rock properties that a geologist examines to evaluate the quality of a reservoir. The types of phrases are explained as follows. First, **oil stain** is a kind of "oil show", and is left on rock samples. It is an indicator for a high possibility of bearing oil. For instance, "very strong oil odor" and "rare oil staining" are positive and negative examples of oil stain, respectively. Second, **porosity** is the volume of the non-solid portion of the rock filled with fluids, divided by the total volume of the rock. If the porosity is high, oil and gas are more likely to permeate from the interior of the rock. For instance, "fine to good porosity" and "very poor porosity" are positive and negative examples of porosity, respectively. Third, **fluorescence cut** describes the following outcome: if hydrocarbons are present in the rock, they will disseminate into the solvent, giving the entire solvent a distinctive color under ultraviolet light. This sheen under UV light is called cut and the color of the cut indicates the quality of the oil. For instance, "fair streaming cut" and "little to no cut" are positive and negative examples of fluoresence cut, respectively.

The above-described phrases can be extracted from a geology report by a dictionary matching. It is possible because the rock description for each depth range, which is always divided by a separator, i.e., a comma, can be parsed into a list of phrases. In addition, a limited number of terminologies are used to describe a type of phrase. For example, a phrase about oil stain often includes keywords such as 'oil', 'stain', 'staining', and 'odor'.

[1] https://www.dmr.nd.gov/oilgas/

B. Rock Color Extraction

The target of rock-color extraction is to generate a rock-color matrix for each well. The geologist records rock colors of the rock samples taken from each depth range along the wellbore. The changes in rock colors along the wellbore reflect the changes in geological formation to a certain degree. It is assumed that near wells may have high rock color similarity due to the fact that near wells may share similar geological formations.

In the rock color extraction, we first extract colors from the description of rocks in each depth range, and generate a binary vector by matching the colors with a color dictionary, entries of which have fixed color definitions. When the value of an entry in the vector is equal to 1, it means that the color shows with rock in the depth, and vice versa. By matching all depth ranges, we can obtain a binary color matrix for each geology report. Note that a dictionary for nine rock colors, including brown, gray, yellow and cyan is defined. Therefore, the number of columns in the rock-color matrix is nine, and the number of rows is equivalent to the number of depth ranges of a wellbore.

V. Feature Extraction

The target of feature extraction is to convert the extracted phrases into sets of numerical values that can discriminate wells by degree of oil or gas bearing. To make feature extraction easy, a phrases is categorized into one sentiment label by sentiment classification. Therefore, for the phrases over depths, a sentiment label sequence along the wellbore can be obtained. Then, we extract five feature sets, namely *base* feature, *ngram* feature, *p2v* feature, *hmm* feature, and *graphsim* feature, from the sentiment label sequence.

A. Sentiment Classification

Due to diverse expressions of the phrases about porosity, oil stain, and fluorescence cut, in the sentiment analysis, sentiment labels are designed to represent different degrees of positive. In the simplest case, two labels, which are positive and negative, can be designed. If necessary, more granular labels can be also defined. For example, four labels can be designed, such as positive, weak positive, weak negative, and negative. It should be noted that a $null$ label is used in the case that a depth range has no corresponding phrase. As for feature extraction, five kinds of features, named by *base*, *ngram*, *p2v*, *hmm*, and *graphsim* are considered. The fundamental idea of feature extraction is to summarize the sentiments for the three kinds of phrases over depth range, so that the features can depict geological characteristic of the subsurface.

B. base feature

In the case of the *base* feature, the frequencies of sentiment labels for each aspect are simply calculated. However, in this case, the frequencies of sentiment labels do not hold orders of sentiment labels. It might happen that even different sentiment-label sequences result in the same frequencies of each sentiment label. It is therefore natural to consider extracting the features concerned with label transition over depths, namely *ngram* feature, *p2v* feature, *hmm* feature, and *graphsim* feature.

C. ngram feature

In contrast to the typical use of n-gram in the research field of Natural Language Processing (NLP), in which words in a sentence are considered, a sentiment label is regarded as a word in the present study. The assumption of using n-gram feature is that patterns of n-grams in wells with different well production might differ. More specifically, the frequencies of n-grams in high-production wells may significantly differ from those in low-production wells.

In the case of the $ngram$ feature, sentiment labels of three kinds of phrases are combined into an integrated label. For example, let sentiment labels be pos, neg, and pos for oil stain, porosity, and fluorescence cut, respectively. Then, the three sentiment labels are combined into one label, e.g., pos_neg_pos. This combined label in a certain depth range is taken as a word. Therefore, a well can be similarly regarded as a sentence composed of combined labels, and a set of wells can be seen as a document composed of those sentences.

Referring to the space of n-grams, it is supposed that each phrase can be classified as one of the M sentiment labels. The word space for a combined label in one depth range is M^3. The space of n-gram, namely M^{3n} can be easily derived. It is thus clear that the space of n-grams exponentially increases with the value of n. Therefore, it is necessary to choose the most discriminative n-grams for the production estimation task. The wells can be partitioned into three groups, namely high production, normal production, and low production. Then, a supervised feature selection algorithm, such as gradient boosting [9], can be used to select the most informative n-grams. The $ngram$ feature discussed in Section. VIII-B1 in details.

D. p2v feature

In the case of the *p2v* feature, a combined sentiment label for the three aspects is also regarded as a word in the label sequence. A technique of Deep Learning, named Paragraph Vector (PV) [10] is utilized to extract the features. It is an unsupervised algorithm that learns fixed-length feature representations from variable-length pieces of texts, such as sentences. This kind of representation is different from the bag-of-words model, which loses ordering of the word and ignores semantics of words. The PV model is based on either of two schemes, namely distributed memory (PV-DM) and distributed bag of words (PV-DBOW). Since PV-DM considers the word order within a sliding window over a paragraph, this scheme is used for the PV model in this study.

As for using the *p2v* feature, it is assumed that the distributed representations learned by PV reflect the semantic meaning of a label sequence, which can further discriminate the high-production wells and low-production wells. An example of semantic meaning can be the degree of positives for both the sentiment label and the sentiment label sequence. For example, a positive degree for a combined label can be simply considered as the number of pos. A detailed analysis of the *p2v* feature is discussed in Section. VIII-B2.

E. hmm feature

As for the *hmm* feature, the basic idea is to summarize the hidden states of sentiment label sequence by using a Hidden

Markov Model (HMM), which is widely used in the fields of Natural Language Processing (NLP) and Speech Processing. The hidden state is a kind of summarization of the three different labels in each depth range. From the viewpoint of a HMM model, sentiment labels over depths are taken as observations, e.g, pos_neg_pos and $null\_neg\_pos$. A HMM is used to infer transitions among the hidden states with respect to the observation sequence. The hidden states can be roughly interpreted as a status with either positive attitude or negative attitude. For example, in the case of pos_neg_pos and $null\_neg\_pos$, the hidden state of pos_neg_pos is expected to hold a more positive degree than that of $null\_neg\_pos$. It is because that pos_neg_pos has two positives in the label, while $null\_neg\_pos$ has only one positive.

After the HMM is trained, we make statistics on the hidden state sequence as the ***hmm*** feature. The statistics include the frequencies of the states and the state transition frequencies. The state transition simply means the frequencies from a state to its self or another state in the state sequence. The analysis about hmm feature is discussed in Section. VIII-B3.

F. graphsim feature

The assumption in the *graphsim* feature is that if two wells have similar well productions, the similarity between transition graphs is likely to be high. The transition graph is defined as a directed graph, in which the number of vertices is equal to the number of specified labels, and the edge from one vertex to another vertex is associated with the transition frequency between the two vertices. The label used in the *graphsim* feature is also the combined label from the three aspects. For the three aspects of phrases, each of which has three possible states, such as pos, neg, and $null$, the total number of the combined labels should be 27. However, comparison for transition graphs of 27 vertices might be sensitive to the noise in the label sequence.

In order to build up a more robust and semantic transition graph, a combined label in a certain depth range is represented in the form of the degree of positives. 7 states are designed, including pos_pos_pos, pos_pos_null, pos_pos_neg, pos_null_null, pos_null_neg, pos_neg_neg, and no_pos. The 7 states were ranked by the degree of positives. For example, pos_pos_pos means that three positives exist. pos_pos_null means that two positives and one $null$ label exist. no_pos means that there is not any positive label. It should be noticed that the new representation does not hold the order for the three aspects. For example, pos_pos_null, pos_null_pos, and $null\_pos\_pos$ are different labels. However, in the new 7 states, they are the same, because they both have two positives label and a $null$ label. Note that the labels without any positive, such as $null\_neg\_null$ and $null\_null\_null$, belong to no_pos.

We make statistics on the frequencies among these 7 states. We found that the states' frequencies themselves might not be fully capable of discriminating the wells with different production. For example, the frequencies of pos_pos_pos may be similar among all the wells. However, the ratio of states with less degree of positives is higher in the low production wells than high production wells. Therefore, five features for a well are designed on the basis of the transition graph with the new 7 states. They are listed as follows:

1) The self-transition frequencies for states with two positives.
2) The ratio of the state transition frequencies among states with at least two positives to those among states with less than two positives.
3) The ratio of the state transition frequencies among states with at least two positives to those among states without positives.
4) The ratio of the state self-transition frequencies among states with at least two positives to those among states with only one positives.
5) The ratio of the state transition frequencies among states with at least one positive to those among states without positives.

The graph-transition similarity between two wells can be represented by the similarity between five features of the two wells. More specifically, the similarities between the five features of a well and those of wells with high well production are used in estimating well production. More detailed analysis of the *graphsim* feature is discussed in Section. VIII-B4.

VI. Production Estimation

We utilize the extracted features and rock color information among wells to estimate well production. The most popular techniques for estimation is regression, which is widely used in a variety of areas, such as agriculture and economics. Since the information about geological location is available in the shale wells, we focus on geology-based regression models, such as Geographically Weighted Regression (GWR) [11]. Its fundamental idea in the context of shale well is that the production of near wells have greater influence on a target well than distant wells. This assumption applies to shale wells, in which shale wells in a drilling space unit influence production each other, called well placement problem [12].

In GWR, only geological distance is used to measure if two wells are near. Under inspiration by the work [13], GWR is extended to a spatial-color space that considers both the geological distance and the rock-color distance. The assumption of using the rock color distance is that near wells have similar geological formations, so that their rock color similarity is high. An detailed analysis of the relationship among rock-color similarity, trajectory distance, and oil production is made in Section VIII-A to support this assumption.

The extension of GWR, named GCWR, can be expressed as:

$$Y_i = \beta_0(l_i, c_i) + \sum_k \beta_k(l_i, c_i) X_{ik} \tag{1}$$

where X_{ik} represents the value of the k-th variable in sample X_i, and Y_i is the response value for sample X_i. $\beta_k(l_i, c_i)$ is a set of weights at point i. l_i and c_i represent the geological location and color information of the i-th point, respectively. $\beta_0(l_i, c_i)$ is the weight for the intercept. The estimation of $\beta_k(l_i, c_i)$ can be calculated as:

$$\widehat{\beta}_{l_i,c_i} = [XW(l_i, c_i)X]^{-1} X^T W(l_i, c_i) Y \tag{2}$$

where $W(l_i, c_i) = \text{diag}(\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in})$ and n is the number of observations. diag() is a diagonal matrix in which the entries outside the main diagonal are all zero. The diagonal elements

α_{ij} are the weights at a spatial and rock-color space, and are formulated as follows:

$$\alpha_{ij} = \exp\{-\frac{d_t(i,j)^2}{2h_t^2}\} \times \exp\{-\frac{d_c(i,j)^2}{2h_c^2}\} \quad (3)$$

where $d_t(i,j)$ represents the geological distance. In our problem, trajectory distance between two wells is used as the geological distance. This is because different wells have different shape of wellbore, and considering the distance between two wellheads might lose information about the wellbore, such as wellbore shape, and drilling direction. $d_c(i,j)$ is a term related to distance on rock color. Similar to the parameter h_t for geological distance is important for GWR, it should be noted that the parameter setting for rock color distance h_c influences to the performance. The way to set the value of h_c is discussed in Section VIII-A3

1) Trajectory Distance: To simplify the calculation of trajectory distance, we suppose the vertical section of a wellbore is truly vertical, divide the horizontal sections of two wellbores into L sections evenly, and average the distances between the $L+1$ points in the horizontal sections. It is also supposed that the vertical section of a wellbore is truly vertical and the horizontal sections of wells keep a constant difference in vertical depths. The trajectory distance between two wells can be calculated as:

$$dist_{traj}(P,Q) = \frac{1}{L+1}\sum_{k=1}^{L+1} pd(gd(L_P^i, L_Q^i), D_{PQ}) \quad (4)$$

where N represents the number of evenly divided sections in the horizontal sections of wells. L_P^i and L_Q^i are the pairs of longitude and latitude in the i-th point in the horizontal sections for wells P and Q, respectively. $gd(x,y)$ represents a geological distance, such as Haversine distance, between the two locations, i.e., x and y. D_{PQ} is the difference between vertical depth of the two wells. $pd(x,y)$ represents a distance according to the popular Pythagorean theorem, stating the square root of the summation of the squares for x and y, respectively. Since the wellbore for shale well is not only vertical, the direction of the horizontal sections of two wells significantly affects the distance between the two wells.

2) Rock Color Similarity: The distance about rock colors $d_c(i,j)$ in Eq. (3) by simply obtained by setting $1-sim_c(i,j)$, where $sim_c(i,j)$ is defined as the rock color similarity.

Due to different number of rows of rock-color matrices, rock-color similarity between two wells can be calculated by sliding a color matrix along with the other. At each comparison during the sliding, we calculate the ratio of the number of 1 matchings to the size of the smaller color matrix, and average the ratios in all comparisons to give the color similarity between the two wells. The rock-color similarity between two wells is calculated as follows:

$$sim_c(P,Q) = \frac{1}{K}\sum_{k=0}^{K}\frac{1}{s(Q)}\sum_{i,j}\mathbf{1}(P_{ij}^{k\cdot r(Q)}, Q_{ij}) \quad (5)$$

where P and Q represent the color matrices, and the number of rows of P, $r(P)$ is larger than that of Q, $r(Q)$. K is the quotient of $r(P)$ divided by $r(Q)$. $s(Q)$ denotes the number of entries in Q. $\mathbf{1}(x,y)$ is a function whose value is equal to 1 only in the case of $x=y=1$. Note that we can also choose another way of comparing two rock-color matrices.

TABLE I. SENTIMENT CLASSIFICATION IN THE CASES OF 2 CLASSES AND 4 CLASSES.

Aspect	2 classes		4 classes	
	SVM	LR	SVM	LR
stain	.9626	.9626	.8724	.8638
porosity	.9451	.9484	.8667	.8611
cut	.8586	.8542	.8399	.8171

VII. EXPERIMENTAL RESULTS

In this section, we present the experimental results of well production estimation using the information extracted from geology report. First, we explain which data is used and how data is preprocessed. Then, we show the results of the sentiment classification. Last, we discuss the results of well production estimation using the features extracted from geology report.

A. Experimental Data

We study the data of shale oil & gas wells permitted in North Dakota in the United States, which is public at a government site (see footnote 1). This website includes a huge amount of information about wells, including well summarization, well trajectory, well logging data and scanned reports.

In this experiment, we focus on three categories of information, which are well summarization information, well trajectory, and geology report in scanned reports. Well summarizations record the basic information about each well, including wellhead location, well type, total measured depth, and formation tops, etc. Note that formation tops data record depths of a number of specific formations.

We mainly use formations top data as the structured data. For scanned geology reports, we use a commercial OCR software, named Abbyy Fine Reader, together with layout analysis to extract the list of depth and rock description for each well. Since some wells have missing trajectory data, scanned geology report, and formation tops, we prepare 1764 wells for well production estimation.

B. Results of Sentiment Classification

To classify sentiments, manually labeled data is required. For each category of phrases, e.g., oil stain, porosity, or fluorescence cut, we manually label the phrases that appear more than 20 times. At the beginning, each kind of phrase is labeled by using four sentiment labels to represent degrees of sentiments, namely positive (PT), weak positive (WP), weak negative (WN), and negative (NG).

To learn the classification model, we try most popular classification methods, such as Support Vector Machine (SVM) and Logistic Regression (LR). Classification performances are evaluated by using 5×10 cross-validation and setting the best parameters for each classification method. F-score is used as criterion of classification performance. On the other hand, two sentiment labels, i.e., positive and negative, are also used. In other words, both PT and WP are regarded as PT, and both WP and NG are regarded as NG. Table I shows the classification performances in the case of two classes and four classes. It can be seen that the classification performance of each aspect in the 2-class case is reasonably good. However, four classes

can represent richer degrees of sentiment than two classes. In the trade-off between sentiment classification accuracy and number of degrees in sentiment, the former is given a higher priority, since inferior classification accuracy will have a side-effect on the following feature extraction. Therefore, in the experiment, only two sentiment labels, together with a $null$ label, are used.

C. Production Estimation Results

Oil and gas production is estimated by using formation-tops data together with features and rock-color information from geology reports. The rock-color information is only used for the GCWR method, while the features from geology reports are applicable to all regression methods. For estimating oil production, we use averaged bbls in the first 12 months, which is calculated by dividing the total bbls in the first 12 months with number of effective working days. Similarly, for estimating gas production, we use averaged mcfs in the first 12 months. Note that *bbl* and *mcf* are units for oil and natural gas volumes, respectively. In the estimation of oil and gas production, four criteria, i.e., Mean Absolute Error (MAE), Squared Mean Squared Error (RMSE), and Correlation Coefficient R, are used. According to the definitions of the criteria, lower values of MAE and RMSE represent better estimation results, while larger values of R represent better estimation results.

To clarify the effectiveness of the features extracted from geology reports, different number of combinations for the features are used together with the structured data, i.e., formation-tops data. Note that the value i in i-F ($i = 1, 2, 3, 4, 5$) denotes the number of combined features. For example, 1-F means that one feature is selected from the five to concatenate with the structured data. In addition, the dimension of each feature is optimized by maximizing the regression performance in the 1-F case. For comparison with our proposed method GCWR, we examine estimation performances of popular regression methods, such as Kernel Ridge Regression (KRR), Support Vector Regression (SVR), Ordinary Least Square (OLS), and GWR, in terms of 5×10 cross validation. For each method, parameters are tuned. In addition, we use 3-gram. The value L is set to be 30 for each well in the trajectory distance, since most of wells have 25 or 30 stages in average.

Table II lists the results of the oil and gas production estimation of each method with different number of combined features. The column labeled 'structure' represents performances estimated with structured data only, i.e., formation-tops data. The columns labeled i-F represent performances estimated by using both structured data and features from geology reports. Note that the values in this table are averages in the cross validation, each of which is the best result for each method after tuning parameters. For each feature setting, i.e., structure and i-F, we examine if the best performance is significantly better than the others. Note that in Table II, the values with bold fonts denote the best performances, and the values marked with a star show that they are not significantly inferior to the best performances at significance level of 0.1.

Four key points can be drawn from Table II. First, the performances of KRR decrease when features from geology reports are used. Second, the performances of OLS, SVR, GWR, and GCWR increase when features from geology reports are used. However, they do not always increase with increasing number of combined features. For example, OLS, SVR, GWR, and GCWR tend to get their best performances at 3-F or 4-F, and their performances degenerate if the number of features increases. Third, GWR and GCWR, which are kinds of geology-based regression methods, outperform OLS and SVR. It can be seen that the neighboring relationship measured by geological locations has an positive effect on estimating well production. Fourth, the performances of GCWR are always better than that of GWR in both oil and gas production estimation. This result indicates that the rock color information is effective and helpful in regard to improving the performance of well production estimation.

TABLE III. R VALUES OF REGRESSION METHODS WHEN USING ONE FEATURE WITH STRUCTURED DATA.

Type	Method	structure	*base*	*hmm*	*ngram*	*graphsim*	*p2v*
Oil	OLS	.3296	.3402	.3429	.3965	.3498	.3567
	SVR	.3244	.3268	.3255	.4012	.3426	.3469
	GWR	.5117	.5148	**.5105**	.5205	.5174	.5189
	GCWR	.5285	.5286	.5311	.5335	.5346	.5406
Gas	OLS	.3104	.3345	.3387	.4103	.3329	.3230
	SVR	.3287	.3430	.3391	.4286	.3492	.3414
	GWR	.4837	**.4765**	.4862	.4867	.4908	.4982
	GCWR	.4905	**.4865**	.5004	.4914	.4958	.5065

TABLE IV. COUNT OF FEATURES IN BEST-FEATURE COMBINATIONS OF i-F ($i = 1, 2, 3, 4$) FOR EACH METHOD.

Type	Method	*base*	*hmm*	*ngram*	*graphsim*	*p2v*
Oil	OLS	2	3	4	1	0
	SVR	0	2	4	3	1
	GWR	2	0	4	1	3
	GCWR	1	1	1	1	3
Gas	OLS	2	3	4	1	0
	SVR	0	3	4	2	1
	GWR	2	2	0	2	4
	GCWR	2	2	1	1	4
	ALL #	11	16	22	12	14

The effects of each feature on the performance of each method, except KRR, are investigated, since KRR does not attain better performance when the features from geology reports are integrated. First, to understand the effectiveness of combing each feature with the structured data, R values of each method in the 1-F setting are investigated. The results are listed in Table III. It is clear that in most cases, features from the geology reports improve the performance of estimation. Second, to understand the overall effectiveness of each feature, which feature contributes to which method most and the contribution of each feature to all methods are investigated. We examine the optimal combinations of features in 1-F, 2-F, 3-F, and 4-F settings for each method with respect to the metric R, and count the number of appearances of each feature in each method. According to Table IV, it is interesting to note that the $ngram$ feature seems to be more effective in regard to OLS and SVR, and the $p2v$ feature seems to be more effective in regard to GWR and GCWR.

VIII. EXPERIMENT DISCUSSIONS

A deep empirical analysis of the rock colors and the features from geology report is performed, indicating the reason that they effectively improve the performance of production estimation. For simplicity, only oil production data is used as a target variable in this analysis.

TABLE II. OIL AND GAS PRODUCTION ESTIMATION USING FEATURES EXTRACTED FROM GEOLOGY REPORTS.

Method	Metric	Oil Production						Gas Production					
		Structure	1-F	2-F	3-F	4-F	5-F	Structure	1-F	2-F	3-F	4-F	5-F
KRR	MAE	83.44	86.12	86.63	87.47	88.50	90.97	**79.42**	79.94*	81.18	82.94	84.09	86.66
	RMSE	**107.13**	110.75	112.19	112.88	113.19	115.96	**102.78**	**102.96**	**104.73**	107.26	108.50	109.85
	R	0.5282*	0.4758	0.4621	0.4498	0.4365	0.3902	**0.5276**	**0.5244**	0.5031*	0.4756	0.4558	0.4168
OLS	MAE	95.14	91.71	91.29	91.27	91.38	91.46	90.38	85.80	85.56	85.74	85.94	86.37
	RMSE	119.03	116.28	115.75	115.79	115.90	116.10	115.30	111.40	110.69	110.70	110.88	111.21
	R	0.3296	0.3965	0.4036	0.4058	0.4045	0.4029	0.3104	0.4103	0.4204	0.4231	0.4208	0.4173
SVR	MAE	93.73	89.75	89.59	89.65	89.70	89.70	87.59	82.72	82.68	82.73	82.80	83.10
	RMSE	120.61	116.46	116.42	116.53	116.67	116.94	116.43	111.20	111.10	111.20	111.24	111.47
	R	0.3244	0.4012	0.4042	0.4019	0.3999	0.3965	0.3287	0.4286	0.4294	0.4299	0.4292	0.4274
GWR	MAE	83.62	83.17	82.57	83.03	83.40	84.04	81.70	80.74	80.10	80.69	80.65	81.19*
	RMSE	108.59	107.70*	107.07*	107.28*	107.80*	108.59*	107.38	106.59	105.93*	106.30*	106.54*	106.96*
	R	0.5117	0.5205	0.5289	0.5281*	0.5238	0.5156	0.4837	0.4982	0.4996*	0.5093*	0.5145*	0.5049*
GCWR	MAE	**82.56**	**81.65**	**81.33**	**81.81**	**82.13**	**83.00**	80.13*	**79.30**	**78.97**	**79.22**	**79.46**	**80.23**
	RMSE	107.26*	**106.41**	**106.09**	**106.71**	**107.04**	**108.18**	107.23	106.02	105.55*	**105.40**	**106.09**	**106.89**
	R	**0.5285**	**0.5406**	**0.5438**	**0.5378**	**0.5368**	**0.5265**	0.4905	0.5065*	**0.5092**	**0.5235**	**0.5230**	**0.5075**

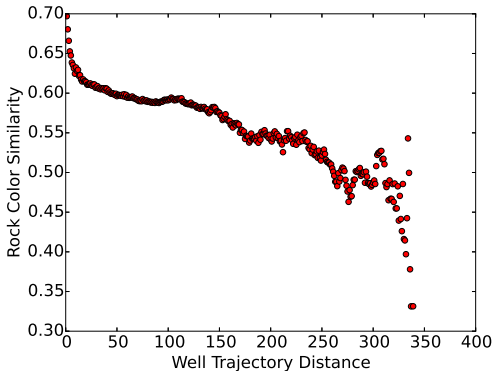


Fig. 2. Relationship between rock-color similarity and trajectory distance.

A. Analysis of Rock Colors

Rock color was experimentally analyzed. Two relationships are examined, namely that between rock-color similarity and trajectory distance, and that between rock-color similarity and oil production. These two relationships implicitly clarify the reason of integrating the rock-color similarities between wells into the GWR model.

1) Relationship between rock-color similarity and trajectory distance: The relationship between rock-color similarity and trajectory distance is investigated as follows. For each pair of wells, we calculate the trajectory distance from Eq. (4) and the rock color similarity from Eq. (5). Their relationship is shown in Figure 2, in which the x-axis is the trajectory distance between two wells, and the y-axis is the color similarity between two wells. We divide the trajectory distance range $(0, 400]$ into equal bins with size of 1 km. Each point in Figure 2 is the averaged value of rock-color similarities among the wells if the trajectory distance between two wells fall into the bins. It can be seen from Figure 2 that when the trajectory distance between two wells increases, their corresponding rock-color similarity decreases. It is noted that the rock-color similarity almost ends around 0.5, meaning that the rock colors of the two wells have little correlation. This is because, as shown in Eq. (5), if the value in two color matrices are randomly either 0 or 1, the similarity between the two matrices should be statistically around 0.5.

2) Relationship between rock-color similarity and oil production: The assumption behind this relationship is that if the rock color similarity between two wells is high, oil productions of the two wells are similar. To clarify this relationship, the following calculation is performed: first, given a well and a distance threshold th_d, find neighbors of the well if the trajectory distance between a neighboring well and the given well is shorter than th_d; second, calculate the rock color similarity between the given well and the neighboring well; third, calculate the production change ratio from

$$r_{prod}(t, n) = \frac{|prod_t - prod_n|}{prod_t} \tag{6}$$

where $prod_t$ is oil production of the given well, and $prod_n$ is oil production of its neighbor well. It should be noted that if $prod_t$ is equal to $prod_n$, the oil-production-change ratio $r_{prod}(t, n)$ will be 0; fourth, divide $[0, 1]$ of the rock-color similarity space into equal bins with size of 0.001, and average the production change ratios if the color similarity between two wells fall into the range of a bin. Set th_d to be 20 km. The relationship between rock-color similarity and oil-production-change ratio for different distance thresholds is shown in Figure 3. It is clear that when rock-color similarity between two wells is higher than around 0.4, the oil-production-change ratio decreases. This result implicitly validates our assumption that two wells that have the similar rock color may have similar oil production. This validation is reasonable, because rock colors reflect geological formations of the subsurface, and similar geological formations may result in similar oil production.

3) Analysis of parameter behavior about rock colors in GCWR: As shown in Eq. (3), the weight between two wells is determined by two terms, which are related to trajectory distance and rock-color similarity. Through the above analysis of the rock colors, it is known that if two wells have high rock-color similarity, their oil productions are similar. This means that if the rock-color similarity between two wells approaches 1, the second term on the right side of Eq. (3) is close to 1; if the rock-color similarity between two wells approaches 0, the second term on the right side of Eq. (3) is close to 0. Based on this observation, the behavior of h_c in the GCWR model from another viewpoint is explained. It is supposed that, if the rock-color similarity sim_c is larger than a threshold th_c, the second term on the right side of Eq. (3) gives a weight in the range of $(\gamma, 1]$ where $\gamma \in (0, 1)$; otherwise, the term directly

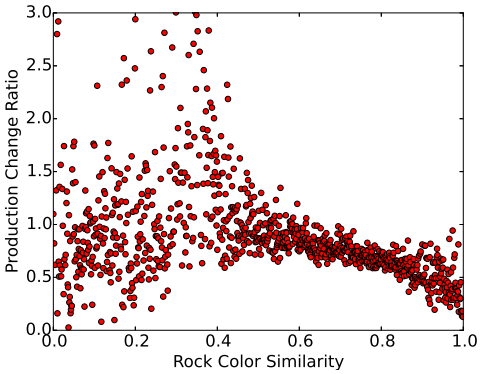


Fig. 3. Relationship between rock-color similarity and oil production.

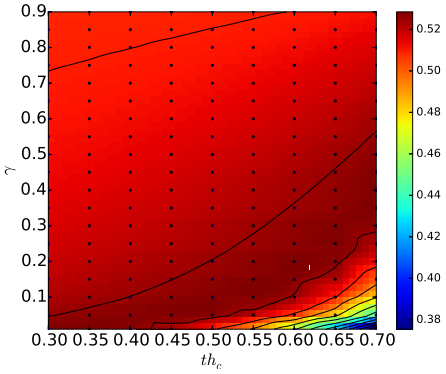


Fig. 4. R changes with respect to both th_c and γ.

gives a zero weight.

R changes with respect to both th_c and γ are shown in Figure 4. It can be seen that when th_c is lower than around 0.4, R tends to get the best performances when γ is lower than 0.1. When th_c increases from 0.4, R tends to achieve the best performances while γ is increasing. The reason might be that setting a low value of γ and a high value of th_c tends to weaken the effect of the high rock-color similarity.

How rock-color similarity and trajectory distance affect α_{ij} in Eq. (3) is explained in the following typical three cases. First, two near wells, w_p and w_q, have the same trajectory distance with a target well w_t, and the color similarity between w_p and w_t, $sim_c(p,t)$, is larger than that between w_q and w_t, $sim_c(q,t)$, where $sim_c(p,t) > sim_c(q,t) > th_c$. In this case, it follows $\alpha_{at} > \alpha_{bt}$. Second, if $th_c > sim_c(p,t) > sim_c(q,t)$, $\alpha_{pt} \simeq \alpha_{qt} \simeq 0$ holds even if the first term on the right side of Eq. (3) is larger than 0. Third, if well w_m is very distant from the target well, w_t, and $sim_c(m,t) > th_c$, $\alpha_{mt} \simeq 0$ may hold because the first term on the right side of Eq. (3) approaches 0 due to the large distance.

B. Analysis of Features

In this subsection, we make analysis of the *ngram* feature, *p2v* feature, *hmm* feature, and *graphsim* feature

TABLE V. KL DIVERGENCE ON FREQUENCIES DISTRIBUTIONS AMONG CATEGORIES.

	all 3-grams			selected 3-grams			positive degrees		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
1st	0.	.384	.596	0.	.137	.282	0.	.032	.099
2nd	.453	0.	.229	.111	0.	.046	.029	0.	.019
3rd	.480	0.136	0.	.224	.043	0.	.083	.017	0.

TABLE VI. CLASSIFICATION PERFORMANCES WITH DIFFERENT DIMENSIONS OF THE SELECTED 3-GRAMS AND 5-GRAMS.

	3-grams		5-grams	
Dim	SVM	LR	SVM	LR
10	.6115	.6219	.4490	.4791
30	.6794	.6809	.6905	.6920
50	**.7097**	.7043	.6976	**.7087**
70	.7016	.6936	.6586	.6535
100	.6715	.6670	.6520	.6421

1) Analysis of ngram feature: In this analysis, we investigate if the *ngram* feature is able to discriminate wells with different oil production. 3-grams is used as an example to explain. Wells are partitioned into three groups, namely wells with high oil production, normal oil production, and low oil production. Oil production is normalized into range $[0, 1]$, and frequencies of wells for each bin with size of 0.1 are counted. It is found that more than one-third of wells have oil production in range $[0.1, 0.2]$. Therefore, it is reasonable to make the following three partitions. If oil productions of three wells fall into ranges $[0, 0.1)$, $[0.1, 0.4)$, and $[0.4, 1]$, those wells are regarded as a low-production well, a normal-production well, and a high-production, respectively.

Three forms related to 3-grams are considered, namely full 3-grams, selected 3 grams, and positive degree. Full 3-grams are all possible 3-grams. For each form, we calculate the frequencies. Selected 3-grams are the 3-grams that are most discriminative for the three groups of wells. Gradient boosting [9] is used to select 64 3-grams. Positive degree for a 3-gram is defined as an ordered tuple, and each of element is the number of pos in a label. For example, in the case of 3-gram, such as (pos_pos_neg, neg_pos_neg, $null\_neg\_pos$), its positive degree is an ordered tuple $(2, 1, 1)$. The positive degree for a 3-gram provides a way to explain the meaning of a 3-gram. Note that for a 3-gram, it can be easily shown that 64 possible positive degrees exist, because each entry in the tuple has 4 possible values, namely 3, 2, 1, and 0.

To compare differences between frequency distributions of three groups for each form, KL divergence is used as the criterion. As listed in Table V, it can be seen that the difference between the frequency distribution of the three groups are obvious in all three forms. The selected 3-grams have smaller difference in frequency distribution of the $1st$ and the $3rd$ groups than that in full 3-grams. However, they are able to reduce the number of feature dimensions to a large extent. It is interesting to note that the positive degrees for the full 3-grams can even discriminate the three groups as well.

The optimal n for the $ngram$ feature is experimentally investigated. The wells are partitioned into two groups with equal size, and two popular classification methods, i.e., SVM and LR, are used to determined the classification performances in terms of F-score. Table VI shows the performances of the selected n-grams ($n = \{3, 5\}$) at different dimensions that changes from 10 to 100. It can be seen that the performances

TABLE VII. CLASSIFICATION PERFORMANCES OF *p2v* FEATURE. (D, W) IS THE PAIR COMPOSED OF A DIMENSION AND A WINDOW SIZE.

	2 classes		2 class variant	
(D, W)	SVM	LR	SVM	LR
(10, 25)	.6705	.4751	.6723	.4874
(15, 5)	.6707	.5587	.7611	.5593
(20, 5)	.6826	.5795	.7568	.5641
(20, 10)	**.7879**	**.6180**	**.8393**	**.6261**

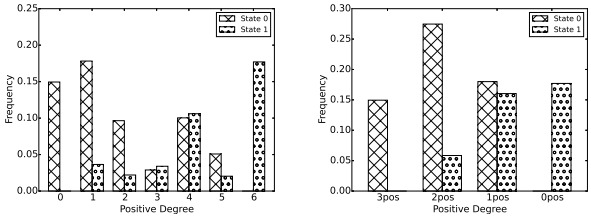
of 3-grams are better than those of 5-grams in most cases.

2) Analysis of p2v feature: It is found that the PV model is even able to learn sentiment of labels. We regard the positive degrees in the *gramsim* feature as a kind of sentiment. We calculate the similarity between pos_pos_pos and other positive degrees. For example, the similarity between two positive degrees, such as pos_pos_pos and pos_pos_null, is the averaged similarities between the label pos_pos_pos and the labels with 2 positives and one $null$ label, such as pos_pos_null, pos_null_pos, $null\_pos\_pos$. In the PV model, the window size and the number of dimension are two important parameters. The two parameters are tuned, and the positive degrees are ranked by descending the similarities with respect to pos_pos_pos. The order is determined as follows: pos_pos_null, pos_pos_neg, pos_null_neg, pos_null_null, pos_neg_neg, and no_pos, when the window size is set to be 10 and the number of dimension is set to be 20. It can be inferred that the sentiment information is implicitly embedded in the representation vector learned by the PV model.

Classification performance using the *p2v* feature is also evaluated. Two cases are considered, namely 2-class and 2-class variant. In the case of 2-class, the wells are partitioned into two groups with equal size based on oil production. In the case of 2-class variant, the wells are partitioned into three groups with equal size, which are low production, normal production, and high production. Only two groups are kept, i.e., low production and high production. The dimensions of the learned feature are tuned in the range $[10, 15, 20]$, and window size is tuned in the range $[5, 10, 15, 20, 25]$. The classification results of the two cases with different pairs of dimension and window size are listed in Table VII when the ranking for the positive degrees with respect to pos_pos_pos is good enough. It can be seen from Table VII that the performances in the case of 2 class variant are better than those in the case of 2 class. It indicates that the *p2v* feature can discriminate wells with different oil productions.

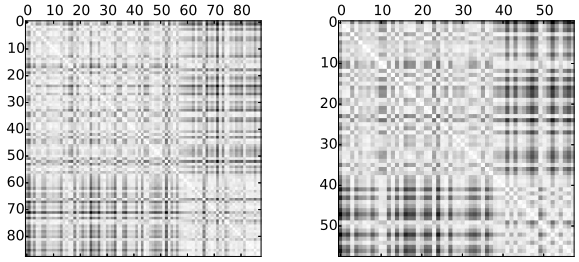
3) Analysis of hmm feature: In this analysis, we examine if the states learned by the HMM are capable of reflecting positive degree of each label. Two types of positive degree are used. The first type is the 7 positive degrees used in the *graphsim* features, namely pos_pos_pos, pos_pos_null, pos_pos_neg, pos_null_null, pos_null_neg, pos_neg_neg, and no_pos. The second type is at a more abstract level. It simply count the number of pos for each label, which are three positives, two positives, one positive, no positive.

In HMM, the number of states is set to be 2, which is often the optimal states for regression methods in Section VII. We first get the positive degree for each observation, which is a integrated label. The frequencies of positive degrees for each state is then calculated according to the state sequence. As shown in Figure 5, it can be seen that the frequencies of positive degrees implies the meaning of the two states. For example, the state 0 represents the labels with more positive degrees, while the state 1 represents the labels with less positive degrees. Note that the x-axis in Figure VIII-B3 is the index of each positive degree. For example, the index 0 means pos_pos_pos and the index 6 means no_pos. It can be seen that the HMM can learn the state that depicts the degree of positive for the label derived from the three aspects, namely oil stain, porosity, and fluorescence cut.



(a) Seven positive degrees (b) Four positive degrees

Fig. 5. Frequency distributions of positive degrees for states.



(a) size=20 (b) size=30

Fig. 6. Similarity graph with 1-norm distance for different window sizes.

4) Analysis of graphsim feature: Similarity between the five features is used to represent the similarity between label transition graph. In the similarity matrix, the degree of gray represents the similarity. For example, the white point shows the high similarity, while black point shows low similarity. To draw the similarity matrix, the wells are ordered in descending oil production. To focus on the trend of similarity change, we use a simple smoothing technique that averages the features within different size of windows. Figure 6 shows the similarity matrix with different window sizes using 1-norm distance as a metric. Two points can be drawn from Figure 6. First, when the size of the window increases, the patterns of white areas become clearer. Second, when the size set to be 30, the two white areas indicate that the similarities among the averaged features can differentiate wells with different oil productions. Therefore, we compare the five features of a well with averaged five features in top d windows, and use their similarities as the *graphsim* feature for the well. Note that d is the number of dimension for the *graphsim* feature.

IX. Related Work

Our research work is mainly related to sentiment analysis [14] [15] [16]. In recent years, techniques of sentiment analysis has been widely used in analyzing various kinds of documents, such as texts in social media [17] and product reviews [18]. Brendan et al. [19] analyzed the sentiment of tweets, and discovered a correlation between the sentiment of the tweets and public opinion pools. Kushal et al. [18] studied opinion extraction and classification for product reviews by identifying appropriate features and scoring methods. As far as it is known, sentiment analysis for geology reports has not been studied.

Sentiment or opinion summarization is also related to our work, since our concern is to summarize a sentiment sequence into a set of numerical values that show geological characteristics. The traditional methods used for sentiment summarization can be mainly divided into two categories, including multi-document text summarization [21] and aspect-based summarization [5][22]. The aspect-based summarization is similar to the $base$ feature used in our study. In the example of camera reviews, its basic idea is to identify different production attributes of cameras, such as picture quality and size, collect the text in the reviews about these specific attributes, and classify the reviews into positive or negative labels. The ratios of positive and negative labels can be easily illustrated for each aspect.

However, the previous techniques in sentiment analysis can not be applied to our problem. It is because a geology report is a depth-series data, and is processed into a list of sentiment label over depths. Compared with the stock price prediction using the time-series sentiments, production prediction for each depth range using the sentiment sequence does not make sense from the application viewpoint. In addition, sentiment summarization for geology reports is essentially different that for product reviews. The sentiment change over depths in geology reports reflects a geological change in the subsurface. The sentiments in the local areas in the subsurface is related to each other, while sentiments from product reviews are almost independent of individuals. The numerical features extracted by the sentiment summarization or aggregation are expected to reflect the geological characteristic of the well, which facilitates the further analysis of wells, such as well clustering and estimation of well production.

On the other hand, production estimation is related to geology-based prediction. Fotheringham et al. [11] proposed a prediction model based on geological locations, named Geographically Weighted Regression (GWR). This prediction model is a weighted version of a typical linear regression model, in which the weights of samples are influenced by the geological distances. It has been applied to various areas, such as crime analysis and estimation of agriculture yield. Huang et al. [13] proposed a prediction model that considers distances in both spatial and temporal spaces. This model produces a promising prediction accuracy of house prices over time. In our work, the GWR model is extended to a spatial-color space, which measures both the trajectory distance and the rock-color distance between wells.

X. Conclusion

In this paper, we propose a framework for well production estimation by using information extracted from geology report. The sentiments of the phrases in the geology reports are extracted and then summarized into features. The rock-color similarity among wellbores is used as a distance metric to be integrated into a geology-based regression method. Extensive experiments show the effectiveness of the sentiment-based features and the rock-color similarity in the well production estimation.

References

[1] A. Scollard., "Petrel Shale-Engineered for Exploiting Shale Resources," in *URTec*, 2014.

[2] S. D. Mohaghegh, "Reservoir Simulation and Modeling Based on Pattern Recognition," in *SPE*, 2011.

[3] K. Holdaway, *Harness Oil and Gas Big Data with Analytics: Optimize Exploration and Production with Data Driven Models*. Wiley, 2014.

[4] S. Pirson, *Handbook of Well Log Analysis for Oil and Gas Formation Evaluation*. Prentice-Hall, Inc.,Englewood Cliffs, NJ, 1963.

[5] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," in *KDD*, pp. 168–177, 2004.

[6] W. W. Weiss, R. S. Balch, and B. A. Stubbs, " How Artificial Intelligence Methods Can Forecast Oil Production," in *SPE*, 2002.

[7] S. Al-Fattah and R. Startzman, "Predicting Natural Gas Production Using Artificial Neural Network," in *SPE*, 2002.

[8] R. Balch, D. Hart, W. Weiss, and R. Broadhead, "Regional Data Analysis to Better Predict Drilling Success: Brushy Canyon Formation, Delaware Basin, New Mexico," in *SPE*, 2002.

[9] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, pp. 1189–1232, 2000.

[10] Q. V. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," in *ICML*, 2014.

[11] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, 2002.

[12] W. Bangerth, H. Klie, M. F. Wheeler, P. L. Stoffa, and M. K. Sen, "On Optimization Algorithms for the Reservoir Oil Well Placement Problem," *Computational Geosciences*, vol. 10, pp. 303–319, 2006.

[13] B. Huang, B. Wu, and M. Barry, "Geographically and Temporally Weighted Likelihood Regression for Modeling Spatio-temporal variation in House Prices," *International Journal of Geographical Information Science*, vol. 24, no. 3, pp. 383–401, 2010.

[14] H. D. Kim, G. K. A., S. Parikshit, and Z. ChengXiang, "Comprehensive Review of Opinion Summarization," *Technical report, University of Illinois at Urbana-Champaign*, 2011.

[15] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[16] E. Cambria, B. W. Schuller, B. Liu, H. Wang, and C. Havasi, "Knowledge-Based Approaches to Concept-Level Sentiment Analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 12–14, 2013.

[17] A. Bakliwal, J. Foster, J. van der Puil, R. O'Brien, L. Tounsi, and M. Hughes, "Sentiment Analysis of Political Tweets: Towards an Accurate Classifier," in *Proceedings of the Workshop on Language Analysis in Social Media*, pp. 49–58, 2013.

[18] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," in *WWW*, pp. 519–528, 2003.

[19] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," in *ICWSM*, 2010.

[20] J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1 – 8, 2011.

[21] K. Lerman, S. Blair-Goldensohn, and R. McDonald, "Sentiment Summarization: Evaluating and Learning User Preferences," in *EACL*, pp. 514–522, 2009.

[22] S. Blair-goldensohn, T. Neylon, K. Hannan, G. A. Reis, R. Mcdonald, and J. Reynar, "Building a Sentiment Summarizer for Local Service Reviews," in *WWW workshop on NLP in the Information Explosion Era*, 2008.