COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Judy Fox (University Virginia) January 30, 2024

Title: Expeditions: Collaborative Research: Global Pervasive Computational Epidemiology

Judy Fox CIC Database Profile

NSF Award #: 2151597

YouTube Recording with Slides

Winter 2024 CIC Webinar Information

Transcript Editor: Lauren Close

Transcript

स्लाइड *1*

धन्यवाद, लॉरेन, आपके दयालु परिचय के लिए, और 2024 COVID सूचना कॉमन्स मेजबानों और हमारे साथी वक्ताओं, छात्रों और कर्मचारियों को धन्यवाद। मैं आज हमारे एनएसएफ वित्त पोषित अभियान परियोजना पर रिपोर्ट करने जा रहा हूं। यह वैश्विक व्यापक कम्प्यूटेशनल महामारी विज्ञान के लिए एक सहयोगी शोध है। मैं जूडी फॉक्स हूँ। मैं वर्जीनिया विश्वविद्यालय में डेटा साइंस स्कूल और कंप्यूटर विज्ञान में काम कर रहा हूं।

स्लाइड *2*

इस परियोजना की पृष्ठभूमि का एक छोटा सा हिस्सा - यह डॉ माधव मराठे के नेतृत्व में एक बहु-संस्थागत परियोजना है और वह समन्वयक पीआई हैं। हमारे पास वर्जीनिया विश्वविद्यालय और बायोकॉम्पलेक्सिटी इंस्टीट्यूट के सहयोगी हैं। यह एक बहु-संस्थागत सहयोग है और हमारे पास अद्भुत सहयोगी और शोधकर्ता हैं। यह मेरे लिए काफी प्रेरणादायक अनुभव है।

स्लाइड *3*

मैं भविष्य के बारे में बात करना चाहता था। यह आज मेरी प्रस्तुति के बारे में रोमांचक हिस्सा है। 2025 में COVID की कल्पना करें। हम क्या होंगे? हम हस्तक्षेप से रोकथाम की ओर बढ़ना चाहते हैं क्योंकि संक्रामक रोग एक ऐसी सामाजिक समस्या है। 2050 तक, हर साल 10 मिलियन से अधिक मौतों का अनुमान होगा और आर्थिक प्रभाव सौ ट्रिलियन डॉलर से अधिक होगा। कुछ साल पहले, हम इस महामारी से बाहर आए, जिसमें अकेले अमेरिका में 1.1 मिलियन से अधिक लोग मारे गए थे और 100 मिलियन से अधिक मामले हैं। यह संक्रमित परिवारों का लगभग एक-तिहाई है। बेहतर सूचित सरकारी नीति के

माध्यम से इसमें से अधिकांश को रोका जा सकता था। हालाँकि, COVID-19 एक जटिल डेटा समस्या है। सबसे पहले, हमारे पास गैर-स्थिर डेटा आ रहा है। डेटा की कमी, शोर डेटा के साथ रुझानों को सीखना और भविष्यवाणी करना काफी मुश्किल है। संक्रमण के पूर्वानुमान को समझने योग्य या समझाने योग्य बनाने से भी निर्णय लेने में मदद मिल सकती है। वे महत्वपूर्ण भौगोलिक और लौकिक फोकस क्षेत्रों की पहचान करने में मदद कर सकते हैं ताकि हम सरकारों को बीमारी के प्रसार को रोकने के लिए संसाधनों का अधिक प्रभावी आवंटन करने के लिए संकेत दे सकें।

स्लाइड *4*

मैं अपने समूह में किए गए शोध के बारे में बाकी बातों पर ध्यान केंद्रित करना चाहता हूं। हम अमेरिका में काउंटी स्तर के COVID-19 संक्रमणों की व्याख्या करना चाहते थे। हमने ट्रांसफॉर्मर एआई मॉडल लागू किया, जो बड़े भाषा मॉडल द्वारा उपयोग किया जाने वाला एक प्रकार का गहन शिक्षण मॉडल है।

स्लाइड *5*

जिन क्षेत्रों पर हम ध्यान केंद्रित कर रहे हैं उनमें से एक सवाल पूछना है: हमें भविष्यवाणी की आवश्यकता क्यों है? वास्तविक समय डेटा का उपयोग करने वाली भविष्यवाणियों को 2009 में डॉ हार्वे वी फाइनबर्ग और डॉ मैरी एलिजाबेथ विल्सन द्वारा हाइलाइट किया गया है। वे इस महत्वपूर्ण पहलू को आकर्षित करते हैं कि यह क्यों महत्वपूर्ण है - हम नवीनतम डेटा का उपयोग करके अध्ययन करना चाहते हैं - रोग नियंत्रण का अध्ययन करने और निरीक्षण करने और भविष्यवाणी करने का प्रयास करने के लिए। हस्तक्षेप चरम बिंदु पर कार्रवाई होगी, लेकिन हस्तक्षेप समय से पहले वक्र को समतल कर देगा।

स्लाइड *6*

हम डीप लर्निंग मॉडल, टेम्पोरल फ्यूजन ट्रांसफॉर्मर (टीएफटी) का उपयोग करते हैं, और यह मॉडल वास्तविक समय में पूर्वानुमान लगा सकता है। अपने प्रयोगों में, हम पिछले 13 दिनों का उपयोग भविष्य के 15 दिनों की भविष्यवाणी करने के लिए करते हैं. डेटा संग्रह विभिन्न भविष्यवाणी मॉडल से आता है। हम उन्हें स्थिर सह-प्रकार और गतिशील डेटासेट में वर्गीकृत करते हैं, जैसे मामले और मौतें। हमारे पास कुछ ज्ञात इनपुट भी हैं जैसे कि आगामी क्रिसमस और नए साल की छुट्टियां।

स्लाइड *7*

इस तरह के एक मॉडल के साथ, हमारा लक्ष्य यह समझने की कोशिश कर रहा है कि संक्रमण कहां और कब होगा, इसके बारे में ज्ञान और जानकारी प्राप्त करने के लिए इंटरप्रिटेबल एआई का उपयोग कैसे किया जाए। कौन से देश सबसे ज्यादा खतरे में हैं? कमजोर समुदाय कौन हैं? और हम मदद करने की कोशिश करते हैं। इस अध्ययन की यात्रा यह सब है जिसमें बहुत सारे सड़क ब्लॉक हैं जिन्हें हमें दूर करने की आवश्यकता है।

स्लाइड *8*

समस्या यह है कि सटीकता के साथ सामान्य मुद्दा है। हमारा मॉडल कैसे भविष्यवाणी कर सकता है ताकि हमें विश्वास हो कि हमारी भविष्यवाणी सटीक है? इसके अलावा, जब हमारे पास यह भविष्यवाणी है, तो हम नीति निर्माताओं को कैसे समझाएं कि मामलों की वर्तमान वृद्धि के लिए महत्वपूर्ण कारक क्या

हैं? इसके लिए हमें काउंटी स्तर की विशेषताओं की तरह बहुत बेहतर ग्रेड स्तर पर डेटा की गहरी समझ होनी चाहिए। हम वास्तविक समय में भी निर्णय लेना चाहते हैं ताकि इसकी प्रासंगिकता की गारंटी हो।

स्लाइड *9*

यह एक बहुत समृद्ध अध्ययन है क्योंकि स्वास्थ्य असमानता केवल लोगों से संबंधित नहीं है, बल्कि सामाजिक-आर्थिक प्रभाव भी है। हमने 3,142 अमेरिकी काउंटियों के लिए ढाई साल से अधिक डेटा एकत्र किया है। हम डेथ सेट सुविधाओं को बीस से छह विशेषताओं तक वर्गीकृत और कम करते हैं। दो स्थिर विशेषताएं हैं: स्वास्थ्य असमानता और जनसंख्या आयु वर्ग। टीकाकरण, रोग प्रसार, संक्रमणीय मामलों और गतिशीलता सहित अवलोकन योग्य विशेषताएं हैं। हम ज्ञात और अज्ञात घटनाओं में शामिल करते हैं ताकि हमारे पास एक बहुत ही जटिल मल्टीमॉडल आर्टिकुलेटेड एआई मॉडल हो।

स्लाइड *10*

मुझे आपको हमारे कुछ पूर्वानुमान परिणाम दिखाने के लिए गियर स्विच करने दें। हमने टीएफटी मॉडल की तुलना एलएसटीएम से की, जो अनुक्रम मॉडल के लिए एक बुनियादी अनुक्रम है। हम दिखा सकते हैं कि बाईं ओर चार्ट में TFT मॉडल सबसे अच्छा प्रदर्शन करता है। यह त्रुटि संदेश देता है और सटीकता अधिक होती है और त्रुटि कम होती है।

स्लाइड *11*

व्याख्यात्मक एआई की समझ को क्या रेखांकित कर रहा है? यह एक एनकोडर डिकोडर आर्किटेक्चर का उपयोग करके तनाव तंत्र से आता है और तनाव तंत्र चैटजीपीटी सहित वर्तमान बड़े भाषा मॉडल का आधार है। क्योंकि हमारे पास यह बहु-ध्यान तंत्र है, हम समय के साथ बीमारी के संदर्भ को पकड़ सकते हैं और इसलिए यह उस स्थान को परिष्कृत करेगा जहां हम सुविधा को देखते हैं - कारण और प्रभाव। इसलिए हम हॉटस्पॉट क्षेत्रों के साथ स्थानिक और लौकिक पैटर्न के महत्व पर ध्यान केंद्रित कर सकते हैं। दाईं ओर, आप TFT मॉडल का आर्किटेक्चर देख सकते हैं। हम पिछली विशेषताओं को इनपुट करते हैं और भविष्य में होने वाली भविष्यवाणी करने का प्रयास करते हैं। इस मामले में, यह संक्रमण के मामले हैं जो भरोसेमंद समय पैटर्न को पकड़ने के लिए अनुक्रम मॉडल का उपयोग करने में सुविधाओं, विशेष रूप से स्थिर विशेषताओं को एम्बेड करते हैं। हम इन सभी पैटर्नों को आत्म-ध्यान देने के लिए प्रचारित करते हैं ताकि व्याख्यात्मक बहु-सिर ध्यान को मुखौटा करने की कोशिश की जा सके ताकि हम महत्वपूर्ण पैटर्न और क्षेत्रों पर ध्यान केंद्रित कर सकें।

स्लाइड *12*

यह पूर्वानुमान मॉडल हमें चक्रीय पैटर्न प्रदान करने में सक्षम है जो COVID मामलों की रिपोर्टिंग है। यह विशेष समय की घटनाओं के लिए भी जिम्मेदार है, जैसे छुट्टियां, जैसे सप्ताहांत। हम उन्हें इस चार्ट में स्पष्ट रूप से लेबल कर सकते हैं। दाईं ओर हम पीछे की ओर भी देख सकते हैं कि 0 से 13 दिनों की समय सीमा में भविष्य की भविष्यवाणी का किस अवधि पर सबसे अधिक प्रभाव पड़ता है।

स्लाइड *13*

यहां रुझानों का चार्ट दिया गया है जिसकी हम भविष्यवाणी कर सकते हैं। शीर्ष 100 सबसे बड़ी काउंटियों को चुनना, आप देख सकते हैं कि हमारे पास जमीनी सच्चाई की तुलना में भविष्यवाणी है। इसके अलावा, हम छोटी आबादी काउंटियों की तुलना कर सकते हैं - आप देख सकते हैं कि परिणाम बहुत बेहतर मेल खाता है और इसमें कम स्पाइक्स और विभिन्न मूल्य हैं।

स्लाइड *14*

स्थान की जानकारी के बारे में कैसे? यह चार्ट बाईं ओर एआई मॉडल के आत्म-ध्यान को दिखाता है, और काउंटी स्तर पर तीव्रता को कैप्चर करता है। यदि हमारा डेटा राज्य स्तर पर है तो यह संभव नहीं है। दाईं ओर 3,000 से अधिक अमेरिकी काउंटियों के सीडीसी से संचयी मामलों का डेटा प्रतिनिधित्व है। यदि आप इन दो परिणामों को देखते हैं, तो आप इन दो डेटासेट और परिणामों के बीच संबंध देख सकते हैं।

स्लाइड *15*

हमने सहसंबंध को मापा और निष्कर्ष निकाला कि हम काउंटी स्तर पर आत्म-ध्यान भार को कैप्चर करके एआई मॉडल की व्याख्या कर सकते हैं। जमीनी सच्चाई की तुलना में मामलों की भविष्यवाणी बनाम मॉडल व्यवहार के साथ एक मजबूत संबंध है।

स्लाइड *16*

R2 वह जानकारी प्रदान करता है जिसकी नीति निर्माताओं को आवश्यकता होती है। हमारा मानना है कि हॉटस्पॉट ट्रांसमिशन में एक छोटी सी कमी से संक्रमण में बड़ी कमी आ सकती है, खासकर शुरुआती चरण में। वास्तविक समय का पूर्वानुमान लगाना और दैनिक संक्रमण में सबसे महत्वपूर्ण क्षेत्रों पर हमारा ध्यान केंद्रित करना महत्वपूर्ण है। हमारे पास काउंटी स्तर पर इन संक्रमणों को पकड़ने के लिए एक बढ़िया तरीका है जो जोखिम को काफी कम कर देगा। भविष्य के कई काम हैं जो हम इस मौजूदा परिणाम के साथ मिलकर कर सकते हैं। हम भविष्य के काम के लिए कई सामाजिक और आर्थिक प्रभावों और असमानताओं का पता लगा सकते हैं।

स्लाइड *17*

यूवीए में, हमने विज्ञान कार्यक्रम के लिए एआई शुरू किया और भाग लेने के लिए 3,000 से अधिक स्नातक को आकर्षित किया। हमने अपनी परियोजना में शामिल एक दर्जन छात्रों का चयन किया।

स्लाइड *18*

हम मॉरिस आधारित सूचकांक अध्ययन का उपयोग करके जनसंख्या आयु वर्ग संवेदनशीलता का और अध्ययन करना चाहते हैं ताकि प्रत्येक जनसंख्या समूह से यह पता चल सके कि कोविड संक्रमणों के लिए सबसे अधिक संवेदनशील कौन हैं?

स्लाइड *19*

हमने एक समय-श्रृंखला गहन शिक्षण मॉडल चुना क्योंकि उपरोक्त तनाव तंत्र अंतर्दृष्टि दे सकता है और समझ सकता है कि मॉडल की भविष्यवाणी कैसे की जाती है।

स्लाइड *20*

इस यात्रा से हमने कई सबक सीखे हैं। संक्रामक रोग एक सामाजिक-आर्थिक विश्वव्यापी समस्या रही है जो सार्वजनिक स्वास्थ्य और अर्थव्यवस्था को इतने बड़े पैमाने पर प्रभावित करती है। COVID-19 महामारी से, हमारा पहला सबक सीखा है कि महामारी की प्रगति को समझने के लिए परीक्षण महत्वपूर्ण है। दूसरा यह है कि विभिन्न छोटे क्षेत्रों में बहुत सारे संक्रमण बहुत अलग हैं। स्थिति काफी गतिशील है। नीति में इसे संबोधित करने का बेहतर तरीका स्थानीय स्तर पर अनुकूलित है। ऐसे कई तरीके हैं जिनसे हम सुधार कर सकते हैं, जिसमें हमने COVID संक्रमण और भविष्य की संक्रामक बीमारी की सटीक भविष्यवाणी करने के लिए अध्ययन किए गए उपकरण बनाए हैं। इससे नीति निर्माता को विज्ञान आधारित तरीके से हस्तक्षेप करने में मदद मिलेगी। हस्तक्षेप भविष्य है। हम महामारी जैसे भविष्य के संकटों के लिए तैयार और तैयार रहना चाहते हैं।

स्लाइड *21*

इससे पहले कि मैं इसे लपेटूं, मैं वास्तव में हर किसी के साथ साझा करना चाहता हूं कि हमें विश्वास बनाने की आवश्यकता है। यदि पिछली महामारी से कोई सबक सीखा गया है, तो हम जनता, विशेषज्ञों, नीति निर्माताओं को स्पष्टीकरण में सुधार करना चाहते हैं कि क्या हो रहा है और हम नीति को कैसे प्रभावित कर सकते हैं। हम मात्रात्मक रूप से मॉडल बनाने और हमारे तरीकों का मूल्यांकन करने के लिए एक व्याख्यात्मक विधि चाहते हैं। हम अपने मॉडल के व्यवहार और एआई-आधारित भविष्यवाणियों को सभी जनता और छात्रों सहित गैर-विशेषज्ञों को भी समझाना चाहते हैं। सामूहिक रूप से, हम एक भविष्य बनाने की उम्मीद करते हैं ताकि हम भविष्य की घटनाओं के लिए तैयार हों।

स्लाइड *22*

इसके साथ, मैं आज की कार्यशाला में भाग लेने के लिए आप सभी को धन्यवाद देना चाहता हूं। मैं यहां हमारे कुछ काम प्रदान करता हूं। कृपया मुझसे संपर्क करने में संकोच न करें। मैं इस कार्यशाला के अंत में भविष्य की चर्चाओं की प्रतीक्षा कर रहा हूं। धन्यवाद!