Unsupervised machine learning as a velocity picking assistant

Charles Ursenbach and Hassan Odhwani, TGS*

Summary

Our objective is a novel application of unsupervised machine learning (UML) to semblance-based velocity picking. Earlier methods have used UML to identify semblance maxima within a single panel. The proposed method creates groups of related semblance peaks from different CDP locations, so that users can work interactively with whole groups that span the entire survey, rather than with individual picks. This allows the user to stay in control of the outcome, while delegating much of the tedious labor to the UML algorithm.

Introduction

Velocity picking is an obvious candidate for machine learning application in seismic processing. As a manual procedure it can be tedious and repetitive. Automatic pickers have been employed with varying degrees of success; they can identify many reasonable picks but may require QC and editing because of the difficulty for a black-box algorithm to capture the insight of an experienced processor looking at real data. Machine learning methods may fare better than non-learning automatic pickers because complete knowledge of the problem is not required to be programmed into the software; rather, it can be acquired as needed during application.

In the case of supervised learning methods, additional knowledge of the problem is added in the training stage, which requires extensive labeled data, thus not fully solving the problem of a human processor's involvement in tedious and repetitive work for a new project. This approach will not be discussed further in this abstract. Instead, we explore a novel application of unsupervised machine learning (UML) to velocity picking. Previous work along these lines has used UML to detect picks within a single semblance panel (Smith, 2017; Chen, 2018; Wei et al., 2018; bin Waheed et al., 2019). In these methods, points with a semblance value above some threshold group themselves into clusters, with each cluster typically centered about a semblance peak. Each cluster therefore yields an automatic pick. The additional knowledge required here can include choosing appropriate threshold values, and appropriate parameters for the given method (e.g., k-means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), etc.). A QC step would also be required to ensure that picks are reasonable.

The method proposed in this study differs significantly from previous work. Clustering is not used to identify picks in a single semblance panel. Instead, the method begins by collecting all semblance maxima in a panel, without requiring any knowledge about whether a given maximum is a good pick candidate or not. Clustering is then applied to this survey-wide collection of data to form horizon-like groups. It is at this point that additional knowledge is provided by the processor to identify survey-wide horizons of picks as appropriate to include in the model or not. This method thus allows the processor to make high-level judgements for the entire survey, while relegating tedious consideration of individual control points to the UML algorithm.

Theory and Method

We normally look at semblance in two-dimensional panels with velocity (V) and time (t). For the survey as a whole, however, the semblance is four-dimensional, with inline (IL) and crossline (XL) as well. It is useful to think of all semblance peaks associated with a particular seismic event as forming a group in this 4D space. Our objective is to use UML to form groups for several seismic events relevant to velocity picking. The projection of such a group into the 3D space of t-IL-XL will have the appearance of a time horizon. We therefore assume that a UML method such as DBSCAN which is appropriate for thin, sheetlike groups, can be used to form these groups in 4D space (or 5D space if picking η (eta) as well). Here the standard DBSCAN method has been modified with a constraint that a group may contain only one member from each CDP location, which strengthens its similarity to a horizon.

The group building process consists of the following four steps:

1) Scan through each control point in the survey to identify all semblance maxima greater than some threshold. This is performed automatically, requiring no interaction from the processor. At this point we also are not concerned whether a given maximum belongs in the velocity model, so we do not need to build into the algorithm any knowledge of what constitutes a good pick or a bad pick. A variety of methods could be imagined for locating these maxima, including the UML methods cited earlier. In this study we have used a very simple approach of calculating semblance values on a grid, and identifying as maxima those points whose neighbors all have smaller semblance values. This is followed by performing a simple quadratic interpolation to locate a refined maximum near the selected grid point.

2) For each semblance maximum, identify in each neighboring CDP location the semblance maximum closest

to it in t-V (or t-η, or t-V-η) parameter space. For example, in Figure 1 we see cartoons of semblance panels for three consecutive control points, generically labeled n-1, n, and n+1. In each semblance panel are depicted three semblance maxima labeled A_i, B_i, and C_i. The A_i and C_i peaks correspond to shallow and deep reflection events. The B_i peaks correspond to random semblance noise with no geological significance. Each arrow points from a peak in one panel to that peak in a neighboring panel which is closest to it in t-V space.

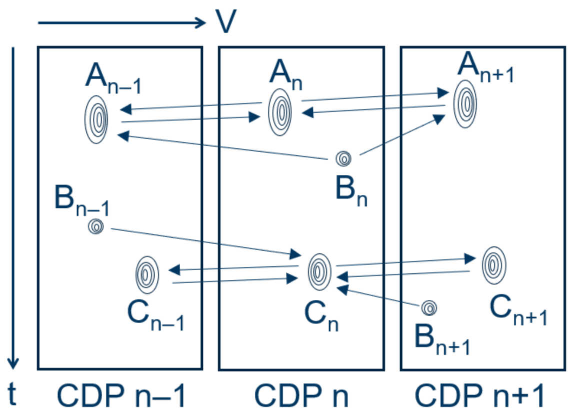

Figure 1: Illustration of steps 2 and 3 of the group building process, as described above and below.

3) Identify pairs of maxima which are in neighboring CDP locations and which are each other's closest neighbors in parameter space. These will be called connected pairs. For instance, in Figure 1 we see that semblance maxima from neighboring panels corresponding to the same geological event tend to be each other's closest neighbor in t-V space, while maxima arising from random semblance noise do not tend to form pairs. By considering only connected pairs, the algorithm focuses on geologically significant information in the semblance panels.

4) Apply a DBSCAN-type method to build groups of connected pairs. The DBSCAN method depends upon a concept of closeness. In this application, two peaks are considered close to each other if they are a connected pair. For a 3D survey each peak can belong to four different connected pairs with its nearest neighbors. This supports the creation of sheetlike groups of all peaks corresponding to the same reflection event.

The above method allows us to create geologically meaningful groups in 4- or 5-dimensional space. While we cannot visualize such groups, we can understand and QC their essential information by using projections into 2- or 3-dimensional space, as illustrated in the next section.

Results

This method has been applied to a land dataset from the Utica basin (Sherrodsville). The method was applied twice, once in each of two stages:

First Stage: η from 5D picking

In the first stage, an anisotropic V+η model was obtained by forming groups in 5-dimensional t-V-η-IL-XL space using the four steps described in the previous section. For efficiency, a velocity function was manually picked at a central control point and was used as a 1D guide function for a %-velocity scan in the first step. After the group building, each point in a group was identified by a quintuple of values, (t_i, V_i, η_i, IL_i, XL_i), and the group could be visualized by projection into various subspaces.

Figure 2a shows projection into t-V space and has the appearance of a typical semblance panel. Appropriate display tools were created to aid the user in selecting groups corresponding to primary events. Each decision by the user was applied to a large portion of the survey at once, rather than requiring many decisions to be made at individual control points.

Figures 2b-2d show projections into t-IL-XL, V-IL-XL, and η-IL-XL spaces respectively. These have the appearance of horizons, with the t-horizon being quite sharp, the V-horizon somewhat fuzzy, and the η-horizon very diffuse. The η model of Figure 2d was spatially smoothed, with a result displayed in Figure 2e. This smooth η result was held fixed for the remainder of the velocity model building, and the vertical velocity values were repicked in the second stage.

Second stage: refined velocity from 4D picking

A refined anisotropic velocity model was obtained by forming groups in 4-dimensional t-V-IL-XL space with the smoothed η model held fixed. This allowed us to use a finer velocity grid for calculation of the 2D anisotropic semblance panel at each control point. Groups were formed using the same four step procedure, which was able to create a good quality anisotropic velocity model, as illustrated in Figure 3.

In the process of obtaining this final result, we applied QC and refinement tools to the initial groups. Several tools were created to assist the user in rejecting outliers, interpolating missing points, combining subgroups, etc. An illustration of this is shown in Figure 4.

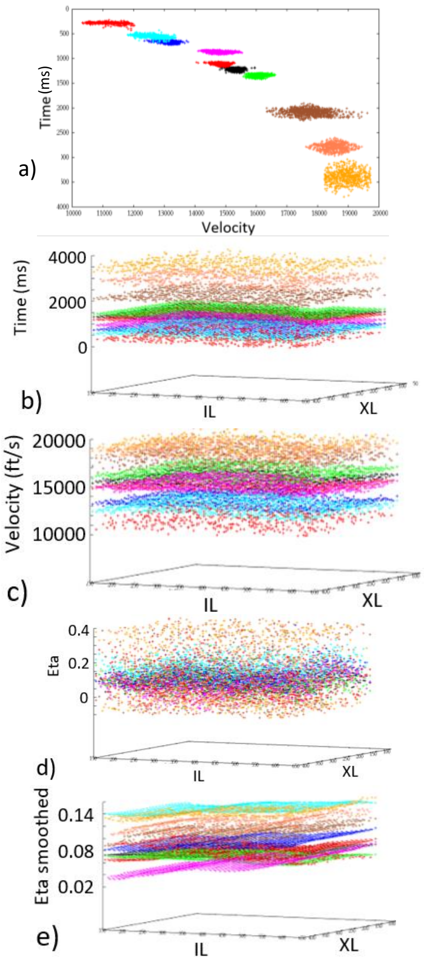

Figure 2. Various projections of the 5-D groups from the first stage of velocity model building: a) projection into the t-V (i.e., $t_{two\text{-}way}$-V_{RMS}) parameter space of a typical semblance panel. Each color represents a different group, and each point represents the semblance maximum at some CDP location. b) projection into t-IL-XL space, where each group resembles a typical time horizon. c) projection into V-IL-XL space, forming velocity horizons. d) projection into η-IL-XL space, forming very diffuse η horizons, because of η scatter. e) same as d), but after spatial smoothing of the η values within each group.

Additional comments

1) Dip volumes can be optionally incorporated into the group building to account for structure in determining closeness of two semblance maxima in adjoining locations.

2) As a byproduct, groups can be used to generate time horizons (Huang et al., 2020). Alternatively, one can use existing horizons as scaffolding to help in forming groups.

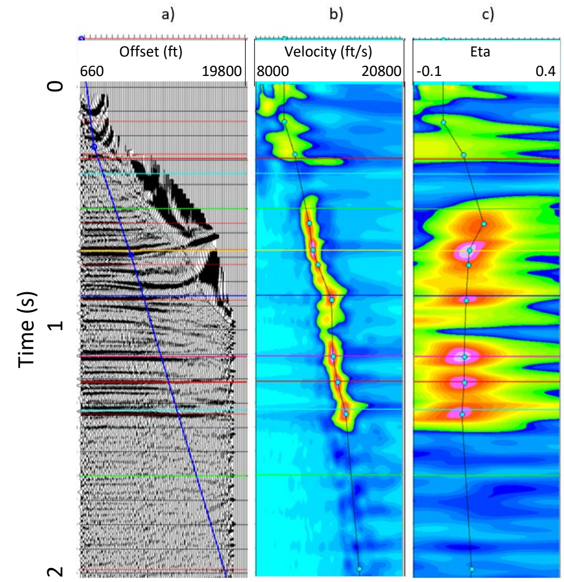

Figure 3. An example of results from this method: a) CDP gather with primaries flattened inside the mute, b) vertical velocity semblance panel, and c) η semblance panel. Horizontal lines in the semblance panels represent horizon values at this location. The horizons were picked using an automatic method, and were used to aid in forming some of the groups.

3) As a benefit, this method lends itself to geologically consistent velocity picking.

4) Another significant benefit is that, because of the nature of the group-building algorithm, dense picking of velocities is easier than sparse picking.

Conclusions

More than just a new technique, this approach represents a novel paradigm for machine learning in seismic processing. Rather than *replace* a human, this method *augments the skills* of experienced processors; UML performs repetitive tasks, while the processor still interactively creates the velocity model. Such a philosophy has also been applied to seismic interpretation (Lowell & Erdogan, 2019), and has been advocated more generally (Kasparov, 2017; De Cremer & Kasparov, 2021).

Expecting black-box automation to replicate the results of capable processors is often unrealistic. This alternative approach simplifies development, speeds up processing, and yields quality products.

Acknowledgements

We thank Tim Seher for bringing the work of Kasparov to our attention.

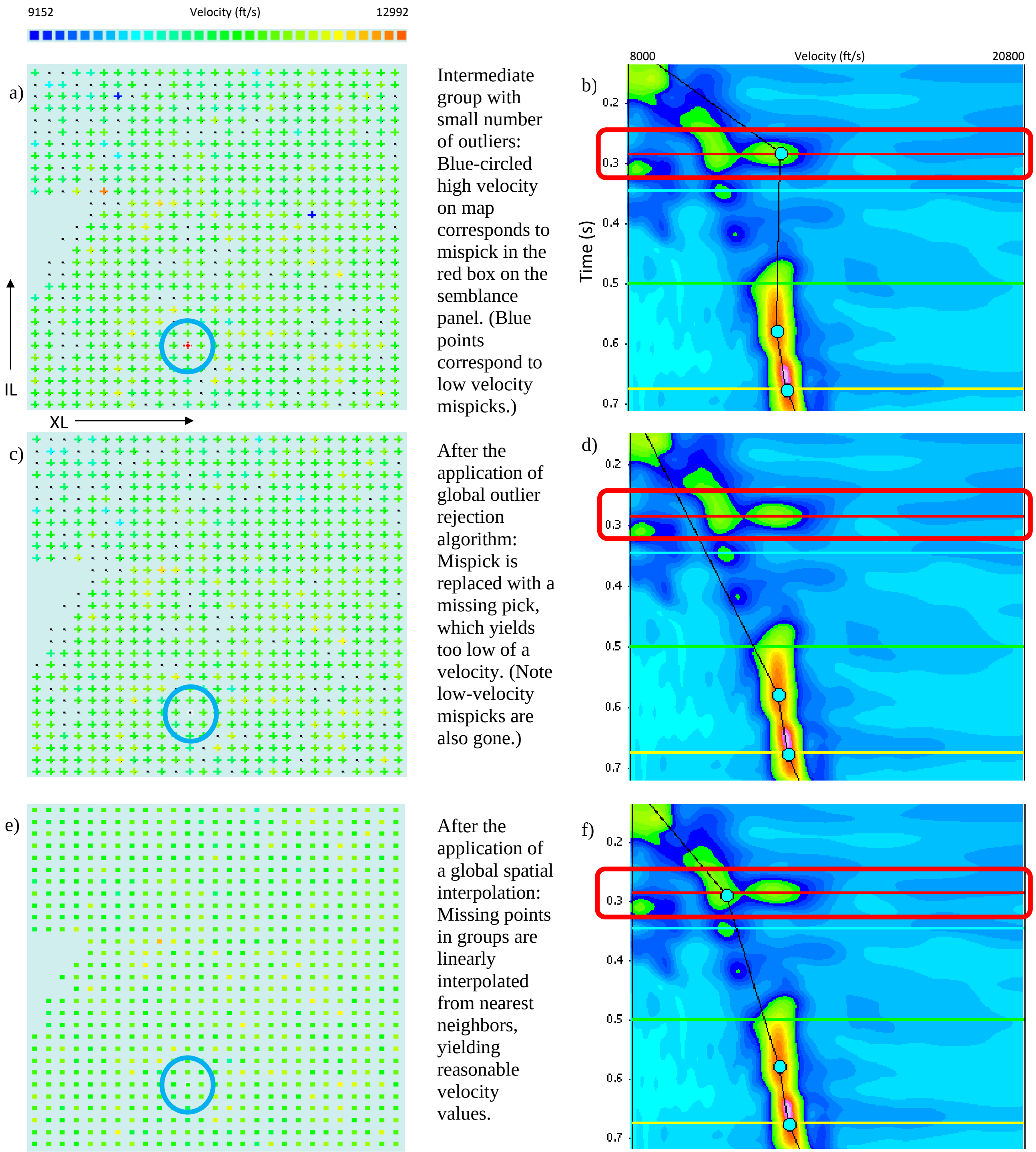

Figure 4. On the left are map displays of velocity near the time marked by the red boxes over the semblance, on the right. (a) and (c) show velocities for a group near this time, with small black dots representing missing picks, i.e., control points not represented in the group. (e) shows a velocity time slice near the same time. The semblance is shown with a mispick in (b), a missing pick after outlier removal in (d), and a reasonable pick after spatial interpolation in (f). No smoothing is applied to the result. This figure illustrates how the processor can use global operations to refine results interactively across the survey without needing to make changes at individual control points.