Learning to fingerprint : physical layer identification

Cyrille Morin¹

Thibaud Vial

Leonardo S. Cardoso¹ Jakob Hoydis² Marie Gorce¹

Jean

CITI Lab Inria

Nokia Bell Labs

02/07/2018

Objectives		
Motivations		

In IOT Context

- Constrained resources
- High cost of overhead

Implicit Identification

- Increase payload capacity
- ID spoofing made difficult
- May be privacy issue

information mathematics

Objectives			
Identify sou	urce of signal		

Possible identification sources

- Transmitter "voice" (RF layer noise)
- Channel characteristics

Why neural networks

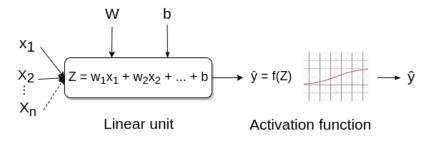
- Task close to Voice recognition problems
- Predominance of Deep learning in modern audio processing

The benefits of CorteXLab

- Online access to several nodes
- Promise of reproducible experiments
- Automatable experiments

Neural Networks 101		

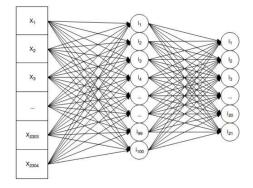
Basic building block : the unit



Approximate $f(x_1, x_2, ...) = y$ by $\hat{y} \approx y$



Multi Layer Perceptron (MLP) or plain neural network



Ínia-

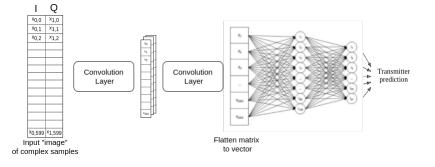
Objective	bjectives				Neural Networks 101							Re	esults				spectives	
Conv	Convolutionnal layer																	
Input matrix												Οι	ıtpı	ut n	nati	rix		
	37 81 55 12 65 85 99							_	·									
	31	91	32	92	70	10	34		F	ilte	r		237	-304	266	140	-426	
	92	73	90	66	80	10	63		-1	-1	-1		191	301	68	268	-272	
	27	10	20	35	41	48	6	*	-1	8	-1	=	-372	-195	-150	-68	-10	
	76	16	58	7	68	82	11		-1	-1	-1		-159	221	-303	163	272	
	29	27	40	88	9	71	97		Å	Â	1	1	-188	-70	379	-397	99	
	55 52 78 64 1 88 80					1	$\langle \rangle$	/										
									\mathbb{V}									

Weights learned

Ínita

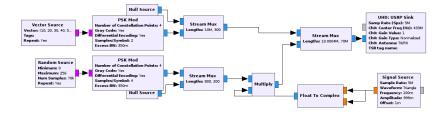
		Experiment	
NI	and the second		

Neural network architecture



Invasion

Generation of training data



Datasets

- Raw I/Q samples recorded by receiver for offline processing
- 21 transmitters and 50000 packets per transmitter
- 50000 * 21 = 1050000 examples in one dataset
- Dataset examples split in 70% training, 10% validation and 20% test sets

Training and testing process

Training process

- Offline
- 100 packet examples per batch
- At least 30000 batches per training

Testin<u>g</u>

- Over same dataset
- Across datasets

Show learning capability Show generalisation capability

Learning ability is good

Precision obtained by training and testing on same dataset										
Dataset ID	DS1	DS2	DS3	DS4	DS5					
Precision	94.23%	92.08%	96.33%	97.24%	95.03%					

Interestis activestis

Perspectives

Learning ability is good

										Cr	nfus	sion	mat	rix									300
	0.0	275	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		0	174	0	130	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		0	0	310	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	2.5	0	0	0	308	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	250
		0	0	0	0	326	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	5	0	0	0	0	0	301	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		0	0	0	0	0	0	301	0	0	0	0	0	23	0	0	0	0	0	0	0	0	
		0	0	0	0	0	0	0	272	0	0	0	0	0	0	0	0	0	0	0	0	0	200
True label		0	0	0	0	0	0	0	0	260	0	0	0	0	0	0	0	0	0	0	8	0	
ab		0	0	0	0	0	0	0	0	0	290	59	0	0	1	0	0	0	0	0	0	0	
5	10	0	0	0	0	0	0	0	0	0	1	270	0	0	0	5	0	0	0	0	0	0	
Ĕ		0	0	0	0	0	0	0	0	0	0	0	290	0	0	0	1	0	0	0	0	0	150
F		0	0	0	0	1	0	0	0	0	0	0	0	313	0	0	0	0	0	0	0	0	
		0	0	0	0	0	0	0	0	0	2	0	0	0	311	0	0	0	0	0	0	0	
		0	0	0	0	0	0	0	0	0	0	2	0	0	0	312	0	0	0	0	0	0	
	15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	271	0	0	0	0	0	100
		0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	268	0	0	0	0	
		0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	285	0	0	0	
		0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	321	0	0	
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	284	0	50
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	298.	30
		0.1					5	P	re	di	ct	ed	1	ab	el		15					20	

Invaria

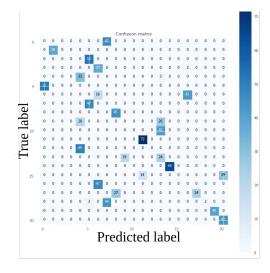
So is generalisation Or is it?

Dataset	Creation date					Test		
DS1	25/07			DS1	DS2	DS3	DS4	DS5
DS2	25/07		DS1	97.94	98.41	97.99	97.21	35.85
			DS2	96.22	97.02	96.32	96.28	36.79
DS3	26/07	Train	DS3	93.34	93.97	95.18	93.45	36.77
DS4	27/07		DS4	96.62	97.10	95.22	97.38	38.53
DS5	16/08		DS5	42.99	43.48	43.16	40.65	94.25

Invite notice the

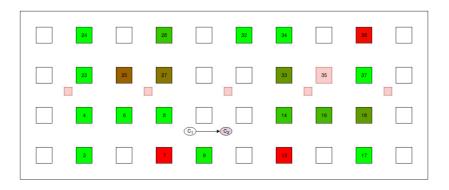
Perspectives

The chair crisis



In all and the

		Results	
The chair cri	isis		



identitis methemetics

We need more complete datasets





80% learning accuracy 70% generalisation

 With Channel variation

- Introduce robot in the room
- Combined with power variation

		Results	Perspectives
Conclusion			

For the neural network approach

- Reliable detection of transmitter
- Highly sensitive to radio channel characteristics

On CorteXLab's side

- Automatable experiment
- Stable channel over time

		Perspectives
Perspectives		

Improvements

- RF "Voice" identification
- Online identification

Beyond

Extension of Deep Learning to other telecommunication aspects

interesting mathematiks