COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by Samson Qian (University of California, San Diego), August 18, 2021

Title: *Generating Explanations for Chest Medical Scan Pneumonia Predictions*

YouTube Recording with Slides

August 2021 CIC Webinar Information

Transcript Editor: Macy Moujabber

Transcript

Lauren Close:

मैं आज हमारे अंतिम वक्ता, सैमसन कियान का परिचय देना चाहता हूं। सैमसन उद्घाटन सीआईसी स्नातक छात्र पेपर चुनौती के हमारे तीन विजेताओं में से एक है जो इस वसंत से पहले आयोजित किया गया था। इसलिए हम सैमसन का स्वागत करते हैं और हम एक उभरते विद्वान के रूप में अपने शोध को साझा करने के लिए बहुत उत्साहित हैं। इसलिए शिमशोन, कृपया आगे बढ़ें और इसे दूर ले जाएं।

Samson Qian:

स्लाइड *1*

धन्यवाद लॉरेन। इसलिए आज मैं एक प्रस्तुति कर रहा हूं जो वायरल और बैक्टीरियल निमोनिया पर मशीन सीखने की भविष्यवाणियों के लिए स्पष्टीकरण उत्पन्न करने के बारे में पिछले वाले से थोड़ा अलग है।

स्लाइड *2*

और इसलिए इस शोध का व्यापक लक्ष्य विभिन्न छाती एक्स-रे रोगियों को लेना है जिनके पास बैक्टीरिया और वायरल निमोनिया के साथ-साथ स्वस्थ रोगी भी हैं और इन छवियों का उपयोग करके इन रोगियों के बीच वर्गीकृत करने के लिए मशीन लर्निंग मॉडल का निर्माण करना है। और इसलिए क्या मशीन लर्निंग मॉडल बनाना संभव है जो इन विभिन्न वर्गों की सटीक पहचान कर सकता है, लेकिन इस मशीन लर्निंग मॉडल की व्याख्या भी कर सकता है ताकि यह समझ सके कि मॉडल भविष्यवाणियां उत्पन्न करते समय कहां देख रहा है? और इसलिए यह व्याख्या एल्गोरिदम का उपयोग करने के विचार का परिचय देता है, जिसे एक्सप्लेनेबल एआई के रूप में भी जाना जाता है, एक मॉडल का विश्लेषण करने के लिए और जिस डेटा पर यह भविष्यवाणी कर रहा है ताकि यह समझ सके कि मॉडल कहां देख रहा है। और अंतिम लक्ष्य यह देखना है कि क्या इस प्रकार का ढांचा रेडियोलॉजिस्ट को विभिन्न रोगियों के निदान में उनके काम में मदद कर सकता है।

स्लाइड *3*

तो व्याख्यात्मक एआई का एक त्वरित अवलोकन। विभिन्न प्रकार के व्याख्यात्मक एआई एल्गोरिदम हैं, जैसे, विभिन्न प्रकार के तरीकों की पारिवारिक शाखा की तरह। इस शोध के लिए, विशेष रूप से, हम पोस्ट-हॉक एल्गोरिदम

पर ध्यान केंद्रित करेंगे, जिसका अर्थ है कि प्रक्रिया के दौरान या उससे पहले डेटा पर प्रशिक्षित होने के बाद जटिल मॉडल का विश्लेषण करने के लिए एल्गोरिदम। और इसका लाभ यह है कि आप एक्स-रे में सुविधाओं को सीखने के लिए अधिक जटिल मॉडल का उपयोग करने में सक्षम हैं, जैसे कि गहरे दृढ़ तंत्रिका नेटवर्क। और कभी-कभी व्याख्या और सटीकता के बीच एक व्यापार-बंद होता है। और इसलिए गहरे और अधिक जटिल मॉडल सरल लोगों की तुलना में व्याख्या करना कठिन है, इसलिए ये एल्गोरिदम इन अधिक जटिल मॉडलों को समझने का एक तरीका प्रदान करते हैं। तो पहले प्रकार के एल्गोरिथ्म को एलआरपी, परत-वार प्रासंगिकता प्रसार के रूप में जाना जाता है, और यह एक ऐसी विधि है जिसके बारे में मैं थोड़ी देर बाद बात करूंगा, लेकिन यह अनिवार्य रूप से मॉडल की परतों और वजन को देखता है और समझता है कि छवि में पिक्सेल मॉडल की आंतरिक संरचना को सक्रिय करने में सबसे अधिक योगदान करते हैं।

दूसरा एक लाइम के रूप में जाना जाता है और यह विधि मॉडल-अज्ञेयवादी है, जिसका अर्थ है कि यह इस बात पर निर्भर नहीं करता है कि आपने किस प्रकार के मॉडल का उपयोग किया है। यह विधि एलआरपी से इस मायने में अलग है कि यह मॉडल के बजाय पहले डेटा से शुरू होती है और यह डेटा के अलग-अलग उपखंडों को देखती है ताकि यह पता लगाया जा सके कि छवि में कौन से क्षेत्र मॉडल की भविष्यवाणी में सबसे अधिक योगदान करते हैं।

और फिर अगले एक को ग्रैड-सीएएम कहा जाता है जो एलआरपी प्रकार के समान है, लेकिन प्रत्येक परतों और प्रत्येक वजन को देखने के बजाय, यह मॉडल में कनवल्शन परतों पर एक नज़र डालता है और फिर यह ग्रेडिएंट पर ले जाता है कि जब आप एक छवि फिट करते हैं, तो अंतिम दृढ़ परत के ग्रेडिएंट, और फिर यह ग्रेडिएंट के सक्रियण मानचित्र की तरह, जैसे, उत्पादन करता है।

और अंतिम प्रकार का एल्गोरिथ्म एक उपन्यास एल्गोरिथ्म की तरह है जिसे कंट्रास्टिव एलआरपी के रूप में जाना जाता है जो नियमित एलआरपी का एक संशोधन है, लेकिन यह प्रासंगिकता लेता है और फिर यह वायरल बैक्टीरियल निमोनिया जैसे विभिन्न वर्गों के बीच भिन्न होता है, इसलिए आप कक्षाओं के बीच अंतर की कल्पना कर सकते हैं।

स्लाइड *4*

इसलिए इस प्रकार के रोगियों में सुविधाओं की पहचान करने के लिए एक दृढ़ तंत्रिका नेटवर्क के निर्माण के लिए VGG16 और ResNet50 जैसी एक जटिल मॉडल संरचना की आवश्यकता होती है जो दो अत्याधुनिक मॉडल हैं जो छवि वर्गीकरण में बहुत अच्छा काम करते हैं, और जैसा कि आप मॉडल संरचनाओं से देख सकते हैं, यह एक बहुत गहरा दृढ़ नेटवर्क है जो छवियों में कई विशेषताओं को लेने में सक्षम है। और इसलिए ये दो मॉडल रोगियों के तीन अलग-अलग वर्गों के बीच वर्गीकृत करने में कुछ बेहतरीन परिणाम देते हैं।

स्लाइड *5*

और इसलिए एक्स-रे छवियों को सभी मेंडली डेटा से एकत्र किया गया था और छवियों पर छवि पूर्व-प्रसंस्करण का एक गुच्छा मॉडल में फिट करने के लिए आवश्यक है। इसलिए यह सुनिश्चित करना कि सभी विभिन्न प्रकार के वर्गों से समान मात्रा में छवि उदाहरण हैं और फिर मॉडल में फिट होने के लिए छवियों पर कुछ पूर्व-प्रसंस्करण कर रहे हैं।

स्लाइड *6*

और यहाँ प्रशिक्षण के दौरान मॉडल के प्रदर्शन के अवलोकन की एक छवि की तरह है। तो बाईं ओर आपके पास छवियों के सत्यापन सेट में प्रशिक्षण सेट के लिए सटीकता प्लॉट है। और जैसे-जैसे आप मॉडल को प्रत्येक पुनरावृत्ति, प्रत्येक युग में अधिक से अधिक प्रशिक्षित करते हैं, सटीकता अधिक बढ़ती जाती है और फिर नुकसान जो हम दाईं ओर कम करने की कोशिश कर रहे हैं, लगातार कम हो रहा है। तो इसका मतलब है- इससे पता चलता है कि हमारा मॉडल डेटा में सुविधाओं को उठा रहा है- जिन छवियों को हम इसे खिला रहे हैं और यह छाती एक्स-रे के बीच वर्गीकृत करने में अच्छा काम कर रहा है।

स्लाइड *7*

तो यहां निर्मित मॉडल पर अधिक विश्लेषण है, और प्रत्येक छवि में मॉडल क्या देख रहा है, इसकी पहचान या जांच करने के लिए एक सटीक मॉडल होना बहुत महत्वपूर्ण है। तो आप बाईं ओर एक भ्रम मैट्रिक्स देख सकते हैं जो दिखाता है कि मॉडल क्या भ्रमित कर रहा है और सामान्य तौर पर वायरल और बैक्टीरियल निमोनिया के बीच अंतर करने के अलावा तीन वर्गों के बीच बहुत अच्छा काम कर रहा है जो कर रहा है- यह कुछ उदाहरणों पर भ्रमित हो रहा है। और इसलिए यह समझने के लिए इन व्याख्यात्मक एल्गोरिदम को चलाना महत्वपूर्ण है कि मॉडल क्या देख रहा है।

स्लाइड *8*

तो यहां एलआरपी का एक संक्षिप्त अवलोकन है, परत-वार प्रासंगिकता प्रसार, फिर से, और अनिवार्य रूप से यह क्या करता है कि आपके पास प्रत्येक परत के साथ आपका तंत्रिका नेटवर्क है- अंदर एक छिपी हुई परत। और यह आपकी छवि को आउटपुट परत से पीछे की ओर प्रचारित करता है। तो आगे प्रचार के बजाय, यह छवि को पीछे की ओर प्रचारित करता है और फिर आप प्रत्येक न्यूरॉन पर प्रासंगिक स्कोर की गणना करते हैं, जो कि ये सर्कल यहीं हैं। और ये प्रासंगिक स्कोर दर्शाते हैं कि छवि में प्रत्येक पिक्सेल मॉडल द्वारा की जाने वाली भविष्यवाणी के लिए कितना महत्वपूर्ण है।

स्लाइड *9*

और इसलिए यहां आप एक नज़र डाल सकते हैं- दाईं ओर एलआरपी का एक उदाहरण है। और जैसा कि आप देख सकते हैं, यह छवि में अलग-अलग पिक्सेल के आधार पर हीट मैप की तरह उत्पादन करता है और यह दर्शाता है कि छवि में प्रत्येक पिक्सेल कक्षा पर मॉडल की भविष्यवाणी में कितना महत्वपूर्ण योगदान दे रहा है। और फिर बाईं ओर आप LIME की तुलना देख सकते हैं जो एक क्षेत्र-आधारित विधि से अधिक है जो छवि में क्षेत्रों को लेता है, और फिर पाता है कि मॉडल भविष्यवाणी करने के लिए कौन से क्षेत्र सबसे महत्वपूर्ण हैं।

स्लाइड *10*

और फिर यहां तीन अलग-अलग वर्गों के लिए LIME के अधिक उदाहरण दिए गए हैं। LIME क्या करता है- यह विशेष रूप से वजन की जांच करने के लिए मॉडल की प्रत्येक परत में नहीं जाता है, बल्कि इसके बजाय यह विशेष रूप से छवि पर चुने गए उप-क्षेत्रों को लेता है और यह निर्धारित करने के लिए मॉडल पर चलता है कि कौन से उप-क्षेत्र मॉडल की भविष्यवाणी के लिए सबसे महत्वपूर्ण हैं और उन विशिष्ट क्षेत्रों पर प्रकाश डाला गया है।

स्लाइड *11*

और फिर एक अन्य प्रकार का एल्गोरिथ्म जिसकी मैंने चर्चा की थी, उसे ग्रेड-सीएएम कहा जाता था और ग्रैड-सीएएम अनिवार्य रूप से एलआरपी की तरह एक एल्गोरिथ्म प्रकार है, लेकिन इसके बजाय यह आपके मॉडल में दृढ़ परतों के ग्रेडिएंट की जांच करता है। और इसलिए यह विधि एक छवि भी लेती है और फिर इसे मॉडल में फीड करती है और फिर एक सक्रियण मानचित्र तैयार करती है जो एक हीट मैप है जो एक दृढ़ परत में सभी ग्रेडिएंट का प्रतिनिधित्व करता है और यह उठाता है कि वास्तव में मॉडल कहां देख रहा है।

स्लाइड *12*

तो यहाँ इसके बजाय ग्रैड-सीएएम के कुछ उदाहरण दिए गए हैं। यह एलआरपी और एलआईएम के बीच संयोजन की तरह है। यह छवि पर एक विशिष्ट क्षेत्र का हीट मैप तैयार करता है, और आप देख सकते हैं कि मॉडल किस पर ध्यान केंद्रित कर रहा है और यह किन क्षेत्रों पर ध्यान केंद्रित नहीं कर रहा है।

स्लाइड *13*

और अंतिम भिन्नता जो सबसे अधिक विपरीत एलआरपी है, मूल एलआरपी विधि में थोड़ा संशोधन है, लेकिन इसके बजाय, यह प्रासंगिकता स्कोर लेता है और एक संशोधन लागू करता है जो विभिन्न प्रकार के वर्गों के बीच प्रासंगिकता

को अलग करता है। और इसलिए आप इन विभिन्न प्रकार के प्रासंगिक स्कोर का उपयोग करके वायरल और बैक्टीरियल निमोनिया के बीच बहुत अधिक स्पष्ट रूप से पहचान या अंतर कर सकते हैं। और इसलिए यह समझना बहुत महत्वपूर्ण है कि आपका मॉडल विशेष रूप से क्या देख रहा है, यह समझने के लिए इन विभिन्न प्रकार की व्याख्यात्मक विधियों को देखना और तुलना करना बहुत महत्वपूर्ण है। और इसलिए कई प्रकार के एल्गोरिदम हैं जिनका आप उपयोग कर सकते हैं, लेकिन आपके मॉडल ने क्या सीखा है, इसकी पूरी तरह से समझने के लिए- जब आपका मॉडल भविष्यवाणियां कर रहा है तो वह क्या देख रहा है- इस प्रकार की तुलना करना महत्वपूर्ण है विधियों और जब आप रोगियों की मदद करने के लिए एक सटीक मॉडल का निर्माण कर रहे हों।

स्लाइड *14*

और अभी कुछ चल रहे शोध- व्याख्यात्मक एआई एक निरंतर बढ़ता हुआ क्षेत्र है। हर दिन अधिक एल्गोरिदम विकसित किए जा रहे हैं और अधिक सटीक मॉडल बनाने के लिए अधिक डेटा एकत्र किया जा रहा है और सीटी स्कैन जैसे विभिन्न प्रकार के डेटा भी रेडियोलॉजिस्ट को रोगियों के निदान में काम करने में मदद कर सकते हैं। इसलिए इन एल्गोरिदम को विभिन्न प्रकार के मॉडलों पर चलाने के लिए यह जांचने के लिए बनाया गया है कि किस मॉडल ने डेटा को दूसरों की तुलना में अधिक सटीक रूप से सीखा है।

स्लाइड *15*

माइकल पज़ानी को स्वीकार करना चाहता हूं, जो अभी अपने शोध करने, डेटा एकत्र करने और निमोनिया की पहचान करने के लिए इन एल्गोरिदम को लागू करने में उनकी सभी मदद के लिए कॉल पर हैं। और धन्यवाद। अगर आप मुझसे संपर्क करना चाहते हैं तो यह मेरा ईमेल है।

स्लाइड *16*

बहुत बहुत धन्यवाद सैमसन। आपके शोध को विकसित होते देखना वास्तव में रोमांचक है और हम, आप जानते हैं, आपके करियर और आपके वैज्ञानिक अनुसंधान को रुचि के साथ देख रहे हैं। हम वास्तव में आपके काम को हमारे साथ साझा करने की सराहना करते हैं, और हमारे छात्र पेपर चुनौती जीतने पर फिर से बधाई देते हैं।