

COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Hong Qin (University of Tennessee, Chattanooga), July 26, 2023



Title: Develop and Evaluate Computational Frameworks to Predict and Prevent Future Coronavirus Pandemics

Hong Qin CIC Database Profile

NSF Award #: 2200138

YouTube Recording with Slides

Summer 2023 CIC Webinar Information

Transcript Editor: Shikhar Johri

Transcript

स्लाइड 1

नमस्ते, मेरा नाम हांग किन है और मैं टेनेसी विश्वविद्यालय, चट्टानूगा में संकाय हूं। यह बात एक ऐसी परियोजना के बारे में है जिस पर मुझे नॉर्थ कैरोलिना ए एंड टी, स्पेलमैन कॉलेज, कैथोलिक यूनिवर्सिटी ऑफ अमेरिका के सहयोगियों की एक बड़ी टीम के साथ काम करने का सौभाग्य मिला। ये कागज पर सहयोगी हैं, लेकिन वास्तव में, कई और भी हैं। हमारे सहयोग में हमारे पास शायद एक दर्जन से अधिक टीम के सदस्य थे जो इस पर काम कर रहे थे।

तो यह एक राष्ट्रीय विज्ञान फाउंडेशन समर्थित परियोजना है। इसे PIPP चरण एक कहा जाता है। मेरा लक्ष्य भविष्य में कोरोनावायरस महामारी की भविष्यवाणी करने और रोकने के लिए एआई आधारित ढांचा विकसित करना है। हालांकि हम कहते हैं कि यह कोरोनावायरस है, मॉडल में अब संभवतः कई अन्य वायरल महामारियों के लिए सामान्य होने की संभावना है।

स्लाइड 2

तो यह है - ठीक है, यह स्वचालित रूप से अग्रेषित हो रहा है - मेरी क्षमायाचना। ठीक है, तो यह शुरुआत में प्रस्तावित हमारे द्वारा प्रस्तावित अवलोकन है, हालांकि हमने अभी बहुत सारे संशोधन किए हैं। विचार अभी भी वही है। हम प्रस्ताव करते हैं कि हम वायरस की भविष्यवाणी कैसे करते हैं - एक नया वायरस महामारी? हमारा विचार है कि पहले हम सभी संभावित SARS-CoV-X उत्पन्न करते हैं जो पुनर्संयोजन पैटर्न या निवास स्थान परिवर्तन के आधार पर अनुमानित हैं। फिर हम उन संभावित SARS-CoV-X प्रतिक्रियाओं की भविष्यवाणी करते हैं। यहां चुनौती है - उन अनुक्रमों को उत्पन्न करना आसान है, लेकिन हम कैसे जानते हैं कि कौन सा वास्तव में मानव आबादी में वायरल या संचारित हो जाएगा? यह वास्तव में यहाँ कुंजी है। इसलिए हम एआई की ओर रुख करते हैं। जब मैं कहता हूं कि एआई एक ब्लैक बॉक्स है, तो इस मामले में यह शायद एक आशीर्वाद भी है क्योंकि हम वास्तव में नहीं जानते कि एक वायरस मेजबान से, एक जानवर से, एक इंसान से कैसे कूदता है। तो एआई एक ब्लैक बॉक्स है लेकिन यह शायद

एक महान उपकरण है जिसका हम उपयोग कर सकते हैं, भले ही यह एक ब्लैक बॉक्स हो। यही मेरा तर्क है।

और मेरी माफि, मुझे कुछ चीजों को ब्लैक आउट करना पड़ा क्योंकि मेरा विश्वविद्यालय इसके कार्य के आधार पर पेटेंट के लिए आवेदन कर रहा है। तो लक्ष्य यहाँ है हम इनपुट दृश्यों हैं - यह सिर्फ प्राथमिक न्यूक्लियोटाइड अनुक्रम है। इस क्रम से, हम कुछ उपयोगी सुविधाएँ उत्पन्न करने के लिए फीचर इंजीनियरिंग करेंगे। फिर हम अपने एआई मॉडल में फीड करने के लिए अनुक्रम सुविधा का उपयोग करते हैं। हमारे एआई मॉडल के आधार पर हम फिनिश की भविष्यवाणी करेंगे। एआई की हालिया प्रगति को देखते हुए, हम कुछ उम्मीदवार जीनों को विषाणु परिवर्तन के संभावित ड्राइवरों के रूप में सत्यापित करने के लिए [अस्पष्ट] गतिशील सिमुलेशन या यहाँ तक कि प्रयोगात्मक तरीकों का भी उपयोग कर सकते हैं। एआई मॉडल को देखते हुए, हम इसे भी सामान्यीकृत करते हैं - मूल रूप से कोरोनविर्यूज से स्थानांतरण सीखने का उपयोग करते हैं, संभवतः इन्फ्लूएंजा या एचआईवी, एचपीवी जैसे अन्य वायरस के लिए। कई वायरस हैं। मैं वास्तव में जीव विज्ञान में विशेषज्ञ नहीं हूं, मैं सिर्फ कंप्यूटर विज्ञान हूं।

और इसलिए हम वास्तव में जिन चुनौतियों से निपट रहे हैं - मेरी क्षमा याचना, यह स्वचालित रूप से फिर से अग्रेषित हो रही है। चुनौती वास्तव में अनुक्रमों से रोगजनकता की भविष्यवाणी करना है। हम एक संभावित नियम भी पा सकते हैं और भविष्य के वायरस की भविष्यवाणी करके सामान्य क्षमता का परीक्षण कर सकते हैं। यह एआई सक्षम होगा, प्रारंभिक चेतावनी प्रणाली। हम वैक्सीन विकास की भविष्यवाणी करने के लिए एआई का उपयोग करना चाहते हैं। यह देखते हुए कि हम जानते हैं कि विषाणु विचरण कैसे बदलता है। तो यह इस की बड़ी तस्वीर है।

स्लाइड 3

तो यहाँ कुछ परिणाम दिए गए हैं जिन्हें हमने विकसित किया है - जिन्हें हमने पूरा किया है। यह जानकारी के आधार पर वायरल फिटनेस का अनुमान है - हमारे पास जो डेटा है। यह वास्तव में ओमिक्रॉन सब-वेरिएंट के लिए है और हम विकास जोड़ीवार तुलना के साथ एक विधि का उपयोग करके ओमिक्रॉन संस्करण की तुलना कर सकते हैं। फिर हम उन वेरिएंट के बीच सापेक्ष अंतर का अनुमान लगा सकते हैं।

स्लाइड 4

इस डेटा के आधार पर हम एक तथाकथित वायरल फिटनेस परिवृश्य उत्पन्न कर सकते हैं। विकास के सिद्धांत से अवगत लोगों के लिए - विकास में एक तथाकथित फिटनेस परिवृश्य है और इसके बारे में वास्तव में बहुत सारे तर्क हैं, लेकिन विकास परिवृश्य के आधार पर बहुत सारे सिद्धांत भी हैं। एक संभावित वायरस के फिटनेस परिवृश्य के साथ, हम सिद्धांत रूप में इसके विकासवादी प्रक्षेपवक्र की भविष्यवाणी कर सकते हैं। यह - हम कर सकते हैं - हमारे पास ऐसा करने की एक विधि है। यह संयुक्त राज्य अमेरिका में अल्फा बीटा डेल्टा ओमिक्रॉन सबवेरिएंट के लिए अनुमानित परिणाम है। हम इसे इन्फ्लूएंजा या चेचक पर भी लागू करने की प्रक्रिया में हैं। हमारे पास चेचक या अन्य वायरस के लिए पर्याप्त डेटा नहीं है, लेकिन इन्फ्लूएंजा आशाजनक प्रतीत होता है। तो यह इस की एक महत्वपूर्ण खोज है - हमारी वर्तमान परियोजना।

स्लाइड 5

यहाँ कुछ मॉडल परिणाम दिए गए हैं। एआई मॉडल में बहुत अधिक विवरण दिए बिना हमने विकसित किया है। वास्तव में बहुत सारे पैरामीटर हैं, लेकिन वर्तमान मॉडल की सटीकता काफी उच्च सटीकता प्राप्त कर सकती है। बेशक, यह हम जो जानते हैं उस पर आधारित है। चुनौती वास्तव में अज्ञात के बारे में

है। भले ही हम अज्ञात की भविष्यवाणी कर रहे हैं, डेटा पहले से ही जात है। मशीन लर्निंग में हम इसे इनपुट प्रशिक्षण और परीक्षण के साथ प्रशिक्षित करते हैं। अज्ञात की भविष्यवाणी करने के बारे में चुनौती अभी भी एक खुला प्रश्न बनी हुई है।

स्लाइड 6

कुछ लोग तर्क दे सकते हैं, अच्छी तरह से हमें गहरी शिक्षा क्यों करनी है? बहुत सारे सांख्यिकीय, आनुवंशिकी, जीनोमिक्स, मात्रात्मक उपचार हैं - बहुत सारे अन्य तरीके हैं और हमें गहरी शिक्षा में क्यों बदलना है? यहां, हमारे पास कुछ सबूत हैं जो गहरी शिक्षा दिखाते हैं, कम से कम, यह मानक जीनोम-वाइड एसोसिएशन अध्ययनों से अलग-अलग हस्ताक्षरों का पता लगाता है। बाईं ओर, हमारे गहरे सीखने के परिणाम का परिणाम है। यह वास्तव में WHO लेबल किए गए विचरण अल्फा, बीटा, डेल्टा, ओमिक्रॉन और अन्य पर आधारित है। हम यह देखना चाहते हैं कि विषाणु की इस वृद्धि में योगदान देने के लिए SARS-CoV-2 जीनोम में कौन से हस्ताक्षर महत्वपूर्ण हैं। वे उच्च संकेत जिन्हें आप SARS-CoV-2 जीनोम के अंत में देख सकते हैं। और इस क्षेत्र के आसपास में लगभग 20,000 से 25,000 पर प्रकाश डाल रहा हूं, यह स्पाइक जीन है जिसे डब्ल्यूएचओ [ने पहचाना है] और अधिकांश प्रतिरक्षा प्रतिक्रिया पर प्रतिक्रिया दी जाएगी। पारंपरिक ज्ञान के आधार पर, वास्तव में यह होना चाहिए कि डब्ल्यूएचओ ने SARS-CoV-2 वेरिएंट को कैसे वर्गीकृत किया। आश्चर्यजनक रूप से एआई मॉडल उन संकेतों को उठाते हैं, लेकिन वे सबसे मजबूत संकेत नहीं हैं, और वे कई अन्य संकेतों को उठाते हैं। कछ मजबूत संकेत नहीं हैं। जीनोम वाइड एसोसिएशन स्टडी जैसे पारंपरिक सांख्यिकीय जीनोमिक तरीकों का उपयोग करके - जो अभी भी स्पाइक जीन को मजबूत संकेत के रूप में चुनता है। कछ अन्य हैं, लेकिन बहुत मजबूत नहीं हैं। तो इस मामले में, एआई और पारंपरिक सांख्यिकीय विधि पिंक, या कम से कम, उन संकेतों को एक अलग वजन देते हैं।

यह एक तरह से आश्चर्यजनक और आश्वस्त करने वाला भी है। इसलिए हमने पारंपरिक ज्ञान, स्पाइकिंग को चुना, लेकिन हमने कुछ अन्य संकेतों को भी चुना जो प्रयोगात्मक विधि द्वारा सत्यापित किए जा सकते हैं या नहीं। लेकिन हम कैसे भविष्यवाणी करते हैं और हम इसे कैसे सत्यापित करते हैं, है ना? इसलिए हम कई चीजों की भविष्यवाणी करने के लिए एआई का उपयोग करते हैं और हम जानते हैं कि एआई बहुत सारी झूठी भविष्यवाणियां उत्पन्न कर सकता है। इस मामले में हम इसे कैसे सत्यापित कर सकते हैं? यह काफी चुनौतीपूर्ण है।

स्लाइड 7

हम सामान्य तौर पर दो अलग-अलग तरीकों से कोशिश कर रहे हैं। एक निष्कर्षों को सत्यापित करने के लिए एक मॉडल प्रयोगात्मक प्रणाली उत्पन्न करना है। हमारी टीम एक जैविक मॉडल का उपयोग कर रही है, यह [अश्रव्य] है, हमारे पास उन जीनों को जीवन चक्र में बदलने और उनकी सापेक्ष गतिविधियों को मापने के लिए एक सेल लाइन सिस्टम भी है।

स्लाइड 8

फिर हमारे पास आणविक गतिशीलता करने के लिए एक कम्प्यूटेशनल व्यक्ति भी है ताकि यह अनुकरण किया जा सके कि वे उत्परिवर्तन उनकी गतिविधि के [अश्रव्य] को कैसे प्रभावित करते हैं। इस मामले में, ACE2 और RBD, लेकिन हमारे पास अन्य [अश्रव्य] भी हैं जिनमें विभिन्न क्षेत्रों में भविष्यवाणी की जाती है कि वे मानव जीन के साथ कैसे प्रतिक्रिया करते हैं - मानव प्रतिरक्षा प्रणाली को लक्षित करने की क्षमता।

स्लाइड 9

हम एआई में सामाजिक विश्वास को बढ़ावा देने के लिए भरोसेमंद एआई को डिजाइन करने के तरीके पर चर्चा करने के लिए एक कार्यशाला और विशेषज्ञ पैनल भी आयोजित करते हैं - समान एआई। COVID महामारी ने दिखाया है कि हमारी वर्तमान प्रणाली में उच्च असमानता है, इसलिए यदि हम भविष्य की महामारियों की भविष्यवाणी करने के लिए AI का उपयोग करते हैं, तो यह हमारी वर्तमान प्रणाली में छिपी असमानताओं को आसानी से बढ़ा सकता है, शायद हमारे वर्तमान डेटा में भी। यदि हम इसे पहचानते हैं और हम एक काफी विविध पैनल का आयोजन करते हैं, और दक्षिण अफ्रीका, केन्या के लोगों सहित, यूरोप और भर से एक व्यक्ति है - वकीलों, चिकित्सकों, नीति निर्माताओं, सरकारी - एनआईएसटी - सरकारी एजेंसियों जैसे विभिन्न पृष्ठभूमि के साथ एआई को बढ़ावा देने और ऐतिहासिक रूप से हाशिए वाले समुदायों में एआई में विश्वास को बढ़ावा देने के लिए सभी प्रकार के परिप्रेक्ष्य हैं।

स्लाइड 10

हम अल्पसंख्यक सेवारत संस्थानों के साथ हमारे सहयोग को देखते हैं बहुत सारे कार्यबल विकास पर भी जोर देते हैं, जिसमें स्पेलमैन कॉलेज, उत्तरी कैरोलिना ए एंड टी, और कैथोलिक यनिवर्सिटी ऑफ अमेरिका शामिल हैं, जिसमें बहुत सारे हिस्पैनिक छात्र शामिल हैं, और मेरा गृह संस्थान टेनेसी विश्वविद्यालय, चट्टानूगा। हम एक अस्पताल प्रणाली ग्लोबल साउथ के साथ भी सहयोग करते हैं।

स्लाइड 11

अंत में मैं इसके समर्थन के लिए एनएसएफ, यूटीसी और एआई टेनेसी पहल को धन्यवाद देना चाहता हूं। और अंत में, मैंने स्लाइड्स में नहीं डाला, एनएसएफ ने एक राष्ट्रीय केंद्र के लिए चरण 2 आवेदन के लिए कॉल जारी किया है, जो शायद 10 गुना बड़ा है, अगर 100 गुना बड़ा नहीं है, तो मैं सहयोगियों की तलाश कर रहा हूं, विशेष रूप से सामुदायिक जुड़ाव, सार्वजनिक नीति के साथ, मैंने किसी को नीति पर काम करते हुए सुना। इसलिए मैं इसके बाद जुड़ना चाहता हूं। तो मैं बस यहाँ बंद कर दूँगा, धन्यवाद।