Automatic Velocity Picking Using Unsupervised Ensemble Learning

Hongtao Wang, Jiangshe Zhang*, Chunxia Zhang, Zixiang Zhao, Chengli Tan, Zhiyu Yang, Weifeng Geng

Abstract—In seismic data processing, accurate and efficient automatic velocity picking algorithms can significantly accelerate the processing, and the main branch is to use velocity spectra for velocity pickup. Recently, machine learning algorithms have been widely used in automatic spectrum picking. Even though deep learning methods can address the problem well in supervised cases, they are often accompanied by expensive computational costs and low interpretability. On the contrast, unsupervised learning methods based on the physical knowledge have great potential to efficiently resolve the task. In this paper, we propose an unsupervised ensemble learning (UEL) method to pick the root mean square (RMS) velocities on the spectrum. In particular, UEL utilizes the information of nearby velocity spectra and the nearest seed velocity curve to assist the selection of effective and reasonable velocity points. To increase the coherence of energy peaks, an information gain method is developed by local normalization. In addition, we designed the attention scale-space filter (ASSF) clustering method to incorporate the coherence information into the picking process. Experiments on three datasets demonstrate that compared to traditional clustering methods, UEL can recognize energy clusters better, especially with smaller blobs. Moreover, the injection of nearby spectra and interval velocity constraint in UEL significantly improves the robustness and accuracy of picking results.

Index Terms—velocity analysis, ensemble learning, clustering

I. INTRODUCTION

VELOCITY analysis in seismic data is critical to seismic inverse problems. In recent years, most of velocity analysis approaches were based on seismic velocity spectrum [1]. The velocity spectrum is a graph of velocity and zero-offset time domain, built by a mapping from the common midpoint (CMP) gather to the spectrum signal coherence with some discriminant criterion (e.g. semblance amplitude criterion) [2]. The high signal coherence points on the velocity spectrum can be regarded as a reflection. Therefore, velocity picking in velocity spectrum can be regarded as a reflection. Therefore, the points which are filtered by the threshold of amplitude are clustered in velocity as well as zero-offset time domain, and the centers of the clusters are regarded as energy peaks. The application of clustering algorithms includes: $K$-means clustering [4], Density-Based Spatial Clustering of Applications with Noise(DBSCAN) algorithm [9], Gaussian Mixture Model (GMM) [10], etc.

Regarding optimization-based methods [3] [4], the objective function is generally chosen empirically and is hard to unify. Even in different geological conditions, the chosen objective functions are different. For the supervised learning approaches, the deep learning models [5] [6] [7] [8] can achieve high picking accuracy. But as far as the current pickup model is concerned, its training tag velocity spectrum energy peak is a huge project and the trained model is difficult to generalize to other geological conditions. However, unsupervised learning approaches have less parameters, and the key parameter is the number of clusters which can be controlled simply, e.g. $K$ in $K$-means, GMM [9], etc. Whereas, current unsupervised learning methods to solve this problem neglected the coherence values of the velocity spectrum and regarded the cluster centers as the energy peaks directly [9]. When true cluster points have different weights, clusters will not be an appropriate estimation of energy peaks. Therefore, we have to consider the coherence information to correct the bias and take a few physical constraints to ensure the picking quality. Moreover, in the low signal noise ratio (SNR) velocity spectra,
multiple noises have great influence on the clustering results. In this case, it is not enough to simply limit the threshold for denoising, and further denoising work is needed.

To solve the above problem, we propose an unsupervised ensemble learning (UEL) method to estimate the root-mean-square (RMS) velocity from a velocity spectrum. In particular, our UEL fuses the information of the current spectrum, the near spectra and a few labeled spectra to solve the velocity analysis problem. The use of additional other information can guide the current spectrum to filter out outliers. To pick the suitable blob centers from the single spectrum, we also propose the attention scale-space filter (ASSF) clustering method. Compared with the base model SSF [11], ASSF makes better use of the coherence information, and pays more attention to the points with large coherence values. Moreover, ensemble learning can improve the robustness of picking results very well, and even on the median SNR dataset, UEL still produces stable picking results.

II. METHODOLOGY

The automatic velocity picking method we proposed is composed of three main parts, as shown in Fig. 1. First, we perform information gain method on the current velocity spectrum. Then, we use the ASSF method to cluster the gain results, and the cluster centers is regarded as the preliminary velocity estimation. Next, we select the near velocity spectra according to the location information of the current spectrum. After gaining the near velocity spectra, the reference velocity is estimated by our proposed low-frequency information extraction method. Finally, we fuse the information of the current velocity spectrum, the near velocity spectra and the nearest seed velocity curve in ensemble learning to obtain the RMS velocity picking of the current spectrum.

A. Velocity Spectrum and Gain Method

The velocity spectrum is generated by performing NMO correction on the CMP gathers (Fig. 2a) through different constant velocities, and then using some criteria to evaluate the coherence [12]. In our work, we choose the semblance method to compute the coherence, using the following function:

\[ C = \frac{\left( \sum_{i=1}^{N} a_i \right)^2}{N \sum_{i=1}^{N} a_i^2}, \]  

where \( a_i \) is the amplitude value in the NMO gather, and \( N \) is the sample number of the amplitude. This method measures the correlation between the parts of the NMO gather and the constants to measure the coherence.

As Fig. 2b shown, the signal strength of the reflection event is different, which leads to the inconsistency of the coherence scale of these energy clusters in the seismic velocity spectrum. To balance the scale, we propose the velocity spectrum gain method using local normalization (LN). We assume that the shape of spectrum value matrix is \( H \times W \). LN method divides each element in the matrix by the local mean, using formula Eq. 2.

\[
C_{i,j}^*(K) = \begin{cases} 
C_{i,j} \cdot \left( \frac{1}{K} \sum_{h=1}^{K} C_{h,j} \right), & 0 < i \leq K \\
\frac{C_{i,j} \cdot (1+2K)}{K} \sum_{h=1}^{K} C_{h,j}, & K < i \leq H - K \\
\frac{C_{i,j} \cdot (H+K-i+1)}{H} \sum_{h=1}^{H-K} C_{h,j}, & H - K < i \leq H 
\end{cases}
\]

As Fig. 2c shown, our gain method balances the global scale very well. The energy blobs of the shallow and deep layers are easier to be observed. Not only that, our gain method does not change the location of the local extreme, which ensures the accuracy of the later automatic picking algorithm.

B. Near Spectra Reference Velocity Picking Method

In the 3D-field seismic data, the geological layers are continuous in space, so the near velocity spectra can also provide the current velocity spectrum with some velocity prior

---

**Fig. 1.** The workflow of our methodology.
information. The reference velocity picking method of the near spectra is composed of four main parts as shown in Fig. 1.

First, we define the velocity spectra satisfying the following two situations as the near velocity spectra:

- The same line number and near CDP numbers;
- The same CDP number and near line numbers.

Then, the near spectra also are gained by the above gain method.

Second, we use a low-pass filter to obtain the low-frequency information of each near spectrum, as Fig. 3b shows. In this paper, we choose the average blur to filter spectra, and the kernel is defined as follows:

\[
K = \frac{1}{w \times w} \begin{bmatrix}
1 & \cdots & 1 \\
\vdots & 1 & \vdots \\
1 & \cdots & 1
\end{bmatrix}
\]  

(3)

where \(w\) is the width of the kernel.

Next, these low-frequency feature maps of the near spectra are stacked to get the common information by the Eq. 4:

\[
S_{i,j}^* = \sum_{k=1}^{K} S_{i,j}^{(k)} / K,
\]

(4)

where \(S_{i,j}^*\) is the element of the stacked feature map and \(S_{i,j}^{(k)}\) is the element of the k-th low-frequency feature map. In Fig. 3d, before estimating the velocity curve, the velocity spectrum is transformed into a three-dimensional point set \(\{(t_i, v_i, c_i) : i = 1, \ldots, N\}\) by using a splitting threshold, and \(c_i\) is the coherence value of point \(i\).

Finally, the reasonable reference velocity is estimated by the attention locally weighted linear regression (ALWLR). Our proposed ALWLR is an improved algorithm of Locally weighted linear regression (LWLR), which is a non-parametric algorithm, and is always used to solve the underfitting problem.

In each local regression of LWLR at time \(t_0\), the loss function is given as follows:

\[
\text{loss}(t_0) = \sum_i w_i (v_i - t_i \cdot \alpha - \beta)^2,
\]

(5)

with

\[
w_i = e^{-\frac{|v_i - t_i|^2}{2h^2}},
\]

(6)

where \(\alpha, \beta\) are the regression parameters, \(t_i, v_i\) are the value of time dimension, velocity dimension at point \(i\), respectively, and \(h\) is the bandwidth of the Gaussian kernel. LWLR uses the local sample location information at the prediction point \(t_0\) to construct a weighted kernel (Eq. 5). While in ALWLR, the weighted kernel pays more attention to the coherence value of each point, as following:

\[
w_i = c_i^\lambda e^{-\frac{|v_i - t_i|^2}{2h^2}},
\]

(7)

where \(c_i\) is the value of coherence dimension at point \(i\) and \(\lambda\) is the weight hyper-parameter between two parts. The implement of ALWLR is shown in Algorithm 1.

**Algorithm 1: ALWLR Algorithm**

**Input:** samples: \(\{(t_1, v_1, c_1)\}_{1=1}^n\);

predicted t set: \(\{t_j\}_{j=1}^m\);

bandwidth of Gaussian kernel: \(h\);

coherece weight: \(\lambda\)

**Output:** predicted velocity: \(\{v_j^*\}_{j=1}^m\)

for \(j = 1:m\) do

\[
w_j = c_j^\lambda e^{-\frac{|v_j^* - t_j|^2}{2h^2}}, \quad i = 1, \ldots, n;
\]

\[
W_j = \text{diag}(w_1, \ldots, w_n);
\]

\[
X_j = \begin{bmatrix}
\begin{bmatrix}
t_1 & \cdots & t_n
\end{bmatrix}^T \\
1 & \cdots & 1
\end{bmatrix};
\]

\[
Y_j = [v_1, \ldots, v_n]^T;
\]

\[
x_j^* = [t_j^*, 1]^T;
\]

\[
v_j^* = x_j^* \cdot (X_j^T W_j X_j)^{-1} X_j^T W_j Y_j;
\]

end

return \(\{v_j^*\}_{j=1}^m\)

As Fig. 3d shown, the velocity estimation of ALWLR considers both the location information as well as the coherence information, and the true RMS velocity points are all in the neighborhood of the curve. This velocity estimation provides the following ensemble method with a good velocity prior information.

**C. Cluster by Attention Scale-Space Filter**

When manually picking up the velocity spectrum, the analyst generally finds the energy centers to be selected first. We hope to imitate the process by which the human eye perceives things to select energy centers. Scale-space filter (SSF) clustering method is a clustering method based on scale space theory, and can find suitable cluster centers in a scale space [11]. Thus, we select SSF method as the base model of our clustering method. The points in the velocity spectrum have additional attribute values, and we want to fuse...
the coherence information into the clustering model, so we propose attention SSF (ASSF) clustering method.

In the scale space theory, there are two important concepts. We take a two-dimensional image as an example to interpret. One is the initial probability density function (PDF) of the samples $p(x) : \mathbb{R}^2 \mapsto \mathbb{R}$. The other is the PDF of the samples at a scale $P(x, \sigma)$, and $\sigma$ is the scale parameter. The scale of observation describes the change in the focal length of the human eye when a person observes an object. For example, when looking closely at close range, $\sigma$ is 0, and when the human eye is looking away from the object, things become blurred and $\sigma$ starts to get larger. $P(x, \sigma)$ is usually represented by the convolution of $p(x)$ with the Gaussian kernel:

$$P(x, \sigma) = p(x) * g(x, \sigma) = \int p(x - y) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{|y|^2}{2\sigma^2}} dy,$$

(8)

where $\sigma$ is the scale parameter, and $g(x, \sigma)$ is the Gaussian kernel as follows:

$$g(x, \sigma) = \frac{1}{(\sigma \sqrt{2\pi})^2} e^{-\frac{|x|^2}{2\sigma^2}}.$$

(9)

Based on the scale space theory, SSF derives the class-center iterative convergence formula at a specific scale $\sigma$:

$$x(n+1) = \sum_{i=1}^{N} x_i e^{-\frac{||x(n) - x_i||^2}{2\sigma^2}} / \sum_{i=1}^{N} e^{-\frac{||x(n) - x_i||^2}{2\sigma^2}},$$

(10)

Eq. (10) is a weighted shift processing, and the shift weight of $x_i$ is based on the Gaussian kernel of the Euclidean distance between $x_i$ and $x(n)$. When the center points converge (Eq. (10)), the nearer center points are merged, and the merging result is used as the cluster center under the scale $\sigma$. The change rate of the scale parameter $\sigma$ is consistent with the blurring process observed by the human eye, which is 0.029 [11].

In ASSF, we pay more attention to the coherence information, and add it into the model assumption. We assume that $p(x_i) \propto c_i$, and $c_i$ is the coherence value of $x_i$. Then, the new class-center iterative convergence formula is as follows:

$$x(n+1) = \sum_{i=1}^{N} x_i c_i e^{-\frac{||x(n) - x_i||^2}{2\sigma^2}} / \sum_{i=1}^{N} c_i e^{-\frac{||x(n) - x_i||^2}{2\sigma^2}},$$

(11)

where $\alpha$ is the weight parameter between the Gaussian kernel and coherence kernel. To confirm the suitable cluster number, we define the lifetime of each cluster number, and its concept is consistent with hierarchical clustering. In ASSF, the lifetime is the longest $\sigma$ iteration time when the cluster number does not change.

In the specific algorithm implementation, we set the minimum number of the centers to ensure the picking number. Moreover, the merging condition and the convergence condition of the centers are set according to geophysical knowledge. When the data is not standardized, the merging threshold is usually set between 100-200, and the convergence threshold is usually set between 10-50. The specific algorithm implementation is shown in Algorithm 2.

D. Ensemble Learning for Velocity Picking

The clustering method can only detect some candidate points like RMS velocity estimation. To ensure the accuracy of automatic picking result, we propose a new unsupervised ensemble learning (UEL) method to fuse the information of the near spectra and the seed velocity with the picking result of ASSF.

**Fig. 3.** Near spectra reference velocity picking method. (a) The gained near spectra. (b) Low-frequency feature maps. (c) Stacking feature map. (d) The visual result of ALWLR.
Algorithm 2: ASSF Clustering Algorithm

Input: samples \( \{x_i(t_i, v_i), c_i\}_{i=1}^N \), initial scale parameter \( \sigma_0 \), bandwidth of Gaussian kernel \( h \), center minimum number \( K \), merging threshold \( T_m \), convergence threshold \( T_c \), coherence weight \( \alpha \).

Output: cluster center set \( C_{final} \)

1. Initialize center set \( C = \{c_0\}_{i=1}^n = \{x_i\}_{i=1}^n \); Initialize the set of center set \( \mathcal{C} = \{C\} \); Initialize the scale parameter \( \sigma = \sigma_0 \);
2. while center number \( > K \) do
   3. \( k = 1, C_0 = C; \)
   4. while \( \min_j \left\{ \left\| c^{k-1}_j - c^k_j \right\|_2 \right\} > T_c \) do
      5. \( c^{k+1}_j = \frac{\sum_{i=1}^N x_i c_{ij} e^{-\|x_i-c_{ij}\|^2 / 2 \sigma^2}}{\sum_{i=1}^N c_{ij} e^{-\|x_i-c_{ij}\|^2 / 2 \sigma^2}} ; \)
      6. \( N = k \)
   7. end
   8. In \( C_N \), merge the points whose pairwise distance \( < T_m \), and get new center set \( C'_N \); Let \( \sigma = 1.029 \sigma, C = C'_N \); Add \( C_N \) into \( \mathcal{C} \)
9. end
10. Search the set \( C_{final} \) which has maximum lifetime;
11. return \( C_{final} \)

First, we define the seed velocity. As Fig. 4 shows, in a 3-D field data, we sample evenly on the space at a certain sampling rate, and note these samples as seed points. The RMS velocities of these seed point velocity spectra are picked up manually in advance. According to the continuity of the stratum, we can use the manual picking of the seed point closest to the current velocity spectrum as the reference velocity.

Then, we construct a confidence area on the current spectrum based on the reference velocity curve of the near seed points \( V_{RS} \) and the reference velocity curve of near spectra \( V_{RN} \). The velocity spectrum can be defined as a two-dimensional point set:

\[
S = \{(t, v) : t \in (T_{min}, T_{max}), v \in (V_{min}, V_{max})\},
\]

where \( T_{min}, T_{max}, V_{min} \) and \( V_{max} \) are the minimum and maximum of the \( t \) domain and \( v \) domain, respectively. We set the confidence parameter \( w \) to control the range of the confidence area, and use the following function to define the confidence area (CA) on the spectrum:

\[
CA = \{x : D(x, V_{RS}) < w, D(x, V_{RN}) < w, x \in S\},
\]

where \( D(x, c) \) is the distance between the point \( x \) and the curve \( c \). Next, the CA can guide the cluster centers obtained from ASSF \( C_1 \) to select reasonable centers. Specifically, we choose the center of the set \( C_1 \) that falls in the CA area, and denote them as \( C_2 \).

Finally, we use interval velocity constraint to remove abnormal points of \( C_2 \). The interval velocity can be computed by RMS Velocities using Dix Formula [14], as follows:

\[
V_i^{int} = \left( \frac{V_{2t_n}^2 - V_{2t_{n-1}}^2}{t_n - t_{n-1}} \right)^{1/2},
\]

where \( V_n, t_n \) and \( V_{n-1}, t_{n-1} \) are the RMS velocity and the time at the \( n \)-th and \((n-1)\)-th picking points, and \( V_i^{int} \) is the interval velocity between \( t_n \) and \( t_{n-1} \). Actually, it is not reasonable for all auto-picking points to perform the interval velocity constraint. There is a min interval time between the auto-picking points. In our work, we set 250ms as the min interval time. When the time interval of two adjacent picking points is below 250ms, one point should be removed. The implement of interval velocity constraint actually is a recursive algorithm, as shown in Algorithm 3.

E. Quality Control

To evaluate the accuracy and robustness of picked RMS velocities, we perform the quality control both quantitatively and qualitatively. Quantitatively, we define the velocity mean absolute error (VMAE) to measure the difference between two discrete RMS velocities curves, as follows:

\[
VMAE(C_1, C_2) = \frac{1}{N} \sum_{i=1}^{N} \left| V_i^1 - V_i^2 \right|,
\]

where \( C_k = \{(t_j, v_j)\}_{i=1}^{N_k}, k = 1, 2 \) is linearly interpolated by the discrete RMS velocity picking point sets \( P_k = \{(t_j, v_j)\}_{j=1}^{N_k}, k = 1, 2 \).

Then, to evaluate the performance on the synthetic seismic data, we define the picking rate (PR) and the mean deviation (MD) of the real RMS velocity points. We consider that the real RMS velocity point is recognized, if there is a picking point in the neighborhood of the real RMS velocity point. Thus, the definition of picking rate (PR) is as follows:

\[
PR(\tau) = \frac{\sum_{i} I\left\{ \min_j \{||p_j - r_i||\} < \tau \right\}}{N},
\]

where \( \tau \) is the recognition threshold, \( \{r_i\}_{i=1}^N \) is the set of the real RMS velocity points, \( \{p_j\}_{j=1}^M \) is the set of the picking RMS velocity point and \( I(\cdot) \) is the indicative function. Next,
**Algorithm 3: Interval Velocity Constraint**

**Input:** Picking result $C_2$, seed reference velocity $V_{RS}$, near spectra reference velocity $V_{RN}$, the range of interval velocity $(V_{min}, V_{max})$, minimum time interval $T_m$

**Output:** Ensemble picking set $C_3$

Initialize $C_3 = C_2$

Sort $C_E = \{C_E^j(t_i, v_i)\}_{i=1}^M$ by time domain;

while True do

// $\#(\cdot)$: count element number
$N = \#(C_E), k = 0$;

for $i = 1 : (N - 1)$ do

  // Remove one point
  if $t_{i+1} - t_i < T_m$ then
    Remove the point $C_j^E$ that is far from both $V_{RS}, V_{RN}$:
    $C_E = C_E - \{C_j^E\}$;
    $k = 1$;
    break // Break from For
  end

  // interval velocity constraint
  if $\text{IntV} \notin \text{IntervalVel}(C_i^E, C_{i+1}^E)$ then
    Remove the point $C_j^E$ that is far from both $V_{RS}, V_{RN}$:
    $C_E = C_E - \{C_j^E\}$;
    $k = 1$;
    break // Break from For
  end

end

if $k = 0$ then
  $C_3 = C_E$;
  break // Break from While
end

return Ensemble picking set $C_3$

we also measure the average deviation of picking among the recognized real RMS velocity points. Thus, we define the mean deviation (MD), as follows:

$$\text{MD}(\tau) = \frac{\sum_j^N \min \{ \| p_j - r^*_j \| \}}{N^*}, \quad (17)$$

where $\{r^*_j\}_{j=1}^{N^*}$ is the recognized real RMS velocity set under the recognition threshold $\tau$, $\{p_j\}_{j=1}^N$ is the set of the picking RMS velocity points.

Qualitatively, to measure the rationality of picked RMS velocity, we use the flatness of the NMO gather which is corrected by RMS velocities, and the clarity of layer structure on the stack section of NMO gathers.

## III. Experiment

### A. Dataset

We generate a high SNR synthetic dataset, a median SNR synthetic dataset, and choose a median SNR field dataset in China to evaluate our methodology. The basic information of each dataset is shown in Tab. I, and all of them are 3-D field datasets. The synthetic datasets are generated by the certain RMS velocity field. In the generation of super gathers, we add noise RMS velocity points in the velocity model and Gaussian noise in every single trace.

| Dataset | Sample Number | Sample Range | SNR |
|---------|---------------|--------------|-----|
| S1      | 26            | 43 1085 99   | 0.20:6960 1300:20:5500  high |
| S2      | 24            | 36 650 52    | 0.20:6980 1400:20:7820  median |
| A       | 52            | 88 4453 99   | 0.20:6960 1300:20:5500  median |

### B. Result

We design three experiments to evaluate our methodology. The first experiment verify the effectiveness of the spectrum gain method, and test its improvement on different clustering method. Then, we evaluate the necessary and accuracy of the near spectra velocity reference. Finally, we evaluate the effectiveness of the UEL.

1) The effectiveness of the spectrum gain method: to verify the effectiveness of the spectrum gain algorithm and the improvement of the ASSF method, we only use the clustering method to pick up the velocity spectrum in dataset S1 with and without gain. Moreover, we also compare ASSF with other clustering methods, such as: Mean Shift [15], K-Means [16] GMM [17], DBSCAN [18], and SSF [11], etc. Since the real RMS velocity points of the synthetic dataset are known, we utilize metrics PR (Eq. [16] and MD (Eq. [17]) to evaluate the quality of picking. As shown in Tab. II, after the gain, the picking effect of each clustering algorithm has been significantly improved. Notably, the picking results of ASSF are significantly improved compared to other clustering algorithms. The performance improvement comes from the addition of coherence information in the ASSF. ASSF can recognize the small energy clusters, and the cluster centers are basically on the energy peak, shown in Fig. 6.

| Clustering Method | No Gain | Gain |
|-------------------|---------|------|
| MD PR              | MD PR   | MD PR |
| Mean Shift         | 111.76  | 97.32%    | 107.97 | 97.70%    |
| K-Means            | 164.91  | 67.72%    | 153.18 | 73.54%    |
| GMM                | 154.40  | 65.75%    | 143.88 | 70.14%    |
| DBSCAN             | 54.12   | 97.94%    | 50.68  | 98.82%    |
| SSF                | 64.94   | 96.91%    | 50.68  | 98.82%    |
| ASSF               | 64.94   | 96.91%    | 45.91  | 98.82%    |

### C. Clustering Methods on Dataset S1

**TABLE II**

| Clustering Method | MD PR | MD PR |
|-------------------|-------|-------|
| Mean Shift         | 111.76| 97.32%| 107.97| 97.70%  |
| K-Means            | 164.91| 67.72%| 153.18| 73.54%  |
| GMM                | 154.40| 65.75%| 143.88| 70.14%  |
| DBSCAN             | 54.12 | 97.94%| 50.68 | 98.82%  |
| SSF                | 64.94 | 96.91%| 50.68 | 98.82%  |
| ASSF               | 64.94 | 96.91%| 45.91 | 98.82%  |
2) Necessity analysis of the near spectra velocity prior: we use the near spectra reference velocity picking method to get the reference velocity on the datasets S1, S2 and A, and compute the VMAE between the reference velocity and the real/manual velocity. As shown in Tab. III, the near spectra reference velocity can be a good prior for the high SNR dataset S1. Even on the median SNR datasets S2 and A, the means of VMAE are both below 100m/s. The minima of VMAE are 38.59m/s and 19.06m/s, and there are still a few reasonable reference velocities. Considering the worst case, the maximum VMAE does not exceed 200ms. To be more conservative, we believe that most reliable pickup points are within 300ms of the reference velocity curve.

3) The effectiveness of the UEL: finally, we test UEL on the datasets S1, S2, and A. In Fig. 5 although the SNR of the deep layer is relatively low, the ASSF method can also ensure accuracy and stability of the pickup result according to the supplement of adjacent speed information. In Fig. 5c and Fig. 5f, the RMS velocity which is picked by ASSF can flatten the CMP gather precisely after NMO correction. We also verify the accuracy of ASSF quantitatively as Tab. IV shown. On datasets S1 and A, the mean VMAE is below 80m/s, and even on S1 its VMAE is 45.89m/s. It means there is only two pixel mean error on the spectrum image. Compared with other clustering methods, our ASSF has accurate and robust results, because of the high picking rate and the low VMAE on these three datasets.

In Fig. 6 we use the automatic picking by ASSF and manual picking to generate the stack section of NMO CMP gather. On the overall formation structure of stack section, the automatic results of UEL are comparable to manual results.
TABLE IV  
COMPARISON OF ASSF AND OTHER CLUSTERING METHODS IN UEL

| Clustering Method | S1 PR | S1 VMAE | S2 PR | S2 VMAE | A PR | A VMAE |
|-------------------|-------|---------|-------|---------|------|--------|
| Mean Shift        | 0.88  | 76.78   | 0.68  | 155.27  | 75.25|        |
| K-Means           | 0.66  | 107.21  | 0.5   | 157.3   | 80.61|        |
| GMM               | 0.64  | 104.23  | 0.5   | 155.03  | 86.54|        |
| DBSCAN            | 0.89  | 78      | 0.68  | 150.93  | 93.2 |        |
| SSF               | 0.89  | 70.63   | 0.68  | 151.29  | 99.76|        |
| ASSF              | 0.89  | 45.89   | 0.7   | 98.80   | 60.50|        |

(a) Line2560AP  (b) Line2560MP  (c) Line2840AP  (d) Line2840MP

Fig. 7. The stack section of NMO CMP gathers on dataset A.

IV. CONCLUSION

In our work, we present an unsupervised ensemble learning method for automatically picking RMS velocity from the velocity spectrum. Through experiments, we verify the following four conclusions. First, the velocity spectrum gain method is based on local equalization, can effectively extract weak signal information. Then, low-frequency information of near velocity spectra can effectively guide the auto-picking problem in the single velocity spectrum. Next, the ASSF cluster method can effectively simulate the process of manually picking up the velocity spectrum and ensure the picking rate well. Finally, our proposed UEL achieves high picking accuracy and even can maintain an average deviation of 60m/s on the field dataset, which is a good result for the unsupervised method.

In future work, we also plan to deem bringing the idea of using near spectra velocity prior into supervised learning to solve the problem of velocity analysis with low SNR. We will consider the velocity analysis as the multi-modal feature fusion problem in deep learning, and fuse a variety of information so that the RMS velocities can be estimated more accurately.

REFERENCES

[1] M. T. Taner and F. Koehler, “Velocity spectra—digital computer derivation applications of velocity functions,” Geophysics, vol. 34, no. 6, pp. 859–881, 1969.
[2] O. Yilmaz, Seismic data analysis: Processing, inversion, and interpretation of seismic data. Society of exploration geophysicists, 2001.
[3] K.-Y. Huang, K.-J. Chen, and J.-R. Yang, “Genetic algorithm for seismic velocity picking,” in The 2013 International Joint Conference on Neural Networks (IJCNN). IEEE, 2013, pp. 1–8.
[4] K.-Y. Huang and J.-R. Yang, “Seismic velocity picking by hopfield neural network,” in 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016, pp. 3190–3193.
[5] H. Zhang, P. Zhu, Y. Gu, and X. Li, “Automatic velocity picking based on deep learning,” in SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, 2019, pp. 2604–2608.
[6] Y. Ma, X. Ji, T. W. Fei, and Y. Luo, “Automatic velocity picking with convolutional neural networks,” in SEG Technical Program Expanded Abstracts 2018. Society of Exploration Geophysicists, 2018, pp. 2066–2070.
[7] R. S. Ferreira, D. A. Oliveira, D. G. Semin, and S. Zaytsev, “Automatic velocity analysis using a hybrid regression approach with convolutional neural networks,” IEEE Transactions on Geoscience and Remote Sensing, 2020.
[8] M. J. Park and M. D. Sacchi, “Automatic velocity analysis using convolutional neural network and transfer learning,” Geophysics, vol. 85, no. 1, pp. V33–V43, 2020.
[9] U. bin Waheed, S. Al-Zahrani, and S. M. Hanafi, “Machine learning algorithms for automatic velocity picking: K-means vs. dbscan,” in SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists, 2019, pp. 5110–5114.
[10] P. Zhang and W. Lu, “Automatic time-domain velocity estimation based on an accelerated clustering method,” Geophysics, vol. 81, no. 4, pp. U13–U23, 2016.
[11] Y. Leung, J.-S. Zhang, and Z.-B. Xu, “Clustering by scale-space filtering,” IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 12, pp. 1396–1410, 2000.
[12] M. T. Taner and F. Koehler, “Velocity spectra—digital computer derivation applications of velocity functions,” Geophysics, vol. 34, no. 6, pp. 859–881, 1969.
[13] W. K. Härdle, M. Mller, S. Sperlich, and A. Werwatz, Nonparametric and semiparametric models. Springer Science & Business Media, 2004.
[14] C. H. Dix, “Seismic velocities from surface measurements,” Geophysics, vol. 20, no. 1, pp. 68–86, 1955.
[15] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 5, pp. 603–619, 2002.
[16] J. A. Hartigan and M. A. Wong, “Algorithm as 136: A k-means clustering algorithm,” Journal of the royal statistical society, series c (applied statistics), vol. 28, no. 1, pp. 100–108, 1979.
[17] C. M. Bishop, Pattern recognition and machine learning, springer, 2006.
[18] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al., “A density-based algorithm for discovering clusters in large spatial databases with noise,” in kdd, vol. 96, no. 34, 1996, pp. 226–231.

Hongtao Wang is currently pursuing the master’s degree in statistics with the School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an, China. His research interests include Bayesian statistics and deep learning.

Jiangshe Zhang was born in 1962. He received the M.S. and Ph.D. degrees in applied mathematics from Xi’an Jiaotong University, Xi’an, China, in 1987 and 1993, respectively, where he is currently a Professor with the Department of Statistics. He has authored and co-authored one monograph and over 80 conference and journal publications on robust clustering, optimization, short-term load forecasting for the electric power system, and remote sensing image processing.

His current research interests include Bayesian statistics, global optimization, ensemble learning, and deep learning.
Chunxia Zhang received her Ph.D degree in Applied Mathematics from Xi’an Jiaotong University, Xi’an, China, in 2010. Currently, she is a Professor in School of Mathematics and Statistics at Xi’an Jiaotong University. She has authored and coauthored about 30 journal papers on ensemble learning techniques, nonparametric regression, etc. Her main interests are in the area of ensemble learning, variable selection and deep learning.

Zixiang Zhao is currently pursuing the Ph.D. degree in statistics with the School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an, China. His research interests include computer vision, deep learning and low-level vision.

Chengli Tan received the B.S. degree in information and computing science and the M.S. degree in statistics from Xi’an Jiaotong University, Xi’an, China, in 2014 and 2017, where he is now pursuing the Ph.D. degree. His current research interests include adversarial learning, Bayesian nonparametrics, and stochastic optimization.

Zhiyu Yang received his master degree from the Department of Computer in North China Electric Power University in 2002. Currently, he is a senior engineer in the Geophysical Technology Research Center of Bureau of Geophysical Prospecting (BGP). He is mainly engaged in research and development work of processing system. His current research interests are algorithm optimization and high performance computing.

Weifeng Geng received his master degree from the College of Geophysics in China University of Petroleum-Beijing in 2002. Currently, he is a senior engineer in the Geophysical Technology Research Center of Bureau of Geophysical Prospecting (BGP). He has been engaged in research of seismic data processing for a long time. His main research interests are VSP (vertical seismic profile) data processing, seismic velocity analysis and high resolution processing.