SnT 2019
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

CTBTO PREPARATORY COMMISSION | PUTTING AN END TO NUCLEAR EXPLOSIONS

Automatic characterization of phase type at three-component seismic stations using neural networks

Radek Hofman

radek.hofman@ctbto.org

T3.5-P12

1. Neural networks at IDC

Introduction:

- According to *Sereno and Patnaik, (1993)*, the neural network based initial wave type classification system was created in early 1990s (as en extension of a two class system P vs. S) and became part of IMS v3 (IMS is here *Intelligent Monitoring System*)
- Station-specific weights were trained using analyst-reviewed data for two three-component (3-C) stations which had enough data
- For the rest of stations **average** weights yielded by training using data from multiple stations were produced
- Around 2002 the networks have been re-trained using data from STKA station which has a low noise, see *(J. Wang, 2002)*
- Since then, these weights are used as weights for all 3C stations in IMS (IMS stands here for *International Monitoring System*)
- m_b SNR screening was introduced later to improve classification performance of the noise
- Our goal is to investigate if the classification performance of 3-C seismic stations can be improved by:
  1. Re-training of legacy multilayer perceptrons taking extracted features on input using more station-specific data
  2. Replacement of the legacy classifier by a new one based on convolution neural networks taking waveform on input

2. Re-training of current neural networks

Problem statement:

- Initial wave type is determined using StaPro (Station Processing) software for all IMS arrivals detected at IDC
  - iwt∈{**noise**, **regional S** (Sn, Lg, Rg, Sx), **regional P** (Pn, Pg, Px), **teleseismic**(P, tx)}
  - initial wave type is further refined to more specific wave types
- For 3-component (3-C) stations the system relies on a cascade of three binary classifiers implemented as neural networks (NNs)
- Current NN weights were derived in 2002 using data from a single station (STKA) and are applied to **all** 3-C stations worldwide
- We now have much more station specific data reviewed by analysts
- Our goal is to investigate if the classification performance of 3-C seismic stations can be improved by re-training using more **station-specific** data

Three stage classifier:

- The system comprises of a cascade of three binary classifiers
  - *Step 1:* the system tries to distinguish between noise **N** and a signal {**regS**, **regP**, **tele**}
  - *Step 2:* if signal: is it **regS** or {**regP**, **tele**}?
  - *Step 3:* if {**regP**, **tele**}: is it **regP** or **tele**?
- All three classifiers are implemented as a multilayer perceptron (neural network)
- After NN classification, there is a revision stage of possible **T→regP** change and **regS** refinement using a Bayesian model

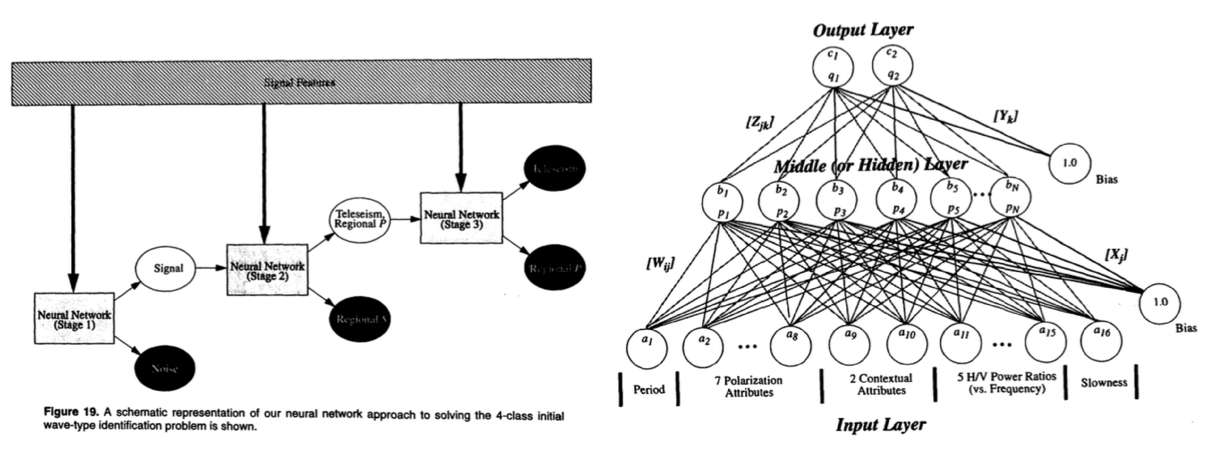


Figure 1: Representation of the three stage classifier and topology of each of its stage. Illustrations from the original paper *(Sereno and Patnaik, 1993)*

Retraining of current neural networks contd.

Classification features:

#	Feature	Description	Source	Norm.
1	*period*	Dominant period of the detected phase	ARRIVAL	
2	*rect*	Signal rectilinearity	APMA	
3	*plans*	Signal planarity	APMA	
4	$inang_1$	Long-axis incidence angle	APMA	/90
5	$inang_3$	Short-axis incidence angle	APMA	/90
6	*hmxmn*	Ratio of the maximum to minimum horizontal amplitude	APMA	$\log_{10}$
7	*hvratp*	Ratio of horizontal-to-vertical power	APMA	$\log_{10}$
8	*hvrat*	Similar to *hvratp*, measured at the time of max 3C amplitude	APMA	$\log_{10}$
9	$N_{after} - N_{before}$	Diff. between the no. of arrivals before and after within $\pm 60s$	on the fly	/10
10	$T_{after} - T_{before}$	Mean time diff. between arrivals before and after within $\pm 60s$	on the fly	/100
11	$htov_1$	Horiz. to vert. power ratio in oct. freq. band centered at 0.25 Hz	AMP3C	$\log_{10}$
12	$htov_2$	Horiz. to vert. power ratio in oct. freq. band centered at 0.5 Hz	AMP3C	$\log_{10}$
13	$htov_3$	Horiz. to vert. power ratio in oct. freq. band centered at 1.0 Hz	AMP3C	$\log_{10}$
14	$htov_4$	Horiz. to vert. power ratio in oct. freq. band centered at 2.0 Hz	AMP3C	$\log_{10}$
15	$htov_5$	Horiz. to vert. power ratio in oct. freq. band centered at 4.0 Hz	AMP3C	$\log_{10}$

2.1 m_b SNR screening

- m_b SNR screening is prepended to NN classifier in StaPro
- All 3-C arrivals are checked by this screening and labeled as noise if $m_b < m_b^{min}$, m_b^{min} is a station specific threshold
- Its purpose is to identify noise phases using m_b and exclude them from classification where significant amount of them is classified as a real phase due to poor classification performance
- **Drawback:** this simple criterion labels also significant number of associated phases as noise (increases *N-phase rate*) and they do not enter NN classification

m_b SNR screening – evaluation of effects:

- Results are produced using current URZ OPS weights without m_b SNR (left) and with m_b SNR (right)
- 17999 arrivals detected Jan – May 2017 (1581 associated in LEB, 16418 noise phases)

	old weights **without** m_b SNR				old weights **with** m_b SNR			
	N	S	P	T	N	S	P	T
N	**4289**	99	9	108	**8312**	113	35	167
S	2371	**134**	0	138	1743	**124**	0	128
P	3305	30	**266**	277	2097	26	**237**	244
T	6452	9	34	**477**	4265	9	37	**461**
	28.7% (assoc. **55.5%**)				**50.7%** (assoc. **51.9%**)			
	N-phase rate **13.6%**				N-phase rate **19.9%**			

2.2 Dataset description

- **Training:** analyst reviewed URZ automatic arrivals **2003 – 2016**
  - Signal phases are all automatic arrivals associated to LEB (arrivals) with re-timing $\leq$ 2 seconds (29252 arrivals)
  - Noise phases are selected randomly from all automatic arrivals classified as noise and not associated to LEB
  - Training data is split on training and validation data in ratio 9:1
  - Dataset is highly imbalanced – we employed weighted loss during training
- **Testing:** analyst reviewed URZ automatic arrivals **Jan – May 2017**

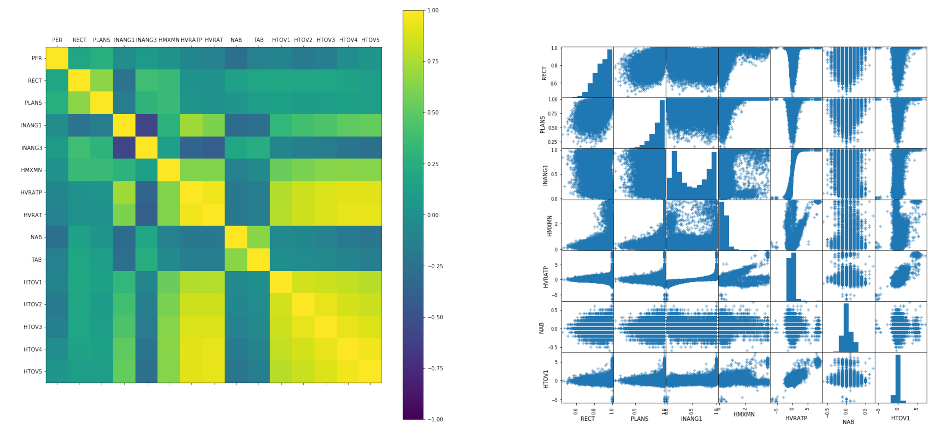


Figure 2: Features properties: correlation matrix (left), scatter plot matrix (right).

2.3 Re-training results

Test set results:

- Offline comparison of confusion matrices for old and new weights Jan - May 2017
- Test set results indicate that we can improve both classification accuracy and N-phase rate at the same time by re-training (without the need of m_b SNR screening)

	old, m_b SNR, T→regP ref.				new weights				new, T→regP ref.			
	N	S	P	T	N	S	P	T	N	S	P	T
N	**8312**	113	35	167	**9740**	48	12	43	**9740**	48	12	43
S	1743	**124**	0	128	2477	**220**	2	221	2477	**220**	2	221
P	2097	26	**237**	244	991	2	**195**	68	1948	3	**254**	191
T	4265	9	37	**461**	3209	2	100	**669**	2252	1	41	**546**
	50.7% (assoc. **51.9%**)				**60.1%** (assoc. **68.5%**)				**59.8%** (assoc. **64.5%**)			
	N-phase rate:**19.9%**				N-phase rate: **6.5%**				N-phase rate: **6.5%**			

Development LAN results:

- 2018/07/25: weights in DVL, m_b SNR screening turned off for URZ (m_b^{min} set to 0)
- Evaluation performed on arrivals between 2018/08/01 and 2018/09/30 (2 months, 4243 arrivals, 494 associated to LEB)
- DVL and OPS arrivals matched by proximity in time and azimuth

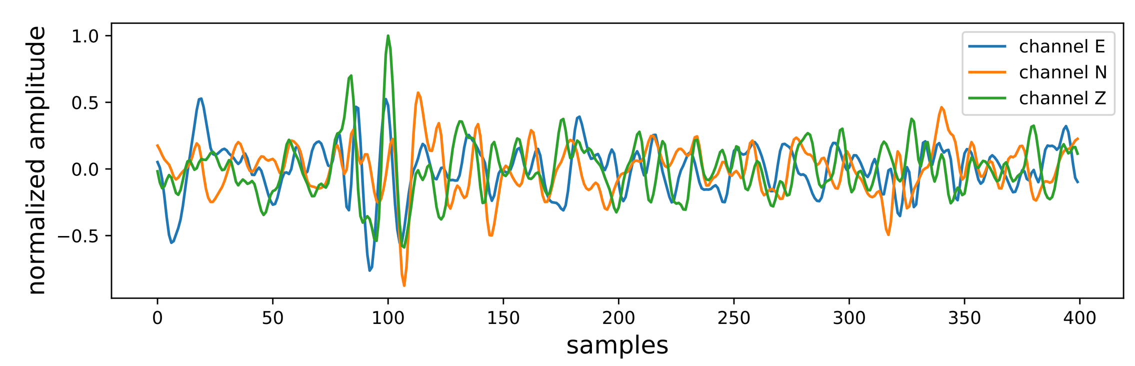
	OPS: old, m_b SNR, T→regP ref.				**DVL**: new, T→regP ref.			
	N	S	P	T	N	S	P	T
N	**1879**	33	9	48	**2042**	8	3	11
S	439	**56**	2	39	719	**83**	2	64
P	497	2	**76**	83	503	1	**84**	60
T	934	1	10	**135**	485	0	8	**170**
	50.6% (assoc. **54.0%**)				**56.1%** (assoc. **68.2%**)			
	N-phase rate: **17.0%**				N-phase rate: **4.5%**			

3. Convolution network

- Recently, works on successful application of convolutional neural network (CNNs) for phase identification have been published, .e.g. *Ross et al. (2018)* and *Woollam et al.(2019)*
- CNNs take waveform on input so the extraction of (subjective and possibly highly correlated) features can be avoided
- We take arrivals from the STA/LTA detector and try to train a single CNN for phase classification, i.e. with probabilities for all four classes on output
- As class labels for training we use labels given to associated phases by analysts, all other arrivals are considered as noise
  - This could bias the dataset — some real seismic phases are not associated because of their redundancy during event creation

Dataset:

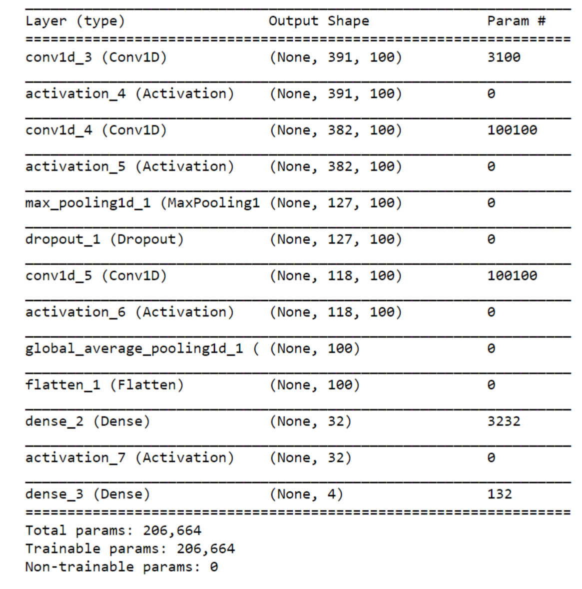
- URZ 2003-2016 data — 10s waveforms (400 samples) around STA/LTA arrivals from all three channels
- Dataset heavily unbalanced: (268103, 1113, 6364, 78289) representatives for (N,S,P,T)
- Channels pass-band filtered normalized by max. abs. amplitude of the sample
- 10% of representatives of each class used for testing



3.1 Convolutional model and results

Convolutional model:

- We use `Keras` to implement and train the network on GPU:
  - Activation function rectified linear (`ReLU`) and `softmax` for the output layer
  - Loss `categorical cross-entropy`, optimizer `Adam`
  - Training stopped by after 48 epochs with with batch size 512
  - Class weights (0.05, 100, 10, 1) for classes (N,S,P,T)



```
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_3 (Conv1D)            (None, 391, 100)          3100
_________________________________________________________________
activation_4 (Activation)    (None, 391, 100)          0
_________________________________________________________________
conv1d_4 (Conv1D)            (None, 382, 100)          100100
_________________________________________________________________
activation_5 (Activation)    (None, 382, 100)          0
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 127, 100)          0
_________________________________________________________________
dropout_1 (Dropout)          (None, 127, 100)          0
_________________________________________________________________
conv1d_5 (Conv1D)            (None, 118, 100)          100100
_________________________________________________________________
activation_6 (Activation)    (None, 118, 100)          0
_________________________________________________________________
global_average_pooling1d_1 ( (None, 100)               0
_________________________________________________________________
flatten_1 (Flatten)          (None, 100)               0
_________________________________________________________________
dense_2 (Dense)              (None, 32)                3232
_________________________________________________________________
activation_7 (Activation)    (None, 32)                0
_________________________________________________________________
dense_3 (Dense)              (None, 4)                 132
=================================================================
Total params: 206,664
Trainable params: 206,664
Non-trainable params: 0
_________________________________________________________________
```

Figure 3: `Keras` model summary of our convolution network.

First results seem to be promising:

	Convolutional network			
	N	S	P	T
N	**16342**	7	2	152
S	2244	**82**	2	116
P	2032	4	**590**	876
T	6192	18	42	**6685**
	accuracy **67.0%** (assoc. **85%**)			
	N-phase rate **1.8%**			

Discussion

Re-training of current classifier:

- Both accuracy and noise improved after retraining using station-specific data
- At the same time, number of associated phases classified as noise in automatic processing decreased mainly due to removal of m_b SNR screening criterion
- Retraining of other 3C stations will follow

Experiments with CNNs:

- Experiments with CNNs give promising results with potential to supersede the old 3-stage multilayer perceptron classifier
- Performance could be probably further improved by reviewing training dataset on seismic phases labeled as noise

References

- Ross, Z. E., et al. *Generalized seismic phase detection with deep learning.* Bulletin of the Seismological Society of America 108.5A (2018): 2894-2901.
- Sereno, T. and Patnaik, G., *Initial Wave-type Identification with Neural Networks and its Contribution to Automated Processing in IMS Version 3.0*, Tech. Rep., SAIC-93/1219, 1993
- Wang, J., *Adaptive training of neural networks for automatic seismic phase identification.* Monitoring the Comprehensive Nuclear-Test-Bant Treaty: Data Processing and Infrasound (2002): 1021-1041.
- Woollam, J., et al. *Conv. neural network for seismic phase classification, performance demonstration over a local seismic network.* Seis. Research Letters 90.2A (2019): 491-502.

Disclaimer