Convolutional Recurrent Neural Networks For Computer Network Analysis

Jakub Nowak [0000–0002–1572–3426], Marcin Korytkowski [0000–0002–6002–2733], Rafal Scherer [0000–0001–9592–262X] **Computer Vision and Data Mining Lab, Institute of Computational Intelligence, Czestochowa University of Technology** Al. Armii Krajowej 36, 42-200 Czestochowa, Poland {jakub.nowak, marcin.korytkowski, rafal.scherer}@iisi.pcz.pl http://iisi.pcz.pl

Introduction

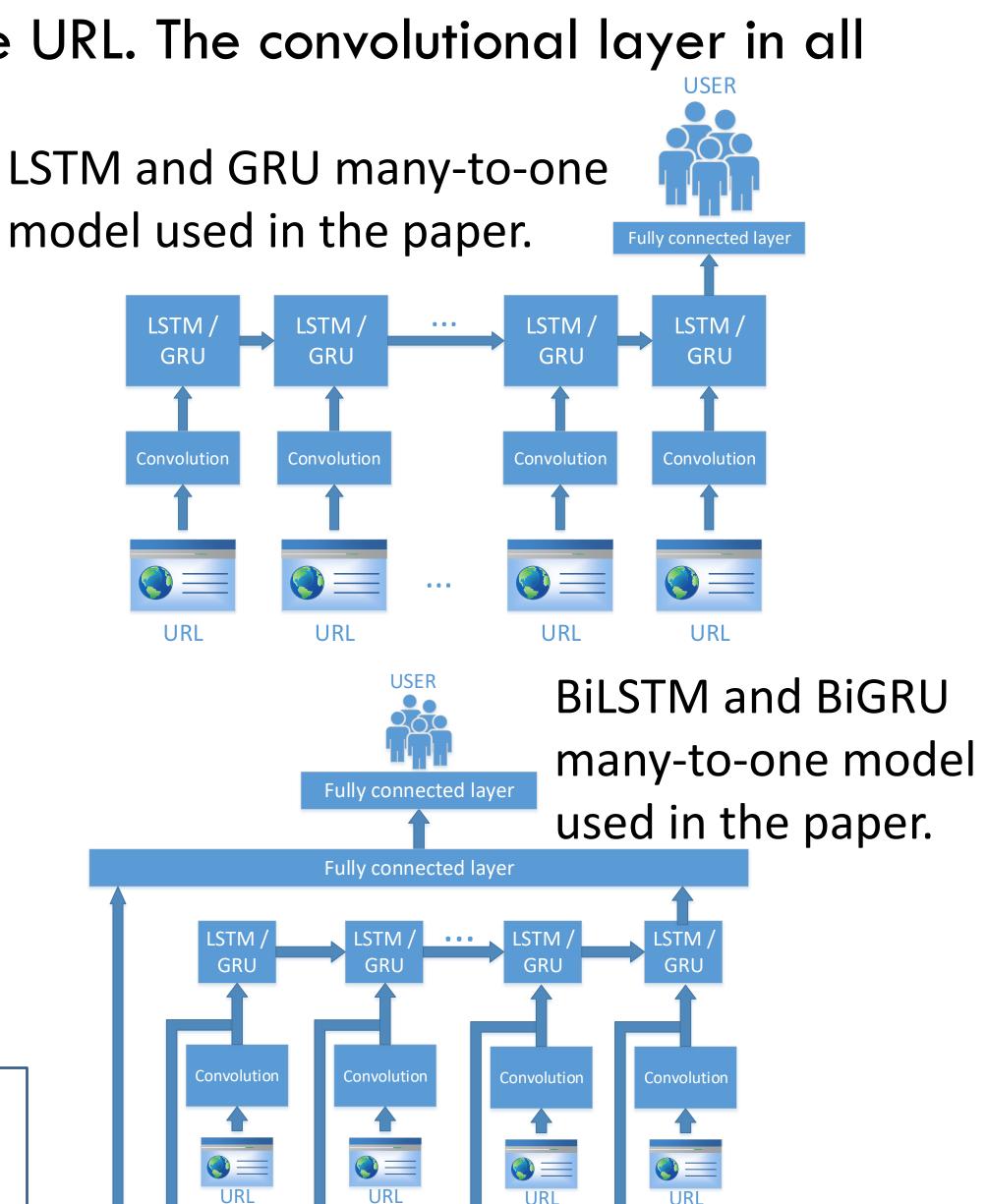
A method of computer network user detection

Neural Networks and Encoding

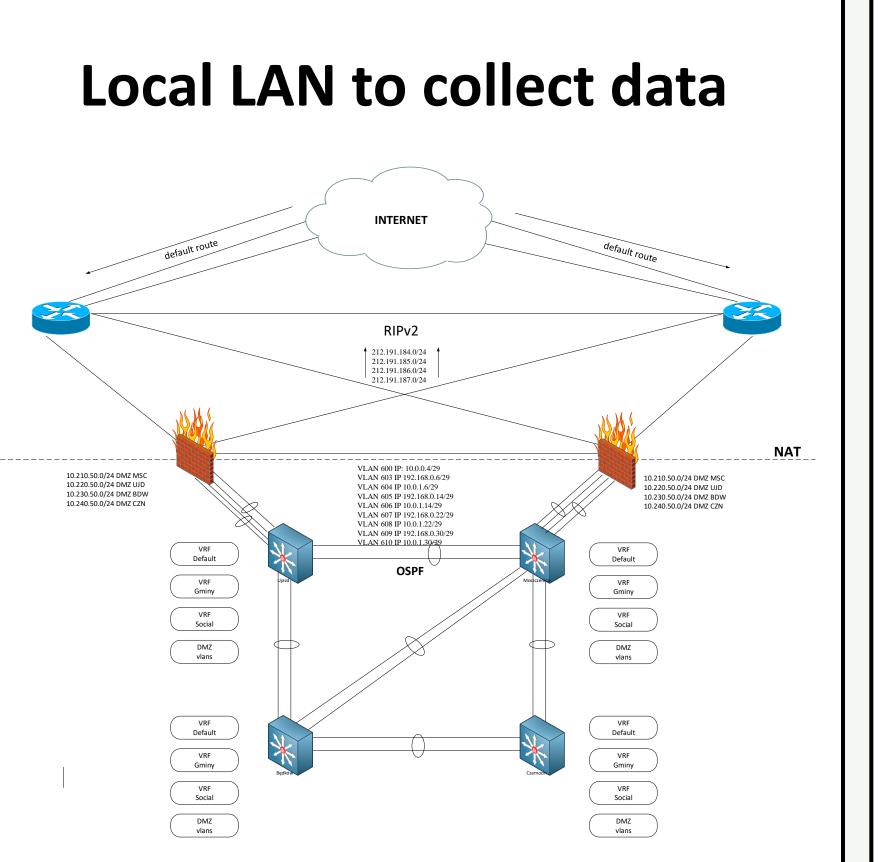
Four types of recurrent neural networks: LSTM, BiLSTM, GRU and BiGRU with convolutional input layers. The input data to the networks was a 70x45-pixel image with character-level one-hot encoding, as we have 70 possible

with recurrent neural networks. We use LSTM and gated recurrent unit neural networks. We added convolutional input layers. We transform requested URLs by onehot character-level encoding. The system was checked on real network data collected in a local municipal network. It can classify network users; hence, it can also detect anomalies and security

characters and up to 45 characters for one URL. The convolutional layer in all RNNs had two versions: Version 1: Input 45x70x1 Embedding 32 (45x32x1) Convolution 128 feature maps, filter 1D size 5 stride 1 Convolution 256 feature maps, filter 1D size 3 stride 1 Convolution 512 feature maps, filter 1D size 2 stride 1 MaxPooling 6 $512 \times 7 = 3584$ (input to LSTM) Version 2: Input 45x70x1 Embedding 32 (45x32x1) Convolution 64 feature maps, filter 1D size 5 stride 1 Convolution 128 feature maps, filter 1D size 3 stride 1 Convolution 256 feature maps, filter 1D size 2 stride 1 MaxPooling 6 2 $56 \times 7 = 1792$ (input to LSTM)



compromises



Example of encoded URL

Results

Testing error for all the neural networks with two variants of input convolutional layers. The networks were tested on future data, not used during training.

Network		Error %
	Version 1	Version 2
LSTM	28.90%	27.10%
BiLSTM	28.00%	28.20%
GRU	26.58%	26.60%
BiGRU	25.40%	26.20%

Conclusions

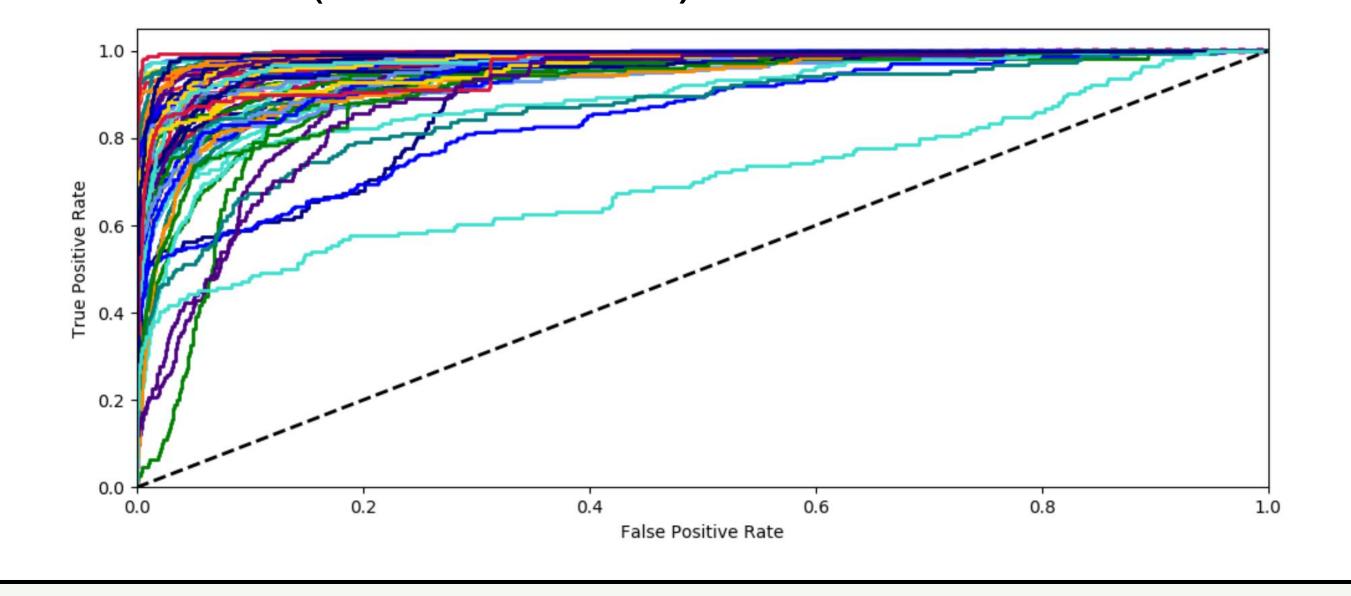
We showed that LSTM and GRU networks with input convolutional layers are suitable for identifying network users based on URLs they requesting.

The convolutional layers and one-hot encoding on the character level we applied, entirely replace the use of a dictionary or other ways of feeding text

Logs from Paloalto firewall:

Logs	(user src eq 'msclutter and a second by a								
Traffic		Receive Time	Category	URL	From Zone	To Zone	Source		
URL Filtering	ID	11/	web-	pagead2.googlesyndication.com/pagead	MSC-	OUTSIDE	10.10.20.31		
WildFire Submission	3	12 more	advertisements	Industry function to search the function of th	gminy-p2p				
Data Filtering	P	11/24 12:2000	web- advertisements	pagead2.googlesyndication.com/pagead	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
Alarms	P	11/ 12:	web- advertisements	pagead2.googlesyndication.com/pagead	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
App Scope	P	11/7 12:	web- advertisements	pagead2.googlesyndication.com/pagead	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
Change Monitor	P	11/ 12:	web- advertisements	googleads4.g.doubleclick.net/pcs/view?x	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
Threat Map	P	11/ 12:	content- delivery- networks	rfx.tagcdn.com/ifr/pl_pl/3390_284/170x	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
Contraffic Map	P	11/7 12	web- advertisements	googleads4.g.doubleclick.net/pcs/view?x	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
PDF Reports	P	11/7	content- delivery- networks	afx.tagcdn.com/ifr/pl_pl/3390_284/170x	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
Sa User Activity Report	P	11/	business-and- economy	csm.nl.eu.criteo.net/	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
Email Scheduler	P	11/1	computer-and- internet-info	js-agent.newrelic.com/	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
M Reports	Þ	11/20	internet- portals	accounts.google.com/	MSC- gminy-p2p	OUTSIDE	10.10.20.31		
	P	11/2	business-and- economy	gapl.hit.gemius.pl/	MSC- gminy-p2p	OUTSIDE	10.10.20.31		

ROC curves for all the users for the BiGRU network (the best one)



to recurrent networks. Such an approach is especially useful in the case of URLs which often do not use regular English (or any other language) words.

