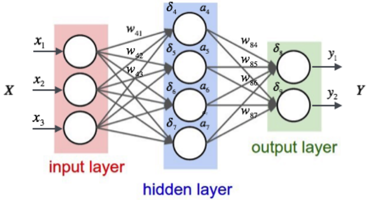
Tuning neural networks for initial wave type classification of detections at 3-component seismic stations in automatic processing

Radek Hofman, Ronan Le Bras, Helmuth Breitenfellner, Cahya Wirawan, Elena Tomuta



International Data Centre
Preparatory Commission for the Comprehensive
Nuclear-Test-Ban Treaty Organization

Problem statement

- Initial wave type (iwt) 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

Neural networks in StaPro

- 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

---

1. SERENO, T. and PATNAIK, G., Initial Wave-type Identification with Neural Networks and its Contribution to Automated Processing in IMS Version 3.0, Technical Report, SAIC-93/1219, 1993
2. 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.

Cascade of three binary classifiers

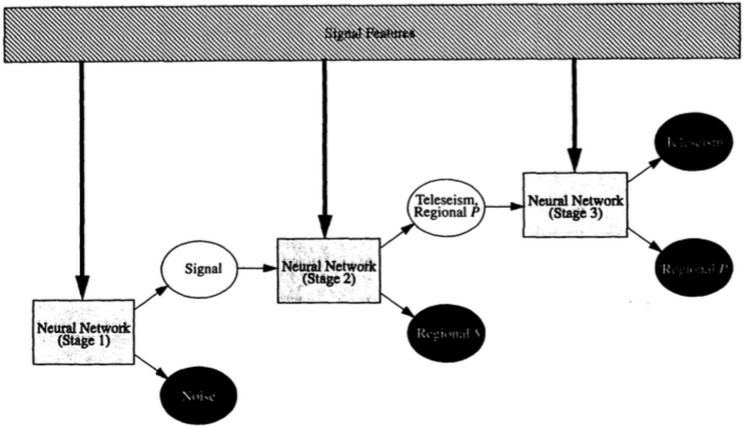
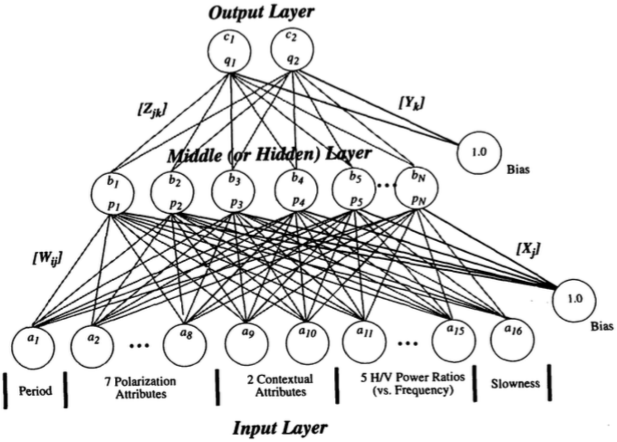


Figure 19. A schematic representation of our neural network approach to solving the 4-class initial wave-type identification problem is shown.

*Figure from (Sereno and Patnaik, 1993)

Multilayer perceptron topology

CTBTO PREPARATORY COMMISSION



- All three NNs in the cascade have the same topology
  - input layer of size 15
  - hidden layer of size 6
  - output layer of size 2
- All layers are dense with sigmoid activation
- Total 110 tunable weights

*Figure from (Sereno and Patnaik, 1993)

Inputs (classification features)

CTBTO PREPARATORY COMMISSION

iwt is determined using a set of features computed along with detection of arrivals:

#	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	$\log_{10}$

A brief history

- According to *Sereno and Patnaik, (1993*), the iwt 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 stands here for *Intelligent Monitoring System*)
- Station-specific weights were trained using analyst-reviewed data for two 3C 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 weights have been re-trained using data from STKA station, 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

m_b SNR screening

CTBTO PREPARATORY COMMISSION

- Currently, there is a m_b SNR screening prepended to NN classifier in StaPro
- All 3-C arrivals are checked by this screening and labeled as noise if a simple condition is met
- For each station, there is a given threshold in configuration file m_b^{min}, detection is labeled as noise if $m_b < m_b^{min}$
- Its purpose is to (probably) identify noise phases using m_b and not let them into NN classification where significant amount of them is classified as an associated phase
- **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

CTBTO PREPARATORY COMMISSION

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

old weights **without** m_b SNR				
	N	S	P	T
N	**4289**	99	9	108
S	2371	**134**	0	138
P	3305	30	**266**	277
T	6452	9	34	**477**
accuracy **28.7%** (assoc. **55.5%**)				
N-phase rate **13.6%**				

old weights **with** m_b SNR				
	N	S	P	T
N	**8312**	113	35	167
S	1743	**124**	0	128
P	2097	26	**237**	244
T	4265	9	37	**461**
accuracy **50.7%** (assoc. **51.9%**)				
N-phase rate **19.9%**				

Why re-training effort?

- Is it correct to use weights derived using a single station data globally?
- Can we do better if the weights were derived for each station individually using station specific data?
- Nowadays, there is much more analyst-reviewed data than before ($\sim$ 13 years for URZ)
- Re-training means just a change in configuration of StaPro – if we keep the original topology of NNs then no code changes are needed
- Possible problems with m_b SNR screening

Ingredients needed for successful re-training

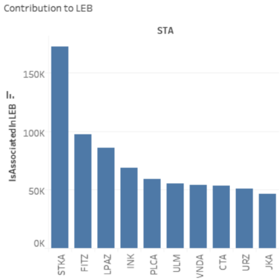
1. **Training data** - For neural network it holds true *the more, the better*
2. **Back-propagation algorithm** - The algorithm for NN training
   - We do not have the original C code used for training. However, the generic nature of NNs and our domain knowledge allows us to implement back-propagation algorithm
3. **A way to translate weights from re-training into StaPro configuration files**
   - We need to produce an alternate weights file and replace with it the original one in StaPro
4. **A simulated pipeline which can be run with the alternate weight file**
   - This gives us an alternate iwt classification which we can evaluate against analyst-reviewed data in a test set

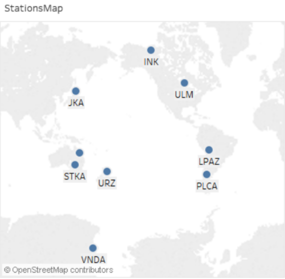
Ingredients needed for successful re-training

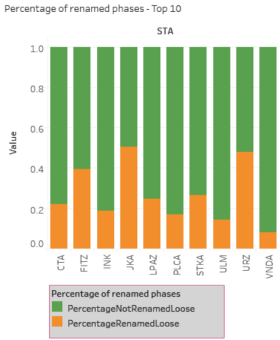
1. **Training data** - We chose to start with URZ station
   - It is one of the most data rich stations and its current iwt classification performance is around 50%
2. **Back-propagation algorithm** - We use Python library Keras backed by Theano
3. **A way to translate weights from re-training into StaPro configuration files**
   - We analyzed the C code and created a script putting the new weights into a weights file template in the right order
4. **A simulated pipeline which can be run with the alternate weight file**
   - We are able to run StaPro using the Station Tuning framework

Experiment setting

- Performance of iwt determination at 3-C stations varies significantly:







- We selected URZ to be retrained — a highly contributing station with above average iwt classification error

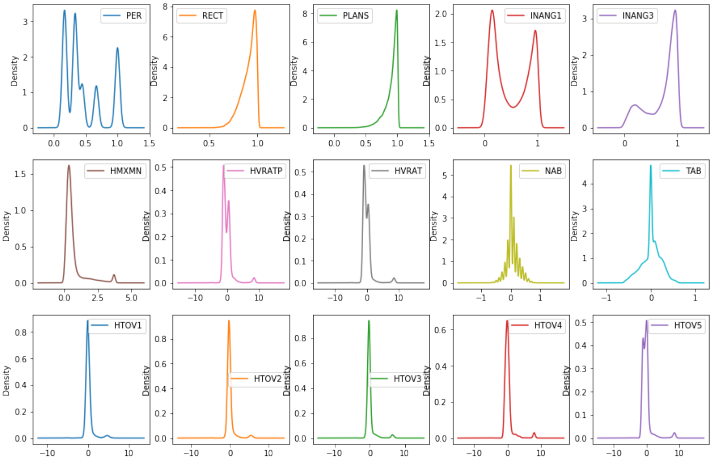
Dataset description

- **Training dataset:** analyst reviewed URZ automatic arrivals **2003 – 2016**
  - Signal phases are all automatic arrivals associated to LEB (arrivals) with re-timing < 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
- **Test dataset:** analyst reviewed URZ automatic arrivals **Jan – May 2017**

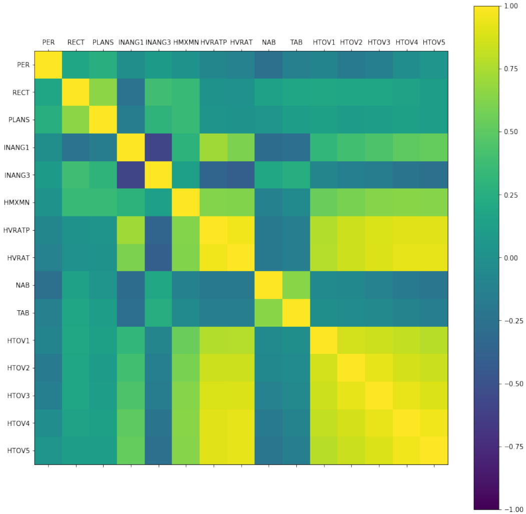
Dataset used by *Wang, (2002)* comprised of $\sim$ 1500 phases for all 4 classes

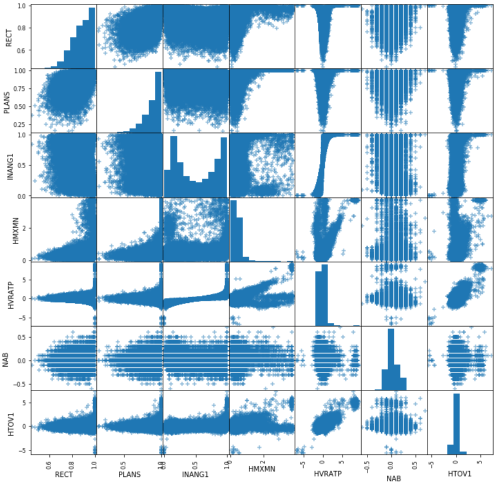
Dataset insights – density

Density plot of all features in training dataset. Some densities are multimodal.



Dataset insights – correlation





Test set results

CTBTO PREPARATORY COMMISSION

- 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 weights, m_b SNR, reg. P→T refined				new weights				new, reg. P→T refined			
	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% (**51.9%**)				**60.1%** (**68.5%**)				**59.8%** (**64.5%**)			
	N-phase rate: **19.9%**				N-phase rate: **6.5%**				N-phase rate: **6.5%**			

Test in automatic processing on DVL pipeline

CTBTO PREPARATORY COMMISSION

- 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 weights, m_b SNR, reg. P→T refined				**DVL**: new weights, reg. P→T refined			
	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% (**54.0%**)				**56.1%** (**68.2%**)			
	N-phase rate: **17.0%**				N-phase rate: **4.5%**			

Conclusion and future plans

- Classification accuracy of both noise and associated phases increased 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
- New URZ weights will be evaluated on TST automatic pipeline
- Retraining of other stations will follow
- Possible research directions:
  - Replace the cascade of three NNs with a single classifier for all four classes
  - Do not use pre-computed features but classify on the level of samples using recurrent or convolution networks