

[COVID Information Commons \(CIC\) Research Lightning Talk](#)



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

Title: [Générer des explications pour les prédictions de pneumonie par scanner thoracique](#)

[YouTube Recording with Slides](#)

[August 2021 CIC Webinar Information](#)

Transcript Editor: Macy Moujabber

Transcript

Lauren Close:

J'aimerais vous présenter notre dernier orateur, Samson Qian. Samson est l'un des trois lauréats du premier concours de rédaction pour étudiants de premier cycle du CIC, qui s'est déroulé au début du printemps. Nous souhaitons donc la bienvenue à Samson et nous sommes ravis de partager ses recherches en tant que chercheur émergent. Samson, allez-y, je vous laisse la parole.

Samson Qian:

Slide 1

Merci Lauren. Aujourd'hui, je vais faire une présentation légèrement différente des précédentes sur la génération d'explications pour les prédictions de l'apprentissage automatique sur la pneumonie virale et bactérienne.

Slide 2

L'objectif global de cette recherche est de prendre des radiographies pulmonaires de patients atteints de pneumonie bactérienne et virale ainsi que de patients sains et de construire un modèle d'apprentissage automatique pour classer ces patients à l'aide de ces images. Est-il possible de construire un modèle d'apprentissage automatique capable d'identifier avec précision ces différentes classes, mais aussi d'interpréter ce modèle d'apprentissage automatique afin de comprendre ce que le modèle regarde lorsqu'il génère des prédictions ? Cela introduit donc l'idée d'utiliser des algorithmes d'explicabilité, également connus sous le nom d'IA explicable, pour analyser un modèle et les données sur lesquelles il prédit afin de comprendre exactement où le modèle regarde. L'objectif ultime est de voir si ce type de cadre peut aider les radiologues à diagnostiquer différents patients.

Slide 3

Voici donc un bref aperçu de l'IA explicable. Il existe différents types d'algorithmes d'IA explicable, comme une branche familiale de différents types de méthodes. Dans le cadre de cette recherche, nous nous concentrerons sur les algorithmes post hoc, c'est-à-dire les algorithmes permettant d'analyser des modèles compliqués après qu'ils ont été formés sur les données, plutôt que pendant le processus ou avant. L'avantage de cette méthode est qu'elle permet d'utiliser des modèles plus complexes, tels que des réseaux neuronaux convolutionnels plus profonds, afin d'apprendre les caractéristiques de la radiographie. Il faut parfois faire un compromis entre l'interprétabilité et la précision. Les modèles plus profonds et plus complexes sont donc plus difficiles à interpréter que les modèles plus simples, et ces algorithmes permettent donc de comprendre ces modèles plus complexes. Le premier type d'algorithme est connu sous le nom de LRP (Layer-wise Relevance Propagation). Il s'agit d'une méthode dont je parlerai plus en détail un peu plus tard, mais elle consiste essentiellement à examiner les couches et les poids du modèle et à comprendre quels pixels de l'image contribuent le plus à l'activation de la structure interne d'un modèle.

La deuxième méthode, connue sous le nom de LIME, est agnostique, c'est-à-dire qu'elle ne dépend pas du type de modèle utilisé. Cette méthode diffère de la LRP en ce sens qu'elle part d'abord des données, plutôt que du modèle, et qu'elle examine des sous-sections individuelles des données pour trouver les régions de l'image qui contribuent le plus à la prédiction d'un modèle.

Le suivant s'appelle Grad-CAM et est similaire à LRP, mais au lieu d'examiner chaque couche et chaque poids, il examine les couches de convolution dans le modèle et prend en compte les gradients qui, lorsque vous adaptez une image, les gradients de la dernière couche de convolution, et produit ensuite une sorte de carte d'activation des gradients.

Le dernier type d'algorithme est un nouvel algorithme connu sous le nom de LRP contrastive. Il s'agit d'une modification de la LRP classique, mais il prend les pertinences et les différencie entre différentes classes telles que la pneumonie virale et bactérienne, ce qui permet de mieux visualiser la différence entre les classes.

Slide 4

La construction d'un réseau neuronal convolutif pour identifier les caractéristiques de ce type de patients nécessite une structure de modèle complexe telle que VGG16 et ResNet50, deux modèles de pointe qui donnent d'excellents résultats en matière de classification d'images et, comme vous pouvez le voir dans les structures de modèle, il s'agit d'un réseau convolutif très profond qui est capable d'identifier de nombreuses caractéristiques dans les images. Ces deux modèles donnent donc certains des meilleurs résultats en matière de classification entre les trois différentes classes de patients.

Slide 5

Les images radiographiques ont toutes été collectées à partir des données de Mendeleev et un certain nombre de prétraitements sont nécessaires pour les adapter au modèle. Il faut donc s'assurer qu'il y a une quantité similaire d'exemples d'images provenant de tous les types de classes, puis effectuer un prétraitement des images pour les intégrer au modèle

Slide 6

Voici un aperçu des performances du modèle au cours de la formation. À gauche, vous avez le graphique de précision pour l'ensemble d'entraînement dans l'ensemble de validation des images. Au fur et à mesure que vous entraînez le modèle, à chaque itération, à chaque époque, la précision augmente et la perte, que nous essayons de minimiser sur le côté droit, diminue régulièrement. Cela signifie que notre modèle identifie les caractéristiques des données, des images que nous lui transmettons, et qu'il fait un bon travail de classification entre les radiographies du thorax.

Slide 7

Il est très important d'avoir un modèle précis afin d'identifier ou d'examiner ce que le modèle regarde dans chaque image. Vous pouvez voir une matrice de confusion sur la gauche qui montre ce que le modèle confond et, en général, il fait un très bon travail entre les trois classes, à l'exception de la distinction entre la pneumonie virale et la pneumonie bactérienne, qui est confuse dans quelques exemples. Il est donc important d'exécuter ces algorithmes d'explicabilité pour comprendre ce que le modèle observe.

Slide 8

Voici un bref aperçu de la LRP (Layer-wise Relevance Propagation). Il s'agit essentiellement d'un réseau neuronal dont chaque couche est une couche cachée. Il propage votre image de la couche de sortie vers l'arrière. Au lieu de propager l'image vers l'avant, il la propage vers l'arrière, puis il calcule des scores pertinents pour chaque neurone, c'est-à-dire ces cercles ici. Ces scores pertinents représentent l'importance de chaque pixel de l'image pour la prédiction effectuée par le modèle.

Slide 9

Ici, vous pouvez jeter un coup d'œil - à droite, un exemple de LRP. Comme vous pouvez le voir, il produit une sorte de carte thermique basée sur les pixels individuels de l'image et qui montre l'importance de la contribution de chaque pixel de l'image à la prédiction du modèle sur la classe. À gauche, vous pouvez voir une comparaison avec LIME, qui est davantage une méthode basée sur les régions, qui prend des régions dans l'image et détermine quelles régions sont les plus importantes pour la prédiction d'un modèle.

Slide 10

Voici d'autres exemples de LIME pour les trois classes différentes. Ce que fait LIME, c'est qu'il ne va pas spécifiquement dans chaque couche du modèle pour examiner les poids, mais qu'il prend des sous-régions choisies spécifiquement sur l'image et qu'il exécute le modèle pour déterminer laquelle des sous-régions est la plus importante pour la prédiction du modèle et qu'il met l'accent sur ces régions spécifiques.

Slide 11

Un autre type d'algorithme dont j'ai parlé s'appelle Grad-CAM. Il s'agit d'un algorithme semblable à LRP, mais qui examine les gradients des couches convolutives de votre modèle. Cette méthode prend également une image et l'introduit dans le modèle, puis produit une carte d'activation, qui est une carte thermique représentant tous les gradients d'une couche convolutive et qui identifie exactement ce que le modèle regarde.

Slide 12

Voici quelques exemples de Grad-CAM. Il s'agit en quelque sorte d'une combinaison entre LRP et LIME. Il produit une carte thermique d'une région spécifique de l'image, et vous pouvez voir ce sur quoi le modèle se concentre et les régions sur lesquelles il ne se concentre pas.

Slide 13

La dernière variante, la plus contrastée, est une légère modification de la méthode LRP originale, mais elle prend les scores de pertinence et applique une modification qui distingue la pertinence entre différents types de classes. Ainsi, il est possible d'identifier ou de distinguer beaucoup plus clairement la pneumonie virale de la pneumonie bactérienne en utilisant ces différents types de scores de pertinence. Il est donc très important d'examiner et de comparer ces différents types de méthodes d'explicabilité afin de comprendre ce que votre modèle examine spécifiquement. Il existe donc de nombreux types d'algorithmes que vous pouvez utiliser, mais pour bien comprendre ce que votre modèle a appris - ce qu'il examine lorsqu'il fait des prédictions - il est important de comparer ces types de méthodes et d'élaborer un modèle précis pour aider les patients.

Slide 14

L'IA explicable est un domaine en constante évolution. De plus en plus d'algorithmes sont développés chaque jour et de plus en plus de données sont collectées pour construire des modèles plus précis. Différents types de données, comme les tomodensitogrammes, peuvent également aider les radiologues à diagnostiquer les patients. L'exécution de ces algorithmes sur différents types de modèles permet donc d'examiner quel modèle a appris les données avec plus de précision que les autres.

Slide 15

Je tiens à remercier le Dr Michael Pazzani, qui participe également à l'appel en ce moment, pour toute l'aide qu'il a apportée à la conduite de ses recherches, à la collecte des données et à la mise en œuvre de ces algorithmes d'identification de la pneumonie. Je vous remercie. Si vous souhaitez me contacter, voici mon adresse électronique.

Slide 16

Merci beaucoup Samson. C'est vraiment passionnant de voir votre recherche évoluer et nous allons, vous le savez, suivre votre carrière et vos recherches scientifiques avec intérêt. Nous apprécions vraiment que vous partagiez votre travail avec nous et nous vous félicitons.