

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

[Transcript of a Presentation by Ajitesh Srivastava \(University of Southern California\) and Viktor K Prasanna \(University of Southern California\), September 2020](#)



[Viktor K. Prasanna CIC Database Profile](#)

Title: *Accurate Predictions and Resource Allocation for COVID-19 Epidemic Response*

NSF Award #: [2027524](#)

[YouTube Recording with Slides](#)

[September 2020 CIC Webinar Information](#)

Transcript Editor: Julie Meunier

Transcript

Katie Naum:

Nous arrivons maintenant au tour de notre dernier intervenant Ajitesh Srivastava de l'université de Californie. Ajitesh, vous pouvez partager votre écran quand vous êtes prêt.

Slide 1

Ajitesh Srivastava:

Super, merci. J'espère que vous pouvez voir mon écran. D'accord. Je m'appelle Ajitesh Srivastava. Je suis le Co-chercheur principal du prix *Rapid Recover: Accurate Fine-grained Predictions and Resource Allocation for COVID-19 Epidemic Response (Guérir: Prédiction précises et allocation des ressources pour la réponse à l'épidémie COVID-19)*.

Slide 2

Les objectifs de ce projet sont les suivants: des prédictions fiables de l'évolution de la COVID-19 à différentes échelles: groupes de pays, Etats, pays, quartiers. Nous aimerions intégrer l'identification, à partir de données, d'un certain nombre de cas non déclarés qui ne sont pas réellement observés, mais qui affectent de manière significative les prévisions à long terme, et enfin, la manière d'utiliser efficacement ces prévisions pour la gestion des ressources pendant la pandémie.

Voici ce que nous avons réalisé jusqu'à présent. Nous disposons de prévisions COVID-19 précises. En quelques secondes, nous pouvons effectuer un entraînement et des prévisions pour 3 000 comtés en moins de 30 secondes. Nous fournissons des garanties de fiabilité sur les facteurs non observés. Par

exemple, l'ampleur de la sous-déclaration. Dans certains cas, nous pouvons identifier ce chiffre et garantir mathématiquement qu'il est proche de la vérité.

Nous avons un site internet accessible à tous et un répertoire Github. Nous fournissons également des évaluations hebdomadaires comparant notre approche aux modèles actuellement utilisés par le CDC. Nous fournissons au CDC nos prévisions de nombre de cas et de décès, incluant des prévisions hebdomadaires aux échelles nationale, fédérale et régionale. Elles sont ensuite utilisées pour informer le grand public et pour choisir les sites de test pour la vaccination.

Nous donnons aussi accès à nos prédictions pour développer des modèles ensemblistes à UMass et au KIT, auquel nous fournissons des prédictions à l'échelle de l'Allemagne et des Etats fédéraux. Voici les publications : deux d'entre elles sont des prépublications et l'une d'entre elles a été acceptée à la conférence KDD.

Slide 3

Notre idée principale est qu'un seul modèle n'est pas suffisant. Avoir un seul modèle complexe n'est pas suffisant. La prévision nécessite de prendre des décisions concernant la stratégie d'apprentissage utilisée, le type de prétraitement des données, les paramètres choisis et les hyperparamètres. Toutes ces décisions peuvent avoir une incidence considérable sur les résultats des prévisions.

En gardant cela à l'esprit, nous souhaiterions avoir un modèle qui capture beaucoup de complexité. Pourtant, nous préférons les stratégies d'apprentissage simples pour éviter le surapprentissage pour garantir que ce que nous apprenons est proche de la vérité et nous voulons vérifier mathématiquement la capacité d'apprentissage - si ce que nous prétendons avoir - s'est simplement adapté aux données passées ? Ou a-t-il réellement appris quelque chose ? Nous voudrions aussi avoir une exécution rapide pour permettre l'exploration de plusieurs scénarios. Vous pouvez lire cet article pour comprendre comment nous partons d'un modèle complexe que nous simplifions ensuite en un modèle mathématiquement équivalent facile à apprendre.

Slide 4

L'un des aspects centraux de notre travail est donc d'aborder la question de la facilité d'apprentissage. Malgré les simplifications, une équation de notre modèle est toujours non-linéaire. Nous pouvons donc toujours adapter vos données à n'importe quel modèle, mais celui-ci reflète-t-il la vérité ? Parce qu'il y a une multitude de solutions. Par exemple, ici pour la sous-déclaration, il y a quatre courbes. Ces quatre modèles représentent très bien les données passées mais conduisent à des résultats différents. Comment savoir laquelle de ces quatre courbes représente la vérité ? Nous avons donc prouvé que ce facteur de sous-déclaration peut être appris de manière fiable à partir de données uniquement sous certaines conditions et tous les détails sont dans cet article.

Slide 5

Nous avons une visualisation en ligne avec laquelle vous pouvez effectuer vous-même des prédictions et vous pouvez également envisager différents scénarios pour déterminer où nous allons et ce qui se passerait si nous faisons le meilleur effort possible par rapport à ce qui se passerait si nous faisons le pire effort possible, c'est-à-dire où nous nous trouverions dans quelques semaines.

Slide 6

Sur notre site, nous fournissons également des comparaisons de notre approche par rapport à d'autres approches du CDC et nous avons toujours été dans les deux ou trois premiers parmi les 25 modèles utilisés à l'heure actuelle.

Slide 7

D'autres travaux pertinents que nous envisageons sont l'apprentissage fiable des effets de diverses politiques. Nous souhaitons également résoudre certains problèmes d'affectation des ressources maintenant que nous disposons de prévisions précises. Ainsi, comment répartir les EPI (équipements de protection individuels) ou comment décider de l'emplacement des tests sur la carte et des sites de vaccination?

Et c'est sur ce point que je terminerai mon intervention. Je vous remercie.