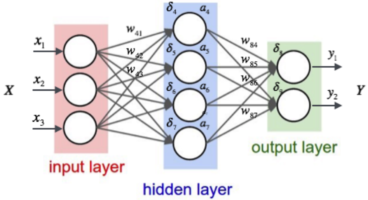
Re-training of primary 3-C stations



Radek Hofman



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

Introduction

- 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 3C stations this relies on a cascade of three binary classifiers implemented as multilayer perceptrons (very simple neural networks)
  - Step 1: the system tries to distinguish between noise 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?
- Finally, there is a revision stage of possible T→regP conversion (if a corresponding reg S phase exists) and regS refinement using a Bayesian model
- *"A neural network, an advanced feature available in StaPro, helps determine the signal type of detections. This feature requires off-line calculation of neural-network weights that define an empirically determined non-linear boundary between signal types and noise"*

Problem statement

- Current NN weights were derived in 2002 using data from a single station (STKA) and are applied to all 3-C stations worldwide
- This is not recommended in StaPro user manual:
  - *"Neural-network weights should be provided for all 3-C and hydroacoustic stations. In some cases, weights from another station may be used as long as the performance is satisfactory compared to that obtainable by the default rules. For example, weights for the 3-C station STKA have been used successfully at station ZAL for some time. It may also be acceptable to use another station's weights for a new station until enough ground-truth information has been acquired to train the neural net to obtain station-specific weights."*
- 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

StaPro IWT classification

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.

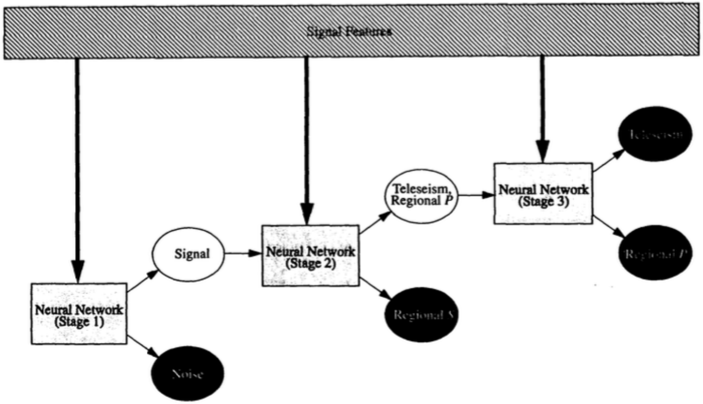
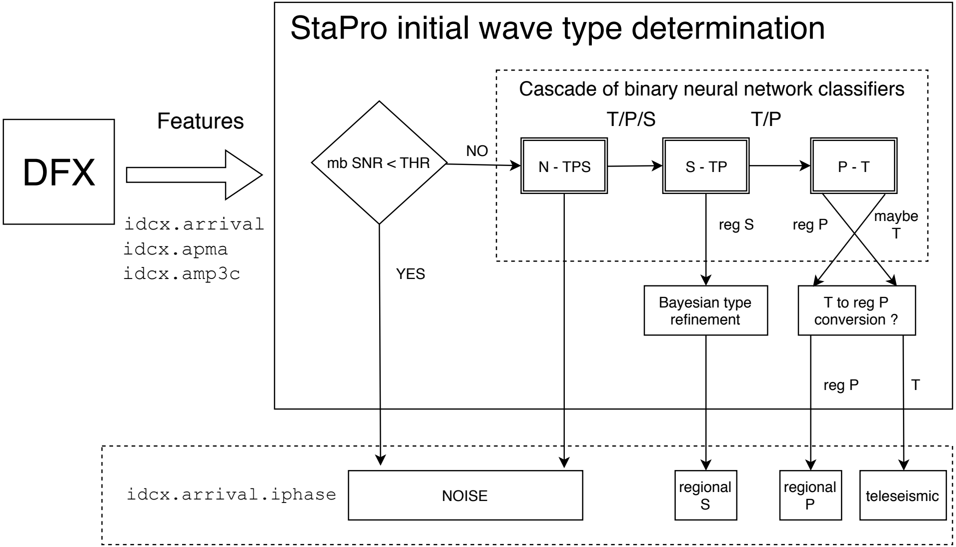


Figure 19. A schematic representation of our neural network approach to solving the 4-class initial wave-type identification problem is shown.

StaPro IWT classification



Classification features

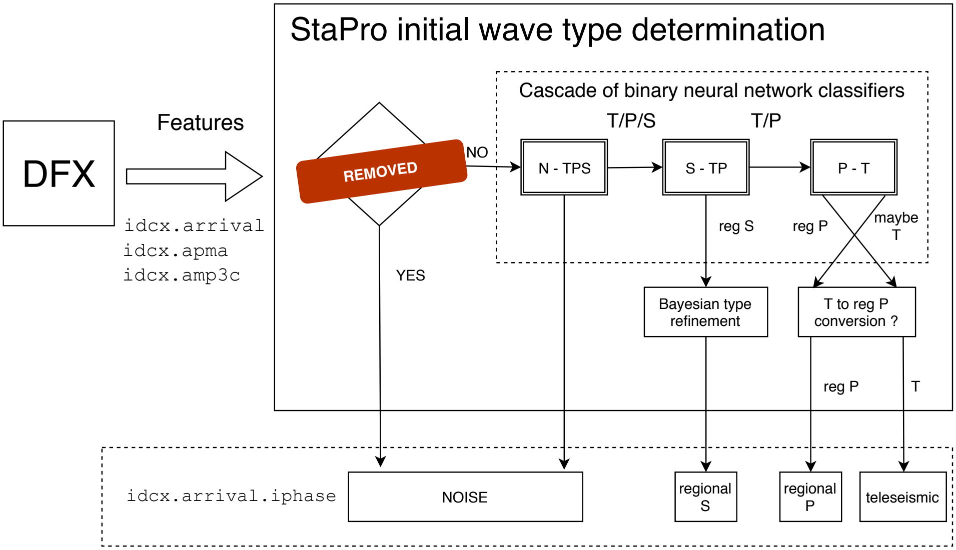
iwt is determined using a set of features computed along with the detection of arrivals:

#	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	$\log_{10}$

SNR_{m_b}- what does it do?

- SNR_{m_b} screening is prepended to multilayer perceptrons
- 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 THR, detection is labeled as noise if SNR in m_b frequency band is smaller than THR
- Its purpose is to (probably) identify noise phases using SNR and not let them into NN classification where significant amount of them is classified as an associated phase - because of generally badly performing N vs. TPS classifier
- In effect, this simple criterion ***misclassifies*** also a significant number of associated phases as noise and they do not enter NN classification
- We propose to remove $SNR_{m_b}$and rely just on newly trained N vs. TPS classifier

SNR_{m_b}- what does it do?



Dataset heavily imbalanced

- Dataset for all re-trained stations contains analyst reviewed LEB associated arrivals
- Not all phases were used, some were missing a subset of DFX features

		LEB assoc. phases for training			
sta	#N	#reg P	#reg S	#T	min lddate
BOSA	777,073	3,151	3,632	44,604	11-Jun-1999
CPUP	1,390,745	3,044	1,226	51,590	23-Jun-1999
KEST	450,929	2,969	670	20,838	31-Jul-2006
LPAZ	1,149,960	7,642	1,358	95,532	11-Jun-1999
NRIK	478,777	125	127	22,307	07-Jan-2011
PLCA	2,279,027	6,823	3,603	58,949	11-Jun-1999
PPT	481,797	11	7	4,517	21-Sep-2001
ROSC	472,657	2,500	1,204	211,99	09-May-2003
STKA	2,889,595	2,095	2,457	195,312	10-Jun-1999
ULM	982,366	3,677	3,046	58,741	10-Jun-1999

Training

- Original training SW is lost:
  - *"StaPro is a self-contained process. A related tool called nnet_train is available for training neural-network weights. This tool is used to train weights for the three-component stations."*
- A new python-based training tool has been created and a simulated automatic pipeline for evaluation of performance after all StaPro steps
- We optimize following criteria which we want to improve (**all of the at the same time**):
  - Overall classification accuracy (noise + associated phases)
  - Classification accuracy of associated phases
  - Γ- percentage of associated phase misclassified as noise
- For some stations not all stages were replaced by new weights just when the performance increased - this can be seen from confusion matrices
- Re-training: provision of new weights for a subset of NN classifiers and disabling of SNR_{m_b}(set `noise-mbsnr` to zero)

Training - confusion matrix analysis example

- A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known
- A perfect confusion matrix has all off-diagonal elements 0

BOSA										
re-trined						OPS baseline				
	N_{true}	T_{true}	P_{true}	S_{true}			N_{true}	T_{true}	P_{true}	S_{true}
N_{iwt}	**45192**	105	27	165		N_{iwt}	**29371**	213	10	115
T_{iwt}	25026	**4423**	5	13		T_{iwt}	33163	**4065**	2	30
P_{iwt}	2134	34	**113**	0		P_{iwt}	18105	361	**333**	111
S_{iwt}	20692	88	201	**270**		S_{iwt}	12405	11	1	**192**
acc: 50.77%; aacc:88.3%						acc: 34.5%; aacc: 84.3%				
$\Gamma = 5.46\%$						$\Gamma = 6.21\%$				

StaPro results (1/1/2018 – 31/12/2018)

			retrained			baseline OPS (%)			re-trained (%)			Δ (%)		
#	sta	SNR_{m_b}	N	S	TP	acc	aacc	Γ	acc	aacc	Γ	acc	aacc	Γ
1	BOSA	1.5	✓		✓	33.5	84.6	5.9	49.6	90.8	5.3	16.1	6.3	-0.6
2	CPUP	2.0	✓		✓	47.6	83.6	8.2	69.3	85.0	7.5	21.7	1.4	-0.7
3	KEST	—	✓			24.7	79.8	3.4	29.6	79.8	2.9	4.9	0	-0.4
4	LPAZ	2.0	✓	✓	✓	31.8	76.8	5.3	47.8	79.5	3.5	16.1	2.7	-1.9
5	NRIK	—	✓		✓	16.8	88.3	1.4	28.5	89.2	1.0	11.7	0.9	-0.3
6	PLCA	1.7	✓			44.1	83.4	7.2	47.6	87.2	2.5	3.5	3.8	-4.7
7	PPT	2.0	✓	✓	✓	57.7	59.2	16.0	78.3	69.5	10.9	20.6	10.3	-5.1
8	ROSC	2.0	✓	✓	✓	42.6	26.0	21.9	55.6	52.0	6.0	12.9	26.0	-15.9
9	STKA	0.9	✓			60.6	75.6	8.3	65.3	81.2	3.8	4.7	5.6	-4.4
10	ULM	2.0	✓	✓		60.1	84.9	6.6	60.8	85.3	6.6	0.6	0.4	0

acc - overall accuracy   aacc - accuracy of associated phases   Γ - percentage of associated phases mislabeled as noise

SNR_{m_b}- what does it do?

ROSC 1/1/2018 - 31/12/2019														
re-training											OPS baseline			
`noise-mbsnr=0.0`					`noise-mbsnr=2.0`						`noise-mbsnr=2.0`			
	N	T	P	S		N	T	P	S			N	T	P	S
N	**12138**	55	8	8	N	**14672**	179	9	17	N	**9482**	232	13	14
T	2882	**486**	22	1	T	1443	**413**	247	0	T	1123	**177**	11	1
P	1859	343	**91**	6	P	1664	300	**85**	4	P	1910	401	**96**	3
S	4892	121	3	**37**	S	3992	113	3	**31**	S	9256	195	4	**34**
acc: 55.6%; aacc: 52.0%					acc: 66.2%; aacc: 44.8%					acc: 42.6%; aacc: 26.0%				
$\Gamma = 6.0\%$					$\Gamma = 17.4\%$					$\Gamma = 21.9\%$				

Interchangeability of weights?

- New weights of each stations were used to classify 2018 data of all other stations:

Accuracy (%) - not StaPro results											
	DATA										
WEIGHTS		BOSA	CPUP	KEST	LPAZ	NRIK	PLCA	PPT	ROSC	STKA	ULM
	BOSA	**49.8**	26.2	33.7	27.7	17.8	59.2	69.4	43.8	69.	42.1
	CPUP	49.8	**69.3**	60.5	38.4	40.8	51.8	87.6	83.	68.8	49.7
	KEST	34.4	30.8	**29.5**	20.7	25.3	17.2	53.4	74.2	42.6	20.8
	LPAZ	60.1	38.5	40.5	**48.2**	25.3	45.8	54.6	38.8	62.5	38.5
	NRIK	52.1	26.9	29.5	27.2	**32.**	34.2	54.2	44.5	57.9	34.5
	PLCA	37.9	38.	33.1	24.9	22.3	**47.7**	63.4	29.3	57.	41.9
	PPT	41.	30.6	35.7	27.9	31.7	27.3	**78.3**	87.2	63.4	24.8
	ROSC	54.2	80.	69.4	64.3	57.8	48.5	52.4	**56.2**	65.5	60.8
	STKA	33.7	38.9	31.8	31.	44.6	54.	48.5	16.9	**62.6**	54.7
	ULM	29.1.	42.7	35.4	31.5	59.	45.	25.	11.	33.7	**61**

Interchangeability of weights?

- New weights of each stations were used to classify 2018 data of all other stations:

Γ (%) - not StaPro results											
	DATA										
		BOSA	CPUP	KEST	LPAZ	NRIK	PLCA	PPT	ROSC	STKA	ULM
WEIGHTS	BOSA	**5.2**	2.6	6.	4.	3.4	5.7	41.	39.7	10.3	5.3
	CPUP	14.5	**7.3**	13.8	8.3	8.1	6.3	33.7	52.1	10.3	6.5
	KEST	8.5	2.9	**3.4**	4.3	5.7	3.2	8.4	23.4	10.2	3.6
	LPAZ	20.	6.7	6.2	**3.6**	1.6	10.	14.	11.3	18.2	8.7
	NRIK	11.2	2.8	4.3	4.2	**1.1**	6.1	20.9	7.2	15.5	4.
	PLCA	3.9	2.8	6.	3.3	2.7	**2.5**	34.6	33.8	6.2	3.5
	PPT	8.	3.6	4.1	6.3	6.7	3.9	**10.8**	30.1	13.4	4.3
	ROSC	51.6	39.1	34.2	25.1	21.2	32.	15.4	**6.1**	46.2	41.
	STKA	3.4	5.4	11.3	6.3	13.3	5.	38.1	41.8	**8.**	6.8
	ULM	3.4	7.4	13.6	7.8	17.2	5.1	25.	14.5	7.1	**6.6**

Conclusion

- Benefits of re-training using stations specific data:
  - higher classification accuracy of not associated and associated phases and better Γ
  - simplification of the algorithm - removal of SNR_{m_b}(`noise-mbsnr=0.0`)
- Application of class_weights during training facilitates good results even with highly imbalanced data
  - It would be useful to have some insights from analysts on relative importance of various classes and types of errors
- Maybe some stations could be grouped together and share data/weights for training/classification not definitely not all of them
- The IWT classifier is deprecated and we could do much better with a new classifier based on convolutional networks
  - Also, dataset can be spoiled by many phases which should be associated but are not
- Would it be possible to evaluate influence (if any) on SEL3 in the spirit of NetVisa evaluation? How about arrays?