Toward High-Confidence and Full Automation of Seismic Data Processing for CTBT Monitoring

Jin Ping
(Northwest Institute of Nuclear Technology, China)
Contents

- Introduction;
- Review of existed technologies;
- Progresses achieved;
- Summary;
1. Introduction

• Seismic monitoring activities, either for CTBT or earthquake-induced disaster control, are facing a new era of big data as the number of seismic stations with data open available or to be dealt with increasing rapidly.
For the CTBT monitoring, the 170 IMS seismic stations would have the capability to detect events of $mb > 3.2$ for a very large part of the globe, equivalent to signals of more than 100,000 events may be yearly registered and to be processed.

Fig.1 Network detection capability map of the 41 International Monitoring System (IMS) primary seismic stations operational by 1 January 2012. 
For earthquake monitoring, as the result of the application of modern digital technologies in the field of seismic observation, many parts of the world are covered by densely-distributed seismic stations. These non-IMS seismic stations can also be used for CTBT verification.

Fig.2 Seismic stations affiliated to the National Seismological Bureau of China.
• To process the huge amount of seismic waveform data, automatic technologies which can reliably and efficiently detect, localize and categorize seismic events is crucial for both CTBT and natural seismicity monitoring;
• However, existing automatic technology for seismic data processing still facing great technical obstacles in improving its reliability and efficiency;
• Summary of accuracy of IDC automatic processing
  ✓ Almost 30% of events which appear in SEL3 have been discarded by analysts as invalid;
  ✓ Analysts have substantially modified a further 50% of SEL3 events;
  ✓ Events added by analysts constitute on average about 20% of events in the reviewed event list (LEB);

• Therefore up to date, routine seismic monitoring still heavily relied on manual interactive analyses, and this reality can hardly fit the situation of the era of rapidly increasing amount of seismic data.
2. Review of existing technologies for automatic seismic data processing.

- Conventional procedure of seismic monitoring for CTBT can be split into four stages:

<table>
<thead>
<tr>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Data Processing (Event Detection and Location)</td>
</tr>
<tr>
<td>Analyst Review (Review and correct automatically produced events one by one)</td>
</tr>
<tr>
<td>Automatic Event Characterization &amp; Screening</td>
</tr>
<tr>
<td>Special Event Analyses</td>
</tr>
</tbody>
</table>

Fig. 3 Principal stages of CTBT seismic monitoring
To improve the automation degree of the monitoring, it is crucial to develop more reliable and effective technologies to detect, localize and categorize seismic events automatically;
For automatic event detection and localization, the conventional method is by the “DIAL” approach;

Fig. 4 Typical approach for automatic seismic data processing.
• This approach usually is characterized by features including:

✓ Signals are detected individually, though in many cases (especially under the situation of regional seismograms) they may arrive collectively at a station as wavetrains;

Fig. 5 A diagram to illustrate wavetrains and detections.
Through the whole processing, seismic signals are principally characterized by features that are local in time domain, features characterizing the whole shape of respective wavetrains are ignored;
Fig. 6 A diagram to illustrate the measurement of local features that are used to identify and associate seismic signals by conventional DIAL approach. During the process, the information about the shape of the whole seismograms is lost and links between different detections is basically cut off.
Decision methods used for signal detection, identification and association basically are based on theories for simple scenarios. Their performances usually are significantly reduced in the situation of real, complex data;

Fig. 7 Illustration of the possible gap between theoretically assumed (left) and practical (right) distribution of feature parameters.
And the processing is basically sequential, other viable explanations are ignored in many cases by this way;

![Diagram showing potential mistakes due to sequential processing.](image)

location uncertainty of seed origin

false location due to false association

correct location

Fig. 8 Illustration of potential mistakes due to sequential processing.
Visit to waveform data is limited to the detection & feature extraction stage, following processing stages are only based on derived, information-incomplete signal parameters;

Fig. 9 The model of waveform data accessing of automatic processing systems.
The low performance of seismic data processing systems is intrinsically related to these features. It can be illustrated by comparing computer processing methods with that used by human being analysts;

Table 1 Comparison of reasoning approaches between automatic processing and interactive analyses

<table>
<thead>
<tr>
<th></th>
<th>Automatic</th>
<th>Interactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association mode</td>
<td>from part to whole;</td>
<td>from whole to part;</td>
</tr>
<tr>
<td>Signal feature used</td>
<td>local only;</td>
<td>both local and integral;</td>
</tr>
<tr>
<td>Decision criteria</td>
<td>sequential, rigorous;</td>
<td>parallel, flexible;</td>
</tr>
<tr>
<td>Visit to waveform data</td>
<td>limited;</td>
<td>through the analyses;</td>
</tr>
</tbody>
</table>
Fig. 9 Illustration of differences in approaches for signal detection, phase type identification and association between computer programs and human analysts. Contrary to the former, the processing procedure used by analysts normally is from whole wavetrains to specific signals, applying both local and integral features.
To develop more reliable and accurate automatic seismic data processing systems, approaches, including signal features and reasoning rules, methods used by human being analysts should be generalized and applied;
3. Progresses Achieved;

- Automatic seismic data processing using both integral and local features of seismograms;
- Progressive signal association algorithm for detecting teleseismic/network-outside events using regional seismic networks;
- Prototype Computer Analyst System for seismic event review;
3.1 Data processing using integral features.

- A novel technique has been developed to incorporate integral features into conventional DIAL approach for automatic seismic data processing; (Jin et al, 2014, A novel technique for automatic seismic data processing using both integral and local feature of seismograms, Earthq. Sci., 27, 337-349.)
• Specialties of the technique.
✓ Detect individual signals together with associated wavetrains;
✓ Detections are categorized into the beginning or the following type;

Fig. 10 Illustration of the beginning detection and following detections of a wavetrain. Normally direct P waves can not be following detections and they can be associated to respective beginning detection straightforwardly.
Integral features are introduced to characterize the full shape of wavetrains;

Fig. 11 Illustration of the concept of integral features of wavetrains (left). The rough envelope shape of seismograms can be inferred from these features (right).
Use the integral features along with conventional local features for seismic phase type identification and association;

Fig. 12 Illustration of the method to search for regional P/S pairs by both integral and local features.
• Very promising results has been achieved by applying this technique with fairly low false and missed event ratio simultaneously obtained.

—Summary of test results using 16 days’ data of Xinjiang Seismic Network (XJSN).

✓ False events ratio: ~8%;
✓ Missed events ratio: <2% for ML>1.0;
✓ Average Location error: 15.6km ; (in the case of the sparse regional network XJSN)
Fig. 14 Left: Statistics on detected Xinjiang Event Bulletin (XJEB) events; Right: Comparison of automatically determined epicenter locations with that listed in XJEB for events of ML > 2.0;
Fig. 15 An example of automatic processing results for a far regional event.
Fig. 16 Examples of automatically detected local/regional events which have been missed by XJEB.
3.2 A novel association algorithm for detecting teleseismic/network-outside events using regional seismic networks

- Monitoring underground nuclear explosions using regional seismic networks is a common practice;
- For various reasons, little attention has been paid to this problem;
- To deal with this issue, we have developed a special progressive association algorithm which can reliably detect teleseismic/network-outside events using regional seismic networks; (Jin et al, 2015, A Novel Progressive Signal Association Algorithm for Detecting Teleseismic/Network-outside Events Using Regional Seismic Networks, Geophys. J. Int., 201, 1950-1960.)
This association algorithm exploits the relationship between the arrival times and the slowness of teleseismic signals as well as the unique slowness range of teleseismic P waves;

$$t_i - t_r = -(x_i - x_r)p_x - (y_i - y_r)p_y + \varepsilon_i$$

Fig.17 Left: Diagram illustrates a teleseismic signal propagating across a regional seismic network; Right: The relationship between slowness and epicentral distances for frequently observable seismic phases.
It takes the so-called primary triangle station arrays (PTAs) as the starting point to search for P waves of teleseismic events progressively;

Fig. 18 Predefined PTAs for XJSN to detect teleseismic events
• This algorithm can effectively and reliably detect network-outside seismic events with very low ratio of false events (<3%);

Fig. 19 Magnitude and distance distributions of (a) detected and (b) undetected IDC REB events using data of XJSN from 2010/06/01 to 2010/06/16. REB events undetected basically have no signal recorded at XJSN stations.
Fig. 20 Comparison of automatically determined epicenters using XJSN with those in IDC REB.
3.3 prototype Computer Analyst System

- To realize the full automation of seismic data processing ultimately, we are making effort to develop a prototype Computer Analyst System (pCAS);

Fig. 21 Proposed approach for full automation of seismic data processing for CTBT monitoring.
• Its function is to substitute human being analysts for seismic event review, including:
  ✓ remove false origins produced by APS and search for missed events;
  ✓ remove falsely associated phases and add previously undetected/unassociated ones;
  ✓ revise mistakes in phase timing and labeling;
  ✓ reconstruct seriously mislocated events produced by APS;
A primitive version of pCAS has been designed for test.

Fig. 22 Principal modules of pCAS and related functions.
Preliminary test results are promising, though there are still a lot of technical challenges to overcome;
Summary of preliminary test results of pCAS;

- Split and false events were removed nearly without the loss of meaningful events;
- Falsely associated phases were reduced by more than 60% while the total number of associated signals averagely increased by 30% for accepted events;
- Severely mislocated events were successfully reconstructed, significantly reducing the upper limit of location errors relative to results of interactive analyses;
Right down: Revised signal pickups and associations by pCAS. The modified epicenter is very close to that determined by interactive analyses.

Fig. 23 Illustration of improvement in signal pickups and associations completed by pCAS for severely mislocated events.
Summary

• High-confidence and full automation of seismic data processing is crucial for effective CTBT monitoring;

• Conventional local-feature based approaches for automatic seismic data processing hardly can satisfy this requirement;

• Novel processing technologies in better patterns and incorporating signals integral features as well as new methods for reasoning are needed;
• With the application of these techniques along with state-of-art theories or technologies in the field of signal and information processing, especially those about artificial intelligence, the full automation of seismic data processing can be expected.
Thank you for your attention!