Introduction

In seismic exploration, first-break picking is the task of determining, given a set of seismic traces, the onsets of the first signal arrivals as accurately as possible. The accurate determination of the first arrivals onset is needed for tomographic imaging, static corrections, velocity analysis, AVO analysis, and geological interpretation. However, the volume of seismic data to be processed and the inconsistency of picks by different people or even by the same person at different times make manual picking a very time consuming and sometimes subjective process. Therefore, developing automatic picking algorithms is important for accelerating the process and making it more consistent.

Automatic first arrival picking methods are always preferred and a variety of these kind of techniques are explored over the decades. Most traditional picking methods can be roughly classified in two categories: sliding window methods and coherence methods (Sabbinone and Velis, 2010). With the extensive application of artificial intelligence technology, on the basis of traditional neural networks (ANN), the emergence of deep learning has opened a new era in the field of seismic data analysis. Recently, several authors have used convolutional neural networks as a classifier, to decide whether there is a first arrival signal or it is a poor picking (Mezyk and Malinowsk, 2018; Yuan, 2018; Loginov, 2019). However, supervised learning requires a lot of labels and spend some time training network. Therefore, Chen (2017) used an unsupervised learning algorithm to automatically classify the seismic signals into several categories according to the characteristics directly, which can not only complete first arrival picking work, but also provide labels for supervised learning. Ma et al. (2019) proposed a novel automatic first arrival picking method by combining the energy ratio method and reinforcement learning, but it lacks detailed descriptions of the reward function and initial state selection, and it is difficult to adapt to complex waveforms.

In this abstract, we formulate the first-break picking as constrained Markov decision processes (CMDPs) in feature space. By designing reasonable actions and transition probability matrices constrained by image segmentation operators (Luo, 2018), the improved method introduces the lateral continuity of the recorded data and reduces the difficulty of selecting the initial state to ensure the picking of first breaks is more accurate and automatic. Testing this approach on field seismic data reveals its properties and shows it can accurately do the automatic seismic events picking, especially the adjacent complex waveforms.

Theory

Markov decision processes

Markov decision processes are defined as controlled stochastic processes satisfying the Markov property and assigning reward values to state transitions. Formally, they are described by the 5-tuple \((S, A, T, P, r)\) where (Sigaud, 2013):

- \(S\) is the state space in which the process’ evolution takes place;
- \(A\) is the set of all possible actions which control the state dynamics;
- \(T\) is the set of time steps where decisions need to be made;
- \(P\) denotes the state transition probability function;
- \(r\) provides the reward function defined on state transitions.

\[
\begin{align*}
S & \\
\rightarrow & \\
P_{ss'}=p(s'|s,a) & \\
\rightarrow & \\
l & \\
r(s,a) & \\
\rightarrow & \\
S' & \\
\end{align*}
\]

\textit{Figure 1 Markov decision process (Sigaud, 2013).}
Figure 1 represents an MDP, drawn as an influence diagram. At every time step \( t \) in \( T \), action \( a_t \) is applied in the current state \( s \), affecting the process in its transition \( p(s'|s,a) \) to the next state \( s' \). Reward \( r \) is then obtained for this transition.

Markov decision processes allow us to model the state evolution dynamics of a stochastic system when this system is controlled by an agent choosing and applying the actions at \( a_t \) every time step \( t \). The procedure of choosing such actions is called an action policy, and is written as \( \pi \). For a policy \( \pi \), we define the value function \( V^\pi(s) \) links any starting state \( s \) to the value of the discounted criterion,

\[
\forall s \in S, \quad V^\pi(s) = E^\pi \left[ \sum_{t=0}^{\infty} \gamma^t r_t | S = s \right] = \sum_{a \in A} \pi(a | s) \left( r^a_s + \gamma \sum_{s' \in S} P^a_{s,s'} V^\pi(s') \right), \quad 0 \leq \gamma \leq 1
\]

Where \( \pi(a | s) \) represents the probability of performing an action in state \( s \). Each state \( s \) has such a \( \pi(a | s) \), and all \( \pi(a | s) \) form the overall policy \( \pi \). \( P^a_{s,s'} \) is probability of transition to state \( s' \) by performing action \( a \) in state \( s' \), and \( r^a_s \) is reward value after transition. The \( \gamma \) factor is discount factor which guarantees that the farther away the reward value has less influence, thus automatically avoiding outliers and miss-picks associated with bad or dead traces. The Markov decision process is to find a suitable policy that can produce the maximum cumulative reward value,

\[
V^*_{\gamma}(s) = \max_{\pi} V^\pi_{\gamma}(s)
\]

Value Iteration is the most common method used to solve MDPs. It relies on the direct resolution of the Bellman optimality equation (1). Detailed explanations are given by Sigaud (2013).

**Designing 5-tuple for first-breaks picking**

To formalize the first-breaks picking into a MDP, we need to identify the suitable 5-tuple \((S,A,T,P,r)\). Starting from the state \( s_0 \), we should find a policy \( \pi \) to ensure that maximizing the total cumulative reward. We set the state \( S \) to be all possible positions in the shot gather. For example, in 2D case, the state is defined as the index \( s_y = (t_y, x_y) \) in the shot gather, where \( t_y \) and \( x_y \) are the time and offset indexes, respectively. Ma et al. (2019) selected the far-offset first-arrival position as the initial state and performed global optimization upwards, but when the SNR of the seismic data is relatively low, there will be problems with selecting the initial state. Therefore, as shown in Table 1, the downward optimization strategy is adopted to avoid the difficulty of seed point selection.

The goal is to automatically track the first-arrival onsets by maximizing the expected total reward. It is necessary to use a well-chosen reward function that gives appropriate “hints” to the algorithm. In order to generate the attributes with high relevance to the first-arrivals, the seismic data needs to be analyzed to construct a feature space containing multiple attributes. For first-breaks picking, its energy attribute are generally the main feature attribute, so we can define the immediate reward:

\[
r(s_y) = \sum_{m=1}^{M} \alpha_m f_m(s_y)
\]

Equation (3) shows that the reward value function can be composed of multiple attributes. \( f(s_y) \), represents some characteristic attribute, \( \alpha \) is weighting factor.
Numerical examples

We illustrate our method using the field dataset from the towed-streamer. This is a 227-trace shot gather with offsetting from 8.75 km to 14.4 km and 4100 sampling points with 2 ms sampling interval shown in Figure 2a. As the offset increases, the energy of first arrival wave decays rapidly, which leads to the energy of direct wave much stronger than that of first arrival wave. Figure 2b shows the reward function calculated with Table 1. The automatic picking result is shown in Figure 2b and Figure 2c. The red dot represents the results of the traditional MDPs based on the initial state position indicated by the red arrow and Table 1, and the green dot is STA / LTA method. With the traditional MDPs, we observe that the resulting picks follow most of the high reward places and avoid the direct-wave energy and outliers due to the discounting mechanism.

Figure 2 (a) The field dataset (from the towed-streamer); (b) the auto-picking result overlaid with the reward function; (c) the auto-picking result overlaid with the shot gather. (green dot: STA / LTA, red dot: the traditional MDPs with the Table 1)

The second example shows a typical marine shot gather from the OBC (Figure 3a). As shown in the yellow box in Figure 3a, there is a refraction arrival with strong energy close to the first breaks. If we adopt the traditional MDPs with the Table 1, the auto-picking result will follow reward places and ignore the weak first breaks as shown in the green dot in the yellow box in Figure 3b. The improved method designs a constrained MDP, as shown in Table 2. The probability $\pi^\pi$ of performing an action in state $s$ is constrained by the structural operator $\delta(s) \cdot \delta(s)$ obtained by image segmentation (Luo, 2018). The red dot shows the results of the first-arrivals using the improved method proposed in this paper. Compare the red and green dot in Figures 3b and 3c, we propose the method combines the advantages of the traditional MDPs and image segmentation, and can do the automatic first-breaks picking more accurately due to the edge of the image.

Table 1 five-tuple in the traditional MDPs

<table>
<thead>
<tr>
<th>state $S$ / $T$</th>
<th>action $A$</th>
<th>transition probability $P^s_{s',a}$</th>
<th>reward function $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i = (i,x_i)$</td>
<td>$A = {\text{Left}; \text{Right}; \text{Up}; \text{Down}}$</td>
<td>$P^s_{s',a} = {P^s_{w^s}; P^s_{w^s}; P^s_{w^s}; P^s_{w^s}}$</td>
<td>$r(s_i) = f_i(s_i)$</td>
</tr>
<tr>
<td>$T_i = {i,...,x_i}$</td>
<td>$\pi^\pi = \pi^\pi = \pi^\pi = \pi^\pi$</td>
<td>$= (10%; 10%; 0%; 80%)$</td>
<td>$= \sum_{j=1}^{n_x} d_j^2$</td>
</tr>
</tbody>
</table>

Table 2 five-tuple in the constrained MDPs

<table>
<thead>
<tr>
<th>state $S$ / $T$</th>
<th>action $A$</th>
<th>transition probability $P^\pi_{s',a}$</th>
<th>reward function $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i = (i,x_i)$</td>
<td>$A = {\text{Left}; \text{Right}; \text{Up}; \text{Down}}$</td>
<td>$P^\pi_{s',a} = {P^\pi_{w^\pi}; P^\pi_{w^\pi}; P^\pi_{w^\pi}; P^\pi_{w^\pi}}$</td>
<td>$r(s_i) = f_i(s_i)$</td>
</tr>
<tr>
<td>$T_i = {i,...,x_i}$</td>
<td>$\pi^\pi = \pi^\pi = \pi^\pi = \pi^\pi$</td>
<td>$= (10%; 10%; 0%; 80%)$</td>
<td>$= \sum_{j=1}^{n_x} d_j^2$</td>
</tr>
</tbody>
</table>
Conclusions

We proposed an unsupervised learning algorithm for picking first breaks based on Markov decision processes (MDPs) under the constraint of image segmentation methods. This method, on the basis of structure-constrained action $\pi^a$ and probability transition matrix $P^a$, can consider the lateral spatial variation of seismic data and reduce the difficulty of selecting the initial state to ensure the picking of first breaks is more accurate and automatic. Test results on field data show that accurate picks can be obtained automatically even in the presence of coherent noise and weak-energy first arrivals.

Acknowledgements

The authors thank the sponsors of WPI group for their financial support and helps. WPI’s research works are also financially supported by National Key R&D Program of China (2018YFA0702503), National Natural Science Foundation of China (41774126), National Science and Technology Major Project of China (2016ZX05024-001, 2016ZX05006-002). All the results in this paper are drawn in Madagascar open-source software (www.ahay.org).

References