ModelChain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks

Tsung-Ting Kuo, Chun-Nan Hsu, and Lucila Ohno-Machado

pSCANNER Face-to-Face Meeting
October 13, 2016

Office of the National Coordinator for Health IT
Predictive Modeling

- Healthcare predictive modeling
  - Machine learning from healthcare data to predict outcomes

- Cross-institutional healthcare predictive modeling
  - More generalizable models
  - Comparative effectiveness research, biomedical discovery, patient-care, etc.
  - Especially for large research network like pSCANNER

Site 1

Patient Records = 10

Hard to predict outcomes

Site 1: Records = 10

Site 2: Records = 1,500

Site 3: Records = 760

Site 4: Records = 380

Feasible to predict outcomes

Patients

Patients
Protecting Privacy of Individuals

- Challenge of cross-institutional predictive modeling
  - Improper disclosure of protected health information (PHI)
- Privacy-preserving algorithms transfer models but not PHI
  - Many methods exist [Wu et al. 2012] [Li et al. 2015] [Wang et al. 2013] [Yan et al. 2013]

Site 1
Site 2
Share observation-level patient healthcare data

Site 3
Site 4
Privacy-preserving predictive modeling algorithms

Site 1
Site 2
Share partially-trained machine learning models

Site 3
Site 4
Centralized architecture

- Institutional policies
- Single-point-of-failure/breach
- Sites cannot join/leave at any time
- Mutable data and records
- Consensus/synchronization issues
The Blockchain Technology

- Desirable features
  - Decentralized architecture
    - Peer-to-peer
    - Sites keep full control of resources
    - No risk of single-point-of-failure
  - Sites can join/leave freely
    - No central server overhead
    - No disruption of learning process
  - Immutable audit trail
    - Tampering is difficult
1. Privacy-preserving online machine learning on blockchains
2. Transaction metadata to transfer partial models and info
3. Proof-of-information algorithm to decide order of learning
Summary

- Blockchain infrastructure as in Bitcoin
- ModelChain: distributed predictive models across institutes
- Privacy protection by exchange of model, not patient data
- No single-point-of-failure
- Verified data provenance
- Secure record of transactions
- Improved interoperability
- White paper and full slides: [https://healthit.gov/blockchain](https://healthit.gov/blockchain)
Acknowledgement

- The authors would like to thank Xiaoqian Jiang, PhD, and Shuang Wang, PhD for very helpful discussions.
- All authors are funded by PCORI CDRN-1306-04819.
- LO-M is funded by NIH U54HL108460, UL1TR001442, and VA I01HX000982.

Tsung-Ting Kuo PhD
Chun-Nan Hsu PhD
Lucila Ohno-Machado MD, PhD
Xiaoqian Jiang PhD
Shuang Wang PhD
Thank you!

Questions?