COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Lalitha Sankar, (Arizona State University), April 2021

Title: FACT: Federated Analytics based Contact Tracing for COVID-19

Lalitha Sankar CIC Database Profile

NSF Award # : 2031799

YouTube Recording with Slides

April 2021 CIC Webinar Information

Transcript Editor: Shikhar Johri

---

Transcript

ललिता शंकर:

स्लाइड *1*

सभी को नमस्कार, और आज मुझे यह अवसर देने के लिए हेलेन और फ्लोरेंस को धन्यवाद। तो यह एक NSF, अनुदान, मेरे भयानक सहयोगियों, Ni Trieu, Ming Zhao, और Visar Brisha के साथ SATSI के माध्यम से वित्त पोषित एक रैपिड अनुदान है। नी और मिंग कंप्यूटर साइंस में हैं और विसार कॉलेज ऑफ हेल्थ सॉल्यूशंस में हैं। जैसा कि नाम से पता चलता है, हम मशीन लर्निंग में बहुत गहरा गोता लगाने जा रहे हैं, और इसका उपयोग अधिक समग्र संपर्क अनुरेखण करने के लिए करते हैं।

स्लाइड *2*

तो वास्तव में हमारे काम के तीन पहलू हैं, हम ब्लूटूथ और जीपीएस दोनों का उपयोग करके मौजूदा संपर्क अनुरेखण प्रोटोकॉल को बढ़ाने की कोशिश कर रहे हैं, और मैं एक पल में इसमें गोता लगाऊंगा। यहां बड़ा विचार यह कहना है कि क्या आप उन सर्वेक्षणों पर सवारी या पिग्गीबैक कर सकते हैं जो संयुक्त राज्य भर में परिसरों में किए जा रहे हैं, विशेष रूप से एएसयू में। हमारे पास एक दैनिक स्वास्थ्य सर्वेक्षण है जिसे हमें लेना है। क्या आप उस तरह के सर्वेक्षण डेटा का उपयोग कर सकते हैं, इसका उपयोग आधारभूत जोखिम मॉडल बनाने के लिए कर सकते हैं, और फिर इसे गतिशीलता और यहां तक कि फोनेशन के साथ जोड़ सकते हैं, जो एक बायोमार्कर है? नाक की बात की तरह, यह पता चला है कि फोनेशन श्वसन और यहां तक कि मानसिक स्वास्थ्य के मुद्दों के पूरे समूह के लिए एक बायोमार्कर है, और इसे गर्म स्पॉट मॉडल बनाने के लिए गठबंधन करता है, और इसी तरह और आगे, और जोखिम का मूल्यांकन करना जारी रखता है और एक जोखिम स्कोर डालता है, और इस काम में हमारा व्यापक लक्ष्य वास्तव में है, इस निर्मित प्रणाली को लें और इसे एएसयू के मोबाइल ऐप में तैनात करें, और यह एक सतत प्रक्रिया है।

स्लाइड *3*

तो बहुत संक्षेप में संपर्क अनुरेखण पर।

स्लाइड *4*

मौजूदा संपर्क अनुरेखण ऐप्स मूल रूप से ब्लूटूथ का उपयोग करते हैं, ऐसे टोकन हैं जिनका उपयोग किया जाता है, मेरे पास इसमें जाने का समय नहीं है, लेकिन वे सुरक्षा हमलों की चपेट में हैं। वे भी, क्योंकि, नहीं करते हैं, क्योंकि वे जीपीएस नहीं करते हैं, वे सेल टावरों और इस तरह की चीजों को छोड़कर संक्रमित हॉटस्पॉट स्थानों का मूल्यांकन नहीं कर सकते हैं। इसलिए वे वास्तव में बहुत सारे समृद्ध उपयोगकर्ता विशिष्ट डेटा का शोषण नहीं कर रहे हैं, और इस परियोजना में हमारा लक्ष्य-

स्लाइड *5*

-वास्तव में जीपीएस के साथ ब्लूटूथ को संयोजित करना था जो मौजूद है, लेकिन हम इस प्रोटोकॉल का निर्माण करने, एंड्रॉइड डिवाइस पर इसका परीक्षण करने और मजबूत गोपनीयता गारंटी प्रदान करने में सक्षम हैं। और दूसरा दुष्प्रभाव यह है कि अब हम पीछे के कई सेवर्स पर सुरक्षित एकत्रीकरण तकनीकों का उपयोग करके हॉटस्पॉट के हिस्टोग्राम की गणना कर सकते हैं, [अस्पष्ट]।

स्लाइड *6*

ठीक है तो यह संपर्क अनुरेखण भाग पर है। डिवाइस के हिस्से पर।

स्लाइड *7*

हमारा लक्ष्य वास्तव में एक जोखिम भविष्यवाणी मॉडल के साथ आना है और हम क्या करने जा रहे हैं, हम एक आधारभूत जोखिम चाहते हैं जो आपके स्वास्थ्य लक्षणों पर आधारित है, लेकिन दैनिक लक्षणों का उपयोग लगातार इस जोखिम का आकलन करने के लिए किया जाएगा। गतिशीलता पैटर्न का उपयोग कर सकते हैं, विशेष रूप से एक विश्वविद्यालय परिसर में यह पता लगाने के लिए कि क्या किसी छात्रावास या किसी भी स्थान पर प्रकोप है, हम कैसे, आप जानते हैं, यहां तक कि यातायात को भी स्थानांतरित कर सकते हैं, और अंततः हमारा लक्ष्य वास्तव में फोनेशन आधारित जोखिम सूचकांक करना है और मेरे पास इस बारे में बहुत विस्तार से बात करने का समय नहीं होगा, लेकिन पूरी बात यह है कि सांस लेने के लिए, आप जानते हैं, हमारे ऐप का उपयोग करने के लिए, "आह" कहने के लिए, आप जानते हैं, कुछ सेकंड, और इसका उपयोग श्वसन दोनों को देखने के लिए करें, श्वसन स्वास्थ्य के मुद्दे सीधे सहसंबद्ध हैं कि आप कैसे फोनेट करते हैं और यह भी स्पष्ट रूप से सहसंबद्ध है मस्तिष्क कोहरे के साथ अब। इसलिए यहां बहुत शोध किया जा रहा है, और विसार, मेरे सहयोगी और मैं इस पर काम कर रहे हैं। ठीक।

स्लाइड *8*

इसलिए, मैं समय के हित में केवल एक चीज में गहरा गोता लगाने जा रहा हूं, जो है, आप कैसे करते हैं, आप जोखिम की भविष्यवाणी करने के लिए सर्वेक्षणों का उपयोग कैसे करते हैं? और हम बहुत भाग्यशाली रहे हैं क्योंकि इस अनुदान और Google के AI से सोशल गुड के लिए अनुदान के लिए धन्यवाद, हम एक प्रतियोगिता में भाग लेने में सक्षम थे जिसे फेसबुक ने होस्ट किया था और फेसबुक से एक डेटा सेट प्राप्त किया था जिसे सीएमयू डेल्फी उनके लिए एकत्र कर रहा है फेसबुक ऐप के माध्यम से और मूल रूप से यह एक सर्वेक्षण है, यह एक दैनिक सर्वेक्षण है और उनके पास 53,000 प्रतिभागियों के साथ 18 मिलियन प्रतिक्रियाएं हैं और पूरा विचार भी लक्षणों पर आधारित सर्वेक्षण, पूर्व स्वास्थ्य स्थितियों, सोशल डिस्टेंसिंग, मानसिक स्वास्थ्य जनसांख्यिकी, और सभी के बारे में

केवल 900,000 हैं जो लोग हैं जिन्होंने परीक्षण लिया है। तो यह वह डेटा है जिसका हम उपयोग करते हैं क्योंकि यही वह लेबल है जिसका उपयोग हम यह अनुमान लगाने के लिए करते हैं कि कैसे- क्या कोई, या आप जानते हैं, आपके लक्षणों के आधार पर, चाहे आपको COVID का खतरा हो।

स्लाइड *9*

तो हम क्या कर रहे हैं, आप जानते हैं, मैं थोड़ा गहराई से गोता लगा रहा हूं कि हम किस प्रकार के मॉडल का उपयोग कर रहे हैं। हम XGBoost का उपयोग करने जा रहे हैं क्योंकि यह थोड़े स्वास्थ्य डेटा सेट है। यह एक मिश्रण मिला है, आप जानते हैं, असतत और निरंतर मूल्य डेटा। लेकिन XGBoost एक शानदार, बहुत प्रसिद्ध मजबूत एल्गोरिथम है। हम जो करने जा रहे हैं, वास्तव में, हमारे द्वारा विकसित किए गए हानि कार्यों की एक पूरी कक्षा का उपयोग करके इसे और भी मजबूत बना दिया है और मुझे इसे ऑफ़लाइन लेने में खुशी हो रही है।

स्लाइड *10*

संक्षेप में, हम जो दिखाने में सक्षम हैं, वह यह है कि यहां तक कि खुद को शीर्ष आठ लक्षणों तक सीमित करना, और यह एक बहुत ही असंतुलित डेटा सेट है, हमारे पास 86% नकारात्मक केवल 14% सकारात्मक लेबल हैं, जिन्होंने परीक्षण किया है। हम अपने नुकसान समारोह का उपयोग करके मौजूदा XGBoost को और भी बेहतर चीजों के साथ बढ़ा सकते हैं।

स्लाइड *11*

इससे भी अधिक दिलचस्प, हम जो करने में सक्षम हैं, वह वास्तव में, सर्वेक्षण डेटा बहुत शोर है। हमने बहुत सारी प्री-प्रोसेसिंग की। लेकिन क्या आप वास्तव में, लगातार ऐसा कर सकते हैं जबकि डेटा शोर है? इसलिए हमने शोर लेबल के खिलाफ हमारे एल्गोरिथ्म का परीक्षण किया, जहां हमने इसे रखे गए लेबल का एक गुच्छा फ़्लिप किया, हमने दो प्रयोग किए, स्तरीकृत और अस्तरीकृत किया, हमने असंतुलन रखा और हमारे पास कुछ बहुत अच्छे परिणाम हैं।

स्लाइड *12*

तो समय के हित में, फिर से, मैं इसे वहीं रोकने जा रहा हूं और इसे ऑफ़लाइन ले जाऊंगा, हमारे पवित्र, आप इस परियोजना की कब्र जानते हैं, अगर मैं कर सकता हूं, तो इन तीनों चीजों को संयोजित करना है, एएसयू ऐप पर डालना और उपयोगकर्ता को सर्वेक्षण जोखिम देने में सक्षम होना, और एएसयू के लिए बहुत सारे बैक एंड डेटा संग्रह करना।

स्लाइड *13*

इसलिए हम इससे थोड़ा दूर हैं, लेकिन हमारे पास हमारा आईआरबी स्वीकृत है, इसलिए हम एक ऐप डालने और सामान करने जा रहे हैं। इसलिए मैं वहीं रुकने जा रहा हूं। बहुत-बहुत धन्यवाद।