**Hướng nghiên cứu: Xây dựng mô hình dự báo dịch Covid 2020**

**1, Vấn đề hiện nay**

Theo tư liệu 1 (Review current modeling publications), tính đến thời điểm 31/3/2020, các mô hình mô phỏng dự báo Covid trên thế giới theo hướng Epi Model đa phần làm theo dạng SIR/ SEIR. Tuy nhiên, do các mô hình này không tính đến khả năng thay đổi ảnh hưởng do tác dụng của các biện pháp cách ly do chính phủ phổ cập trong dân số nên các dự báo thường cao hơn dự kiến. Do vậy, chúng ta sẽ học hỏi một mô hình khác, được đưa ra trong tư liệu 2 (China Epimodel final phase). Tuy nhiên, có khó khăn được đặt ra trong việc ước lượng một số tham số cho mô hình như tốc độ cách ly, độ hiệu quả của việc cách ly, etc. Vì vậy, chúng ta có thể sẽ ước lượng các tham số này dựa trên việc phân tích thái độ dân chúng trên mạng xã hội với việc cách ly. Tóm tắt lại, hướng nghiên cứu hiện nay sẽ có hai bước:

**Hướng nghiên cứu 1: Thu thập phân tích dữ liệu đang có trên mạng xã hội về Covid-19 và phân tích thái độ dân số đang có về virus để ước tính độ hiệu quả của việc cách ly.**

**Hướng nghiên cứu 2: Sử dụng số liệu thu thập được từ hướng nghiên cứu 1 để dự đoán sự phát triển của virus trong thời gian tới**

**1, Hướng nghiên cứu 1: Social Media Analysis**

Mô hình phân tích của hướng nghiên cứu 1 gồm 3 bước:

**Bước 1 – Query tất cả các post trên Twitter tại một vùng địa lý có tồn tại những từ khóa liên quan đến Coronavirus. Query có thể được thực hiện qua data pipeline tại Mỹ hoặc thư viện Python Beautisoup/ ScrapePy**

**Bước 2 – Chuẩn hóa dữ liệu và phân tích các nhóm topic phổ biến sử dụng Latent Direchet Allocation (LDA)**

**Bước 3 – Sử dụng kết quả LDA từ bước 2 để chạy sentiment analysis để phân tích thái độ dân chúng về tác dụng của việc cách ly.**

Bước 1 và Bước 2 đã được làm và tóm tắt với dữ liệu Twitter tháng 1/2/3 trong Jupiter notebook, tư liệu 3 (social media analysis), cũng như trong bài báo tư liệu 4 (Twitter Model)

**2, Hướng nghiên cứu 2: Epidemic Modeling**

Trong epidemic modeling nói chung, có 2 bước:

**Bước 1 – chia dân số thành các nhóm vùng khác nhau mô phỏng bởi các biến thay đổi theo thời gian (ví dụ ở tư liệu 5 (Tóm tắt mô hình stochastic epidemic modeling))**

**Bước 2 – Thu thập các tham số cho mô hình, dựa trên mô phỏng xác suất hoặc số liệu thực tế (một số tham số sử dụng Monte Carlo Simulation, một số tham số thu được từ các thống kê quốc tế, và một số tham số thu được từ hướng nghiên cứu 1)**