Generation and Classification of Illicit Bitcoin Transactions

Pablo de Juan Fidalgo, Carmen Cámara and Pedro Peris López
Universidad Carlos III de Madrid (UC3M) – 2nd Dec 2022

Index

1. Introduction and background
2. Balancing the data set
3. Classification of illicit transactions
4. Experiments and results
5. Conclusions and future work

Index

1. **Introduction and background**
2. Balancing the data set
3. Classification of illicit transactions
4. Experiments and results
5. Conclusions and future work

Introduction and background

- Criminals usually seek for a financial reward
- Authorities try to follow the trace of the money
- Bitcoin is created in 2008
  - Decentralized monetary system
  - Anonimity
  - Publicly available
  - Banking institutions and Law Enforcement Agencies lost power
- AI to perform network forensics

Introduction and background

- Scarcity of labelled data
- Collection of 13,500 Bitcoin addresses related to illegal behaviour
- Elliptic Data Set
  - Result of a research from IBM, MIT and Elliptic professionals
  - More than 200,000 transactions
  - Licit tx (42,019) versus illicit tx (4,545) and unknown tx
  - 166 features (94 local features + 72 aggregated features)

Index

1. Introduction and background
2. **Balancing the data set**
3. Classification of illicit transactions
4. Experiments and results
5. Conclusions and future work

Balancing the data set

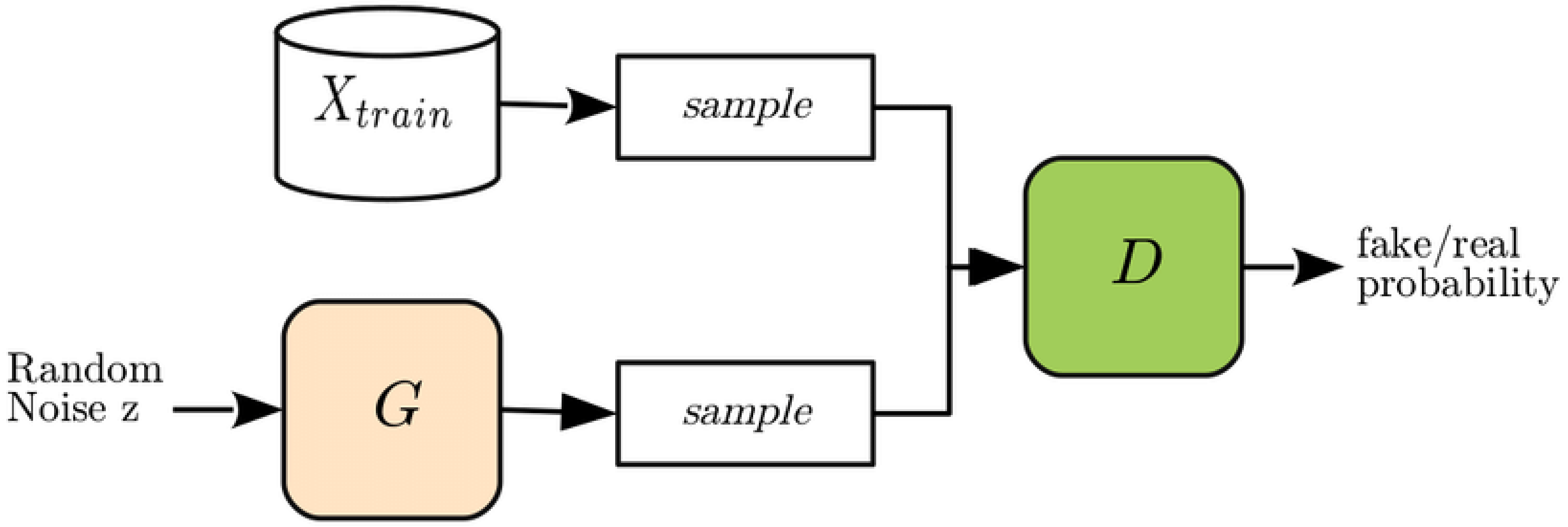
Natural generation of data

- 13,500 illicit Bitcoin addresses
- Kaggle user de-anonymized Elliptic Data Set (1st Jan, 2016 to 2nd Oct, 2017)
- Our work: 25,000 new illicit transactions in that timespan
- Before: 9.8:90.2 illicit/licit ratio
- After: 41.2:58.8 illicit/licit ratio

Balancing the data set

Synthetic generation of data

- Oversampling
- Undersampling
- Generative Adversarial Networks
  - TGAN
  - 25,000 synthetic samples



Index

1. Introduction and background
2. Balancing the data set
3. **Classification of illicit transactions**
4. Experiments and results
5. Conclusions and future work

Classification of illicit Bitcoin transactions

Machine learning

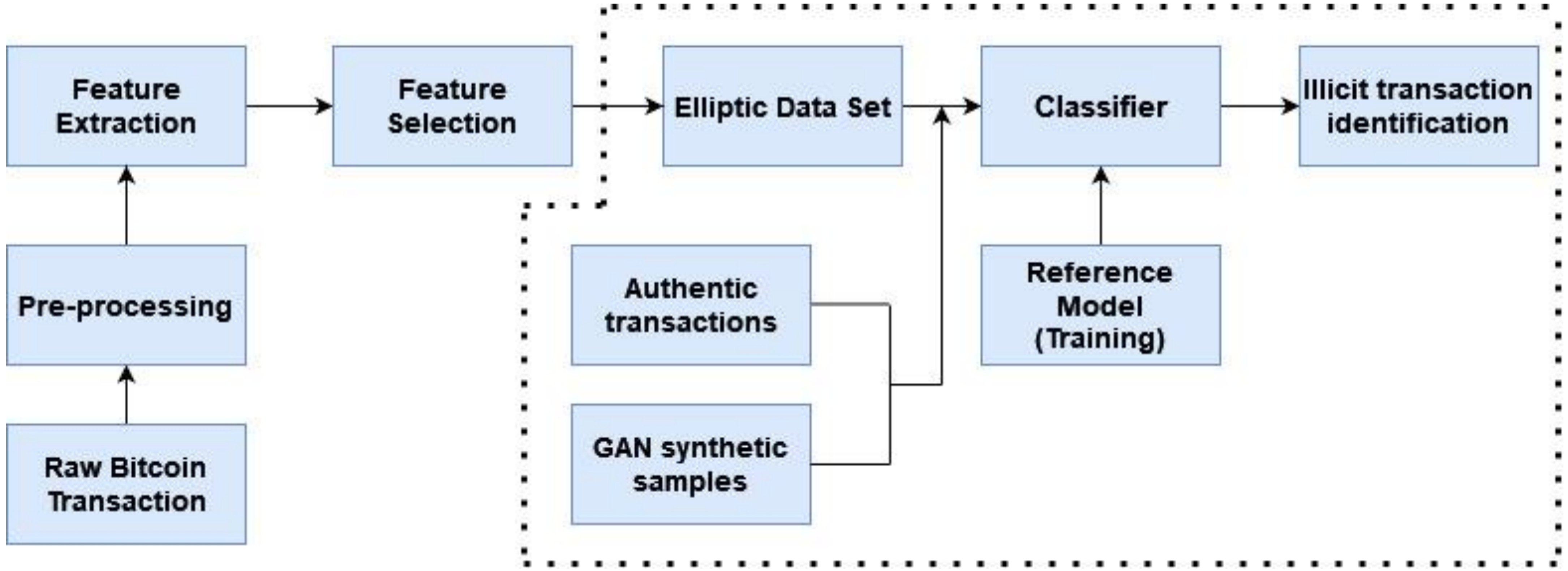
- Branch of AI which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy

- Process can be divided in:

  - Decision process

  - Error function

  - Model optimization process



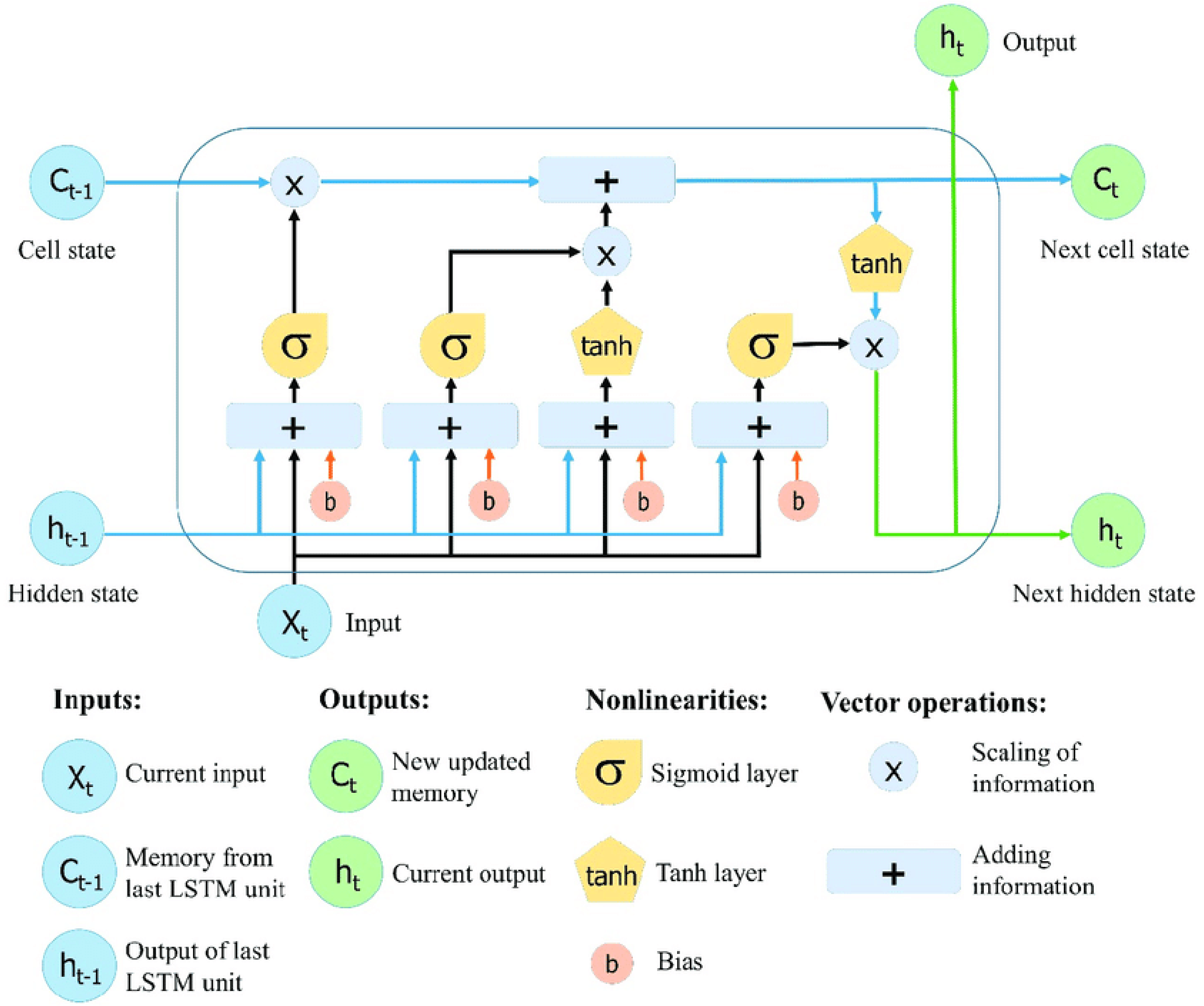
- Random Forest

- Logistic Regression

Classification of illicit Bitcoin transactions

Deep learning

- DL shows better performance than ML
- ANN model human brain structure
  - CNN
  - RNN → LSTM
- LSTM
  - Forget gate
  - Input gate
  - Output gate



Index

1. Introduction and background
2. Balancing the data set
3. Classification of illicit transactions
4. **Experiments and results**
5. Conclusions and future work

Experiments

- All features were used (local + aggregated)
- Random Forest and Logistic Regression
- GCN as DL solution? LSTM!

Experiments

Results with ML

Method	Precision	Recall	F1
Elliptic RF	0.956	0.670	0.788
RF with natural tx	0.985	0.962	0.974
RF with synthetic tx	0.999	0.983	0.991
Elliptic LR	0.404	0.593	0.481
LR with natural tx	0.784	0.824	0.804
LR with synthetic tx	0.951	0.961	0.956

Experiments

Results with DL

Method	Precision	Recall	F1
Elliptic GCN	0.812	0.512	0.628
Elliptic Skip-GCN	0.812	0.623	0.705
Elliptic EvolveGCN	0.850	0.624	0.720
LSTM with Elliptic data	0.908	0.855	0.868
LSTM with natural tx	0.947	0.927	0.934
LSTM with synthetic tx	0.991	0.981	0.985

Index

1. Introduction
2. Background
3. Generation of data
4. Classification of data
5. Experiments and results
6. **Conclusions and future work**

Conclusions

- Reduction of class imbalance in Elliptic data set with nearly 25,000 illicit tx
- GANs are a fantastic synthetic solution for unbalanced data sets
- Better data set >>> better algorithm
- LSTM is a strong alternative for binary classification with time-series data

Future work

- Reverse engineering of the features
- Hyperparameter tuning for a more powerful machine
- Generation of samples with WGAN model

Generation and Classification of Illicit Bitcoin Transactions

Thank you for your attention!

Questions?

Pablo de Juan Fidalgo, Carmen Cámara and Pedro Peris López
Universidad Carlos III de Madrid (UC3M) – 2nd Dec 2022