Vilniaus universitetas

Doktorantas:
Paulius Vaitkevičius

Vadovas:
Dr. Virginijus Marcinkevičius

Antrųjų metų ataskaita
2020 m. spalio 21 d.

Mašininiu mokymusi grįstų atvirųjų šaltinių žvalgybos informacijos išskyrimo ir analizės metodai

Doktorantūros laikotarpis: 2018 - 2022

Vilniaus universitetas

TURINYS

1. Problemos apibrėžimas, tyrimo objektas, tikslai ir planuojami gauti rezultatai
2. Ataskaitinių metų darbo planas ir ataskaita
3. Kitų metų darbo planas
4. Trumpas per metus gautų mokslinių rezultatų pristatymas

PROBLEMOS APIBRĖŽIMAS, TYRIMO OBJEKTAS, TIKSLAI IR PLANUOJAMI GAUTI REZULTATAI

Vilniaus universitetas

Tyrimo tikslas

Sukurti apsimetinėjimo atakoms atsparų metodą, grįstą giliaisiais neuroniniais tinklais ir natūralios kalbos apdorojimo algoritmais, kuris leistų efektyviai ir patikimai atpažinti duomenų išviliojimo internete tinklapius.

Vilniaus universitetas

Tyrimo objektas

1. Mašininio mokymo ir giliojo mašininio mokymo algoritmai, skirti atpažinti duomenų išviliojimo internete (angl. „Phishing“) tinklapius.
2. Atsparūs priešiškoms atakoms algoritmai (angl. „Adversarial Machine Learning“).

Vilniaus universitetas

Tyrimo uždaviniai

1. Atlikti literatūros analizę, išanalizuoti state-of-the-art algoritmus duomenų išviliojimo internete tinklapių atpažinimui.
2. Atkartoti *state-of-the-art* algoritmų rezultatus.
3. Pasiūlyti naują efektyvesnį duomenų išviliojimo internete tinklapių atpažinimo metodą.
4. Sukurti duomenų rinkinius eksperimentų vykdymui.
5. Atlikti eksperimentinius tyrimus, palyginant pasiūlytą metodą su *state-of-the-art* algoritmais.

Vilniaus universitetas

Planuojami rezultatai

1. Atlikta **literatūros analizė**, palyginant pažangiausius tyrimo srities algoritmus;
2. Atlikti **eksperimentiniai tyrimai**:
   - ✓ Mašininio mokymosi algoritmų efektyvumo palyginimas;
   - ✓ Giliojo mašininio mokymosi (GMM) algoritmų efektyvumo palyginimas;
   - ✓ GMM algoritmų (RNN, LSTM, GRU, CNN, kt.) efektyvumo tyrimai, naudojant natūralaus teksto apdorojimo technikas (N-grams, word embeddings, kt.).
   - ❑ Naujo GMM algoritmo kūrimas, sprendžiant apibrėžtus uždavinius.
   - ❑ Pasiūlyto GMM algoritmo eksperimentinis tyrimas analizuojant jo efektyvumą;
   - ❑ Pasiūlyto GMM algoritmo atsparumo priešiškoms atakoms (angl. „Adversarial Machine Learning) eksperimentiniai tyrimai.

ATASKAITINIŲ METŲ DARBO PLANAS IR ATASKAITA

KITŲ METŲ DARBO PLANAS

Vilniaus universitetas

Rezultatai ir terminai	Komentarai	SM-I		SM-II		SM-III		SM-IV	
		S-1	S-2	S-3	S-4	S-5	S-6	S-7	S-8
DALYVAVIMAS KONFERENCIJOSE ir KT.									
1. Dalyvavimas konferencijoje Lietuvoje.	KODI-2019 / 2019-10-04								
2. Dalyvavimas konferencijoje Lietuvoje.	DAMSS-2019 / 2019-11-29								
3. Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.	DL-2019 / 2019-07-26, Varšuva								
4. Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.	DB&IS-2020 / 2020-06-19								
PLANUOJAMAS MOKSLINIŲ TYRIMŲ PUBLIKAVIMAS									
1. Mokslinių tyrimų disertacijos tema apžvalga (konferencijos darbų medžiagoje)	DAMSS-2019 Proceedings Vol 8 (2019)								
2. Teorinio tyrimo publikavimas (recenzuojamoje konferencijos darbų medžiagoje)	DB&IS-2020 Springer CCIS Proceedings								
3. Algoritmų palyginimo tyrimo rezultatų publikavimas (recenzuojamame leidinyje)	INFORMATICA Volume 31, Issue 1 (2020)								
4. Empirinio tyrimo rezultatų publikavimas (recenzuojamame leidinyje)									
STUDIJOS									
1. Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika									
2. Fundamentalieji informatikos ir informatikos inžinerijos metodai									
3. Didžiųjų duomenų analitika									
4. Mašininis mokymasis									
X. Bendruosius gebėjimus stiprinančios veiklos (3 kreditai)									
MOKSLINIŲ TYRIMŲ IR DISERTACIJOS RENGIMAS									
1. Mokslinių tyrimų disertacijos tema apžvalga ir analizė									
2. Mokslinio tyrimo vykdymas:									
2.1. Tyrimo metodikos sudarymas									
2.2. Teorinis tyrimas									
2.3. Empirinis tyrimas									
3. Atskirų daktaro disertacijos dalių parengimas									
4. Daktaro disertacijos parengimas ir svarstymas padalinyje									
5. Daktaro disertacijos gynimas									

pasiekti rezultatai

planuojami rezultatai naujaisiais m. m.

TRUMPAS PER METUS GAUTŲ MOKSLINIŲ REZULTATŲ PRISTATYMAS

Vilniaus universitetas

2019 - 2020 m. tyrimai

1. Sukčiavimo internete tinklapių atpažinimas, naudojant LSTM ir GRU

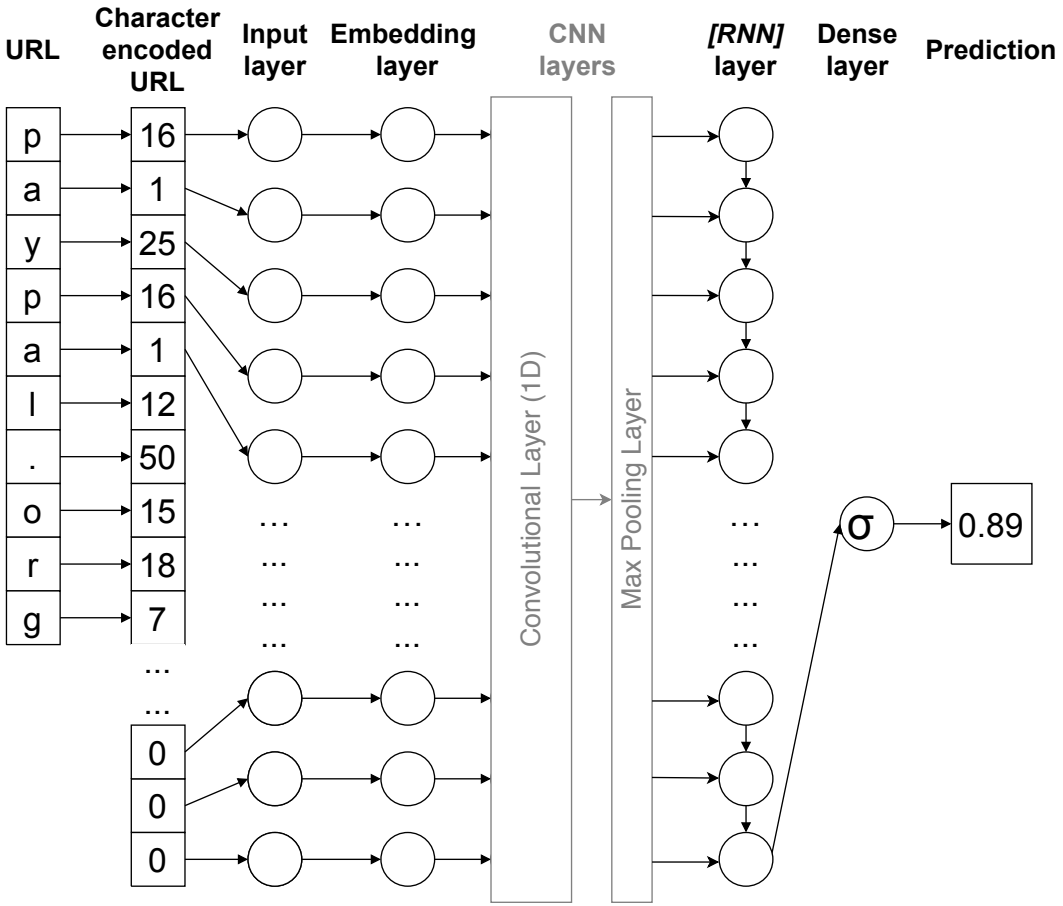
- Pristatyta DAMSS 2019
- Vaitkevicius, P., & Marcinkevicius, V. (2019). Learning Phishing Websites URLs Using Long Short-Term Memory Network and Gated Recurrent Units. *Proceedings of the 11th International Workshop "Data Analysis Methods for Software Systems," 8*, 88. https://doi.org/10.15388/proceedings.2019.8

2. RNN ansamblių kūrimas, sukčiavimo internete tinklapių atpažinimui

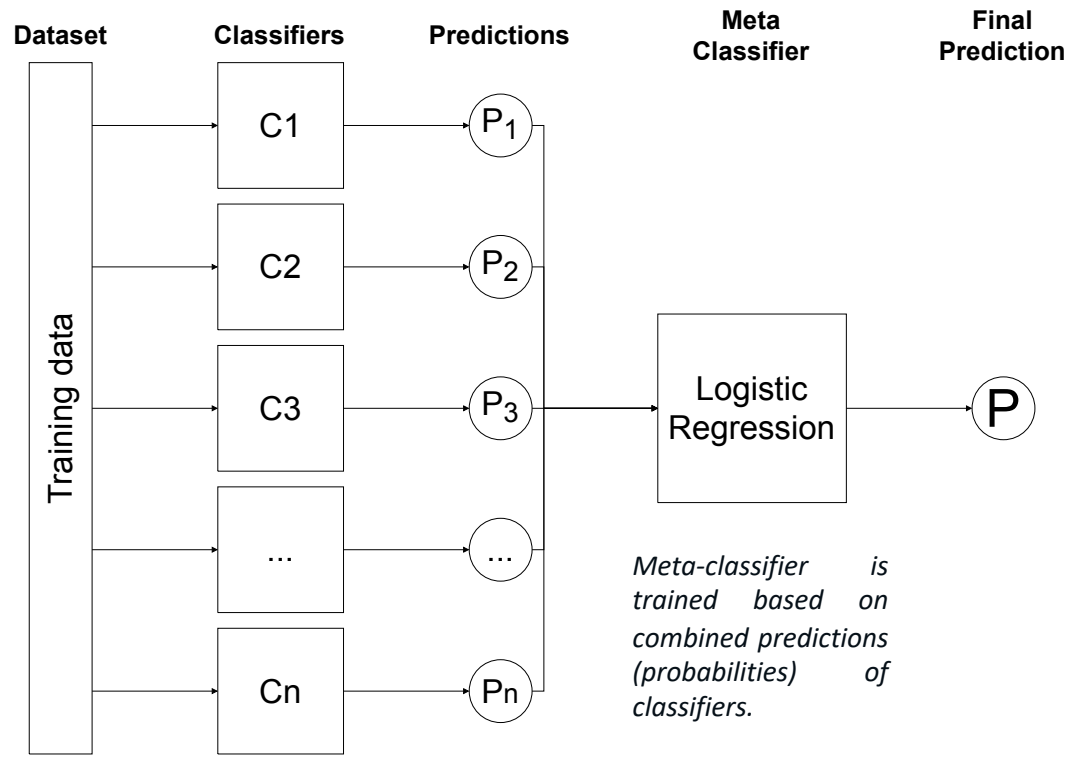
- Pristatyta DB&IS 2020
- Vaitkevicius, P., & Marcinkevicius, V. (2020). Composition of ensembles of recurrent neural networks for phishing websites detection. Communications in Computer and Information Science, 1243 CCIS, 297–310. https://doi.org/10.1007/978-3-030-57672-1_22

Algorithms used in the experiment

Architecture of our RNN (CNN) based methods



Architecture of our stacking ensemble

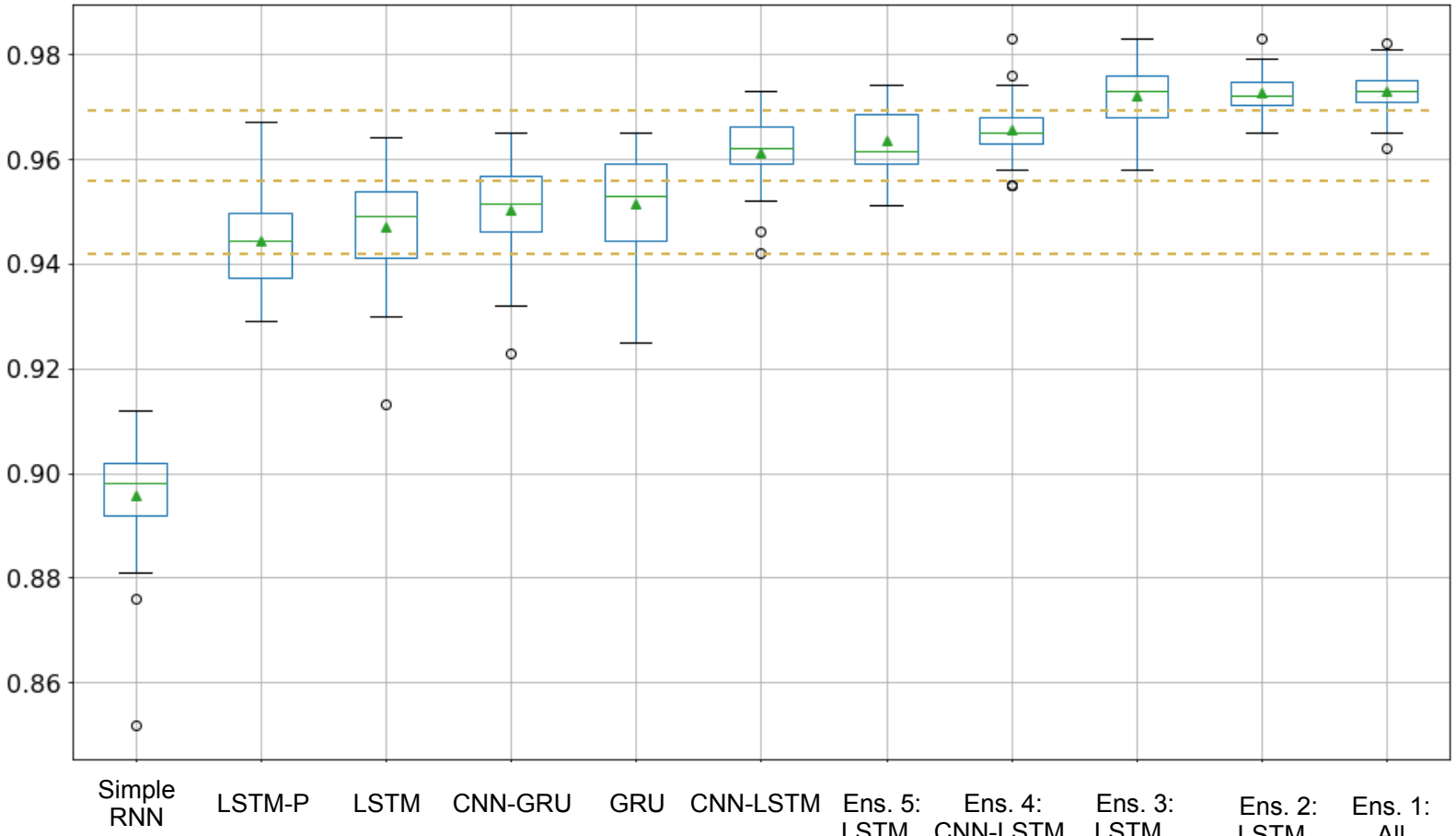


The dataset

Tan, Choon Lin (2018), “Phishing Dataset for Machine Learning: Feature Evaluation” (Universiti Malaysia Sarawak)

- Published: 2018-03-24
- DOI: 10.17632/h3cgnj8hft.1
- 10.000 websites
  (5.000 phishing + 5.000 legitimate):
  - URL
  - Human extracted 48 features

Results



Method	Accuracy
Gradient Tree Boosting	0.9742
Ensemble 1 (LSTM, LSTM-P, GRU, CNN-LSTM, CNN-GRU, Naïve-Bayes)	**0.9730**
AdaBoost	0.9728
Ensemble 2 (LSTM, LSTM-P, GRU, CNN-LSTM, CNN-GRU)	**0.9725**
Ensemble 3 (LSTM, GRU, CNN-LSTM, CNN-GRU)	**0.9721**
Random Forest	0.9715
Multilayer Perceptron	0.9671
CNN-LSTM	**0.9612**
Classification and Regression Trees	0.9574
Ensemble 4 (CNN-LSTM, CNN-GRU, Naïve-Bayes)	**0.9657**
Ensemble 5 (LSTM, LSTM-P, GRU)	**0.9634**
Support Vector Machine	0.9549
GRU	**0.9515**
CNN-GRU	**0.9503**
LSTM	**0.9471**
LSTM-P	**0.9443**
Naïve-Bayes	0.9177
SimpleRNN	**0.8958**
Naïve-Bayes (this experiment)	**0.6056**

Conclusions

1. **Our proposed method**, employing RNN-based ensembles, outperform single RNN methods by at least **0.02** difference in classification accuracy, which is statistically significant.
2. Our proposed RNN ensemble-based method on URL character sequence performs **as well** as classical ensembles with human extracted features on the same dataset.
3. Adding the CNN layer to the method increases classification accuracy by **0.01**, and this difference is statistically significant.
4. For phishing websites' URL classification problem, RNNs with **explicit memory** implementation, like LSTM and GRU, outperform classic RNNs without memory by **0.06** difference in classification accuracy, which is statistically significant.

Vilniaus universitetas

AČIŪ UŽ DĖMESĮ

Paulius Vaitkevičius
VU DMSTI doktorantas
+370 650 83623
paulius.vaitkevicius@mif.vu.lt

ORCID