

Nhóm 11

Thành Viên : Nguyễn Trần Hưng – 0950080122-Trưởng Nhóm

Trần Thanh Lâm-0950080127

Nguyễn Hoàng Minh Tuấn-0950080080

Lớp : 09CNTT04

BÀI BÁO CÁO

I. GIỚI THIỆU

Là người tạo dữ liệu di chuyển, con người tạo dữ liệu hàng ngày. Tất cả chúng ta được kết nối bằng cách chia sẻ dữ liệu từ các mạng xã hội

mạng, thiết bị thông minh, v.v. Các thiết bị viễn thám đã được sử dụng rộng rãi để quan sát hành tinh của chúng ta từ quan điểm khác nhau và để làm cho cuộc sống của chúng tôi dễ dàng hơn. Không sai khi nói rằng toàn bộ Trái đất hiện đang tồn tại được thực hiện bằng kỹ thuật số. Do đó, Trái đất công nghệ số quá trình tạo dữ liệu chuyển động là trình điều khiển chính cho lớn dữ liệu trong viễn thám, có thể được sử dụng để thực hiện hiệu quả hơn (ví dụ: cải thiện dịch vụ chẳng hạn như cảnh sát, y tế và giao thông vận tải) và cũng cho kinh doanh, tức là cải thiện quá trình ra quyết định, sản xuất, đổi mới sản phẩm, trải nghiệm và dịch vụ của người tiêu dùng, v.v.

Theo báo cáo của IBM, 2,5 triệu byte dữ liệu được bây giờ được tạo ra mỗi ngày. Nói cách khác, “90% dữ liệu trên thế giới ngày nay được tạo ra trong hai năm.”Chúng ta đang thực sự sống trong thời đại dữ liệu lớn, và giờ đây các nhà lãnh đạo chính phủ, doanh nghiệp và tổ chức phi lợi nhuận đang nhanh chóng nhận ra rằng điều đó rất quan trọng. Thu thập dữ liệu lớn trong các ngữ cảnh khác nhau. Tuy nhiên, vẫn còn tồn tại một vấn đề phổ biến liên quan đến cách chúng ta có thể

thu được vào dữ liệu lớn. Vấn đề này là một câu hỏi hóc búa: Trên một mặt, vô số dữ liệu lớn có thể mang lại cho chúng ta những cơ hội lớn. Mặt khác, chúng ta vẫn chưa biết cách khai thác lượng dữ liệu lớn như vậy với độ phức tạp, đa dạng và không đồng nhất rất lớn, nhưng vẫn có giá trị tiềm năng cao. Điều này làm cho dữ liệu rất khó xử lý và phân tích trong thời gian hợp lý.

Dữ liệu lớn có thể được đặc trưng chủ yếu bởi ba đặc điểm: khối lượng, sự đa dạng và tốc độ, được định nghĩa là ba chữ “V” kích thước bởi Meta Group (nay là Gartner) vào năm 2001. Nó đáng lưu ý rằng “giá trị” là một phẩm chất quan trọng của dữ liệu lớn, nhưng nó không phải là một đặc điểm xác định. Dữ liệu viễn thám lớn có thể được mô tả bằng các kích thước riêng của nó (sau đây gọi là 3V). Dữ liệu lưu trữ được đặc trưng bởi dung lượng ngày càng tăng, từ terabyte (TB $\frac{1}{4}$ 1024 GB) đến petabyte (PB $\frac{1}{4}$ 1024 TB) và thậm chí đến exa byte (EB $\frac{1}{4}$ 1024 PB). Ví dụ, một lượng lớn dữ liệu viễn thám hiện được cung cấp miễn phí có sẵn từ Chính phủ của NASA, chỉ một trong những tài liệu lưu trữ của NASA.

Hệ thống thông tin và dữ liệu khoa học trái đất (ESDIS), nắm giữ 7,5 PB dữ liệu với gần 7000 bộ dữ liệu duy nhất và 1,5 triệu người dùng trong

2013. Tập này chỉ chứa trong miền dữ liệu viễn thám. Về sự đa dạng, hiện nay chúng ta có thể thấy rằng dữ liệu viễn thám lớn bao gồm đa nguồn (laser, radar, quang học, v.v.), đa thời gian (được thu thập trên ngày khác nhau), và đa giải quyết (khác nhau độ phân giải không gian) cũng như dữ liệu viễn thám là dữ liệu từ các ngành khác nhau tùy thuộc vào một số lĩnh vực ứng dụng. Tốc độ của dữ liệu lớn trong viễn thám không chỉ liên quan đến việc tạo ra dữ liệu với tốc độ nhanh tốc độ tăng trưởng mà còn cả hiệu quả của quá trình xử lý và phân tích dữ liệu. Nói cách khác, dữ liệu phải được phân tích trong một (gần như) thực hoặc hợp lý thời gian để đạt được một nhiệm vụ nhất định, ví dụ: vài giây có thể cứu hàng trăm ngàn

sinh mạng trong một động đất. Mặc dù 3V có thể mô tả dữ liệu lớn, chúng tôi xem xét rằng dữ liệu lớn trong viễn thám là không cần thiết

thỏa mãn cả ba chiều V. Ví dụ, bất kỳ một của khối lượng và vận tốc, khối lượng và sự đa dạng, hoặc sự đa dạng và vận tốc đã có thể xác định một vấn đề dữ liệu lớn. Ngoài trừ những thách thức chung của dữ liệu lớn được đặc trưng bởi 3V, có những thách thức khác đối với điều khiển từ xa các ứng dụng cảm biến, chẳng hạn như khả năng mở rộng để tích hợp

nhiều hệ thống quản lý khác nhau cho các vệ tinh khác nhau cho một trung tâm dữ liệu viễn thám. Đặc biệt tầm quan trọng là giá trị của dữ liệu, một chất lượng quan trọng ẩn trong dữ liệu lớn. Các phương pháp xử lý dữ liệu có thể được sử dụng để khám phá giá trị như vậy, và sau đó là giá trị của lớn dữ liệu có thể được thực hiện trong một ứng dụng viễn thám thực sự.

Do đó, để hiểu rõ hơn về dữ liệu lớn, cần thống nhất ba quan điểm, tức là sở hữu dữ liệu, ứng dụng dữ liệu và phương pháp dữ liệu. Trong bài báo, một bộ ba khung được đề xuất để hiểu rõ hơn về dữ liệu lớn trong

bối cảnh của các ứng dụng viễn thám. Tất cả các khía cạnh của bộ ba như vậy chia sẻ những thách thức chung và khác nhau quan điểm có những thách thức cá nhân của riêng mình. Trong công việc này, những thách thức chung và cá nhân được thảo luận trong ngữ cảnh của các ứng dụng viễn thám. Bất chấp những thách thức lớn như vậy, những tiềm năng lớn

dữ liệu viễn thám được trình bày chi tiết. Những tiềm năng này đã được áp dụng để giải quyết các vấn đề trong thế giới thực khác nhau. Vấn đề ở đây, chẳng hạn như khảo cổ học, đánh giá cây trồng và dự báo năng suất, an ninh lương thực, sức khỏe con người, phát triển và sử dụng đất, quy hoạch đô thị, quản lý và tính bền vững, giám sát rừng, nghiên cứu chiến tranh và xung đột và một số nghiên cứu khác.

Để minh họa hiệu quả của viễn thám lớn dữ liệu, hai trường hợp nghiên cứu thảo luận về việc sử dụng dữ liệu lớn trong viễn thám được thể hiện trong bài báo này. Bên trong trường hợp thử nghiệm đầu tiên, dữ liệu mạng xã hội cùng với điều khiển từ xa hình ảnh cảm biến được xác định bao gồm điều khiển từ xa lớn dữ liệu cảm biến để phát hiện tràn dầu biển tự động và sau đó một phương pháp dữ liệu mới được áp dụng để giải quyết với những thách thức ghi nhận. Trong trường hợp thử nghiệm thứ hai, việc truy xuất thông tin dựa trên nội dung được thực hiện bằng cách sử dụng điện toán hiệu năng cao (HPC) để trích xuất thông tin từ cơ sở dữ liệu lớn về hình ảnh viễn thám, được thu thập sau vụ tấn công khủng bố vào Thương mại Thế giới Trung tâm tại thành phố New York vào ngày 11 tháng 9 năm 2001.

Phần còn lại của bài báo được tổ chức như sau. Phần tiếp theo thảo luận về hiểu biết của chúng ta về dữ liệu lớn trong viễn thám từ ba quan điểm khác nhau. Theo quan điểm của chúng tôi về dữ liệu lớn, Phần III chia lớn thách thức dữ liệu thành thách thức chung cho tất cả các điều khiển từ xa các ứng dụng cảm biến và những thách thức riêng lẻ trong các khía cạnh riêng lẻ của cái gọi là bộ ba dữ liệu lớn. Sau đó, tiềm năng của dữ liệu viễn thám lớn được trình bày trong Mục IV. Mục V trình bày vụ việc nói trên nghiên cứu về dữ liệu lớn trong viễn thám. Cuối cùng, Phần VI

rút ra một số kết luận của công việc và thảo luận về tương lai diễn biến.

II: TÌM HIỂU DỮ LIỆU LỚN TRONG VIỄN THÁM

Nhìn từ góc độ chung, chúng ta có thể hiểu dữ liệu lớn như có ý nghĩa khác nhau liên quan đến những người sở hữu dữ liệu lớn, những người có thể xử lý và phân tích dữ liệu lớn, và những người sử dụng dữ liệu lớn. Theo đó, các phương pháp dữ liệu khác nhau có thể được khai thác để giải quyết các vấn đề lớn dữ liệu những thách thức để khai thác hiệu quả giá trị của những dữ liệu đó. Sau đây, một bộ ba (ba trong một) là được thảo luận để hiểu về dữ liệu lớn (đặc biệt tập trung vào các ứng dụng viễn thám). Ở đây, chúng tôi xác định ba khía cạnh để hiểu dữ

liệu lớn, tức là, sở hữu dữ liệu, phương pháp dữ liệu và ứng dụng dữ liệu, đóng góp cùng nhau vào một vòng đời dữ liệu lớn duy nhất. bộ ba khái niệm về dữ liệu lớn.

Có những thách thức chung và khác nhau trong các khía cạnh riêng lẻ của hiệu dữ liệu lớn, được trình bày chi tiết tiếp theo.

A. Khía cạnh đầu tiên: Sở hữu dữ liệu

Đây là một khía cạnh quan trọng của dữ liệu lớn dựa trên mà chúng ta có thể xác định các ứng dụng và sử dụng hoặc thiết kế phương pháp dữ liệu thích hợp để giải quyết một vấn đề thực tế Các cơ hội tương ứng dựa trên thực tế là dữ liệu đa dạng hơn có thể được mua lại bởi các thiết bị thông minh nơi hầu hết con người hiện có quyền truy cập internet để trở thành công cụ tạo dữ liệu cá nhân và chuyển động. Theo đó, dữ liệu các giá trị có thể được lấy từ các dữ liệu viễn thám phức tạp, đa dạng, không đồng nhất và nhiều chiều đó và các dữ liệu khác từ không gian mạng.

Tuy nhiên, thách thức lớn phát sinh ở mỗi bước khi có được và tổ chức lớn dữ liệu viễn thám. Ví dụ, dữ liệu viễn thám thu được từ vệ tinh, máy bay hoặc cảm biến khác thiết bị trong khi các dạng dữ liệu khác được truy xuất từ không gian mạng. Dữ liệu viễn thám được tiền xử lý bằng hiệu chỉnh hình học và phóng xạ, hội nghị địa lý, loại bỏ tiếng ồn, v.v. [18], và dữ liệu từ không gian mạng cần được làm sạch để giảm lỗi và tiếng ồn, trong chất lượng dữ liệu nào có thể được cải thiện.

Viễn thám dữ liệu phải được truyền từ vệ tinh đến trạm mặt đất và từ trạm mặt đất đến khách hàng. Các vấn đề liên quan khác là nén dữ liệu, lưu trữ dữ liệu, dữ liệu truy xuất, quyền và bảo vệ dữ liệu, v.v. chúng tôi nhấn mạnh dữ liệu đó không có giá trị cho đến khi chúng được sử dụng cho các ứng dụng. Sự khác biệt chính giữa dữ liệu truyền thống và dữ liệu lớn là cách xác định đúng tập dữ liệu và làm thế nào để kết hợp chúng để giải quyết một thách thức hoặc tiểu thuyết vấn đề.

B. Khía cạnh thứ hai: Phương pháp dữ liệu lớn

Một phương pháp dữ liệu lớn nên được thiết kế để giải quyết một cách có hệ thống các vấn đề dữ liệu lớn từ các lĩnh vực viễn thám khác nhau.

Phương pháp như vậy được sử dụng để thiết kế các phương pháp dữ liệu mới cho dữ liệu viễn thám lớn chuẩn bị, triển khai dữ liệu, khai thác thông tin, mô hình hóa dữ liệu, hợp nhất dữ liệu, trực quan hóa dữ liệu và dữ liệu diễn dịch. Những khía cạnh này đặc biệt quan trọng trong ứng dụng viễn thám, trong đó tiền xử lý các

bước cũng quan trọng như khai thác thông tin các bước. Tuy nhiên, xử lý và phân tích dữ liệu đại diện cho một các phương pháp điều khiển dữ liệu và đường ống nhiều bước có thể khác biệt đáng kể so với quan điểm của các ứng dụng và miền cụ thể. Do tính không đồng nhất nói trên và chiều cao của dữ liệu lớn trong viễn thám, chúng tôi cũng đối mặt với những thách thức tính toán và thống kê quan trọng liên quan đến khả năng mở rộng xử lý, tích tụ nhiều, tương quan giả, nội sinh ngẫu nhiên và sai số đo lường [19], [20].

Những thách thức này đòi hỏi mới kỹ thuật tính toán và thống kê để giải quyết vấn đề phân tích và xử lý dữ liệu lớn. Việc phân tích và kỹ thuật xử lý được định hướng dữ liệu và có thể mang lại lợi ích từ các lý thuyết và phương pháp từ các lĩnh vực thống kê, học máy, nhận dạng mẫu, trí tuệ nhân tạo, khai thác dữ liệu, v.v. Kiến thức tên miền là một thứ khác khía cạnh quan trọng cần được liên kết chặt chẽ với dữ liệu phân tích.

C. Khía cạnh thứ ba: Ứng dụng dữ liệu lớn

Mục tiêu chính trong các ứng dụng dữ liệu lớn là xác định dữ liệu phù hợp để giải quyết các vấn đề trong tầm tay, khó giải quyết hoặc hầu như không thể bị thao túng bởi dữ liệu viễn thám truyền thống. Sau đó, vấn đề tiếp theo là làm thế nào để thu thập, sắp xếp và sử dụng những dữ liệu lớn này để đối phó với các vấn đề viễn thám thực tế. Để xác định đúng dữ liệu, chúng ta nên liên kết chặt chẽ đến khía cạnh đầu tiên của việc hiểu dữ liệu lớn. Nói cách khác, để khai thác dữ liệu lớn trước tiên người ta phải lấy dữ liệu từ các tác nhân dữ liệu liên quan (hay nói chung là ngành dữ liệu hoặc tổ chức). Để truy cập dữ liệu, cần tính đến sự hợp tác giữa các miền hoặc tổ chức tài khoản một cách hiệu quả. Đây là một trong những điều cốt yếu thách thức trong ứng dụng viễn thám.

Sau khi có được dữ liệu phù hợp, chẳng hạn như viễn thám dữ liệu, dữ liệu văn bản và hình ảnh từ các mạng xã hội, các phương pháp dữ liệu sáng tạo nên được phát triển để khám phá, nhận ra, và chứng minh giá trị của dữ liệu lớn cho ứng dụng viễn thám.

III. DỮ LIỆU LỚN, THÁCH THỨC LỚN

Những thách thức của dữ liệu lớn trong viễn thám không chỉ liên quan đến việc xử lý khối lượng dữ liệu lớn [21]. Đặc biệt, những thách thức về thu thập, lưu trữ, quản lý và phân tích dữ liệu cũng liên quan đến các vấn đề viễn thám liên quan đến dữ liệu lớn. Trong phần này, chúng tôi đặc biệt phân tích những thách thức của dữ liệu

lớn trong viễn thám liên quan đến các khía cạnh khác nhau của việc hiểu dữ liệu lớn trong phần trước.

Từ những quan điểm khác nhau để hiểu về dữ liệu lớn, chúng ta đang đối mặt với những thách thức lớn trong việc tận dụng giá trị mà dữ liệu mang lại. Trong ba khía cạnh, những thách thức giống nhau được chia sẻ, chẳng hạn như điện toán dữ liệu, cộng tác dữ liệu và phương pháp dữ liệu cho các ứng dụng khác nhau; trong khi chờ đợi, chúng ta đang phải đối mặt với những thách thức khác nhau trong các khía cạnh riêng lẻ của việc hiểu dữ liệu lớn. Hình 2 tóm tắt những thách thức phổ biến và khác nhau, được mô tả chi tiết trong các phần tiếp theo.

A. Những thách thức chung:

Sau đây, ba thách thức phổ biến, Đó là: tính toán dữ liệu lớn, cộng tác dữ liệu lớn và dữ liệu lớn, các phương pháp, được liệt kê theo bộ ba hiểu biết về dữ liệu lớn trong viễn thám.

1) Điện toán dữ liệu lớn: Một thách thức trong việc thiết kế hệ thống hiệu suất cao cho tính toán dữ liệu lớn là để phát triển các hệ thống không đồng nhất hơn có khả năng tích hợp các nguồn tài nguyên ở các địa điểm khác nhau [22]. Mặc dù các hệ thống đặt điện toán đám mây đã được chứng minh là đạt được mức độ cao, hiệu suất tổng hợp trong các ứng dụng viễn thám vẫn còn những thách thức liên quan đến việc kết hợp dần dần khái niệm điện toán đám mây

nguyên cứu viễn thám [23]. Mục tiêu cuối cùng là làm cho các bộ sưu tập dữ liệu dễ dàng truy cập từ những người dùng khác nhau. Tuy nhiên, một thách thức còn lại là mức tiêu thụ năng lượng, vẫn còn khó để tận dụng tối đa trong các nền tảng song song ồ ạt hoặc thậm chí trên tàu các kịch bản xử lý. Giải quyết những thách thức này là quan trọng đối với việc kết hợp đầy đủ các kỹ thuật tính toán dữ liệu lớn vào các ứng dụng viễn thám. Văn chương đối với dữ liệu lớn trong viễn thám chủ yếu tập trung vào vấn đề đồ sộ của tính toán dữ liệu lớn và coi đó là

một vấn đề tính toán cần nhiều dữ liệu [24]. Thông thường, một Mô hình HPC được khai thác cho (gần như) dữ liệu lớn thời gian thực xử lý [20], [23], [25].

2) Hợp tác dữ liệu lớn: Quyền sở hữu dữ liệu trong các vấn đề viễn thám thường bị phân mảnh giữa các đại lý hoặc ngành công nghiệp dữ liệu [26]. Theo đó, truy cập và kết nối dữ liệu có thể là một trở ngại. Có thể nêu lên các mối quan tâm hợp pháp để đạt được sự hợp tác liên ngành thúc đẩy chia sẻ dữ liệu, chẳng hạn như văn bản

xã hội hoặc phương tiện truyền thông xã hội. Tuy nhiên, các cá nhân thường từ chối chia sẻ thông tin cá nhân dữ liệu do bảo mật và quyền riêng tư. Điều này mâu thuẫn với ý tưởng cá nhân hóa dữ liệu. Ngoài ra, nhiều công ty dữ liệu coi dữ liệu lớn là độc quyền và do đó không có động cơ chia sẻ dữ liệu. Đồng thời, đây là một thách thức quan trọng đối với các cơ quan chính phủ trong việc chia sẻ dữ liệu trừ khi tất cả những người tham gia có thể đạt được lợi ích vật chất và các ưu đãi trong việc chia sẻ dữ liệu lớn hơn các rủi ro [27]. Chẳng hạn, ngay cả khi NASA hiện đang chia sẻ một lượng đáng kể dữ liệu viễn thám theo sáng kiến chính phủ mở, 4 hầu hết các hình ảnh có độ phân giải không gian cao, chất lượng cao vẫn không có sẵn cho công chúng.

Do đó, cần phải tìm ra những cách cộng tác mới để cải thiện khả năng truy cập dữ liệu lớn trong các bài toán viễn thám

3) Phương pháp Dữ liệu lớn: Vấn đề phân tích

dữ liệu lớn trong viễn thám có thể được chính thức hóa đơn giản là

theo sau. Gọi X là tập dữ liệu đầu vào và gọi là $f(x)$

chức năng ánh xạ giữa đầu vào $x \in X$ và đầu ra y . Sau đó, một nhiệm vụ phân tích dữ liệu chung có thể được xây dựng thành:

$$y = f(x)$$

nơi xử lý tương ứng có thể được thực hiện trong bộ nhớ của máy tính chứa dữ liệu đầu vào.

Tuy nhiên, phân tích dữ liệu lớn nói chung nên áp dụng một cơ chế phân vùng đầu vào dữ liệu thành một phân tán và hoặc kiến trúc song song, tức là, $X = \{X_1; X_2; \dots; X_n\}$ có nghĩa là chia tập hợp lớn hơn X thành N dữ liệu nhỏ hơn bộ. Các phương pháp hoặc thuật toán dữ liệu được thông qua, tức là, $f(.)$ nên được sửa đổi để đáp ứng các môi trường máy tính mới. Mặc dù nói chung đây là một sự đơn giản hóa

(vì các tập dữ liệu nhỏ hơn có thể không dễ xử lý độc lập và liên quan đến một số đồng bộ hóa và/hoặc giao tiếp trong tác vụ xử lý liên quan), một thách thức quan trọng đối với sơ đồ xử lý này là không phải tất cả các thuật toán thoát đều có thể được phân phối hoặc triển khai hiệu quả trong dạng song song. Ngay cả khi các phương pháp xử lý dữ liệu có thể làm như vậy, thì việc thu thập dữ liệu phân tán và phân phối dữ liệu đó đến đúng nút điện toán là một thách thức. Do đó, xử lý dữ

liệu lớn nói chung (và viễn thám nói riêng) cần các mô hình thống kê và tính toán mới liên quan đến các chiến lược xử lý dữ liệu tiêu chuẩn

B. Thử Thách Cá Nhân

Trong phần này, một vài thách thức quan trọng được thảo luận trong bối cảnh thiết kế tuổi thọ dữ liệu viễn thám lớn như kỳ (xem Hình 3). Sau khi hiểu được nhu cầu kinh doanh (ví dụ: ứng dụng viễn thám liên quan đến dữ liệu lớn), một số bước quan trọng là xác định đúng loại dữ liệu trên các lĩnh vực khác nhau, để triển khai dữ liệu lớn, để sử dụng hoặc thiết kế các phương pháp dữ liệu sáng tạo và cuối cùng là trực quan hóa và giải thích các kết quả thu được. Ở đây, các phương pháp dữ liệu bao gồm phân tích dữ liệu, mô hình hóa dữ liệu, xử lý dữ liệu, v.v.

1) Nhận dạng dữ liệu phù hợp: Dữ liệu viễn thám lớn

thường bao gồm dữ liệu trong miền và dữ liệu ngoài miền. Trong

trước đây, các định dạng dữ liệu khác nhau hiếm khi được kết hợp để đáp ứng các ứng dụng viễn thám/nhiệm vụ. Do đó, dữ liệu không có giá trị trừ khi được khai thác để hoàn thành một nhiệm vụ cụ thể. Theo đó, sự khác biệt chính giữa dữ liệu viễn thám truyền thống và dữ liệu viễn thám lớn nằm ở cách lựa chọn và kết hợp các định dạng dữ liệu khác nhau để giải quyết các vấn đề trong thế giới thực trước đây được cho là nan giải. Đây là một thách thức chính của dữ liệu lớn, tức là làm thế nào để xác định và khai thác dữ liệu thích hợp để giải quyết vấn đề hiện tại.

Trong viễn thám, chúng ta có nhiều loại khác nhau của dữ liệu [28], bao gồm quang học (ví dụ: đa phổ và siêu phổ), radar [ví dụ: radar khẩu độ tổng hợp (SAR)] hoặc laser [ví dụ: phát hiện ánh sáng và phạm vi (LiDAR)] được cung cấp bởi máy bay hoặc vệ tinh hoặc cảm biến mặt đất. Các loại nguồn dữ liệu khác cũng có thể được tích hợp trong điều khiển từ xa

các vấn đề về cảm biến, tức là dữ liệu dạng văn bản trên internet (ví dụ: tin tức, nhật ký web, v.v.) có thể được sử dụng để giúp ghi nhận các mẫu dữ liệu được cung cấp bởi các cảm biến từ xa [3], chẳng hạn như thông qua học tích cực [29] hoặc kỹ thuật cung cấp dịch vụ cộng đồng [30], liên quan đến chi phí thấp hoặc không có. Ngoài ra, dữ liệu hình ảnh do các cá nhân chụp từ mạng xã hội có thể được tính đến để hỗ trợ các nhiệm vụ diễn giải dữ liệu viễn thám. Các định dạng dữ liệu khác như dữ liệu điều tra dân số, dữ liệu khí tượng, dữ liệu giao thông thông minh, dữ liệu đồ họa địa lý có độ trung thực cao, dữ liệu chăm sóc sức khỏe, v.v., có thể giúp

ích đáng kể để giải quyết một vấn đề cụ thể trong thế giới thực, ví dụ: giám sát an ninh lương thực [31]

2) Những thách thức trong việc sở hữu dữ liệu: Sau khi dữ liệu có

được truyền đến trạm mặt đất, những dữ liệu đó sẽ được lưu trữ trong một hệ thống. Một hệ thống lưu trữ dữ liệu thường bao gồm các thành phần phần cứng và phần mềm. Về phần cứng, cơ sở hạ tầng phần cứng phải được điều chỉnh linh hoạt cho các môi trường ứng dụng khác nhau. Về phần mềm, hệ thống lưu trữ dữ liệu thường được trang bị nhiều giao diện, kho lưu trữ dữ liệu, truy vấn từ các dịch vụ web cho tương tác của người dùng. Với sự phát triển nhanh chóng của dữ liệu viễn thám, các hệ thống quản lý cơ sở dữ liệu liên quan đến cấu trúc truyền thống (RDBMS) không thể đáp ứng các yêu cầu quản lý dữ liệu lớn trong viễn thám. Theo đó, điều cấp bách là áp dụng hoặc thiết kế một hệ thống lưu trữ dữ liệu mới có thể đáp ứng sự tăng trưởng nhanh chóng của dữ liệu viễn thám lớn ở quy mô PB hoặc lớn hơn. Thảo luận chung về lưu trữ dữ liệu lớn này có thể được tham khảo [32].

Phân phối dữ liệu cung cấp quyền truy cập vào dữ liệu viễn thám

và siêu dữ liệu cho người dùng, cả tại các trạm mặt đất chính hoặc

mạng thu trạm mặt đất. Thông thường, điều này bao gồm các cổng web đồ họa cung cấp quyền truy cập vào dữ liệu và tìm kiếm siêu dữ liệu cho người dùng. Theo truyền thống, người dùng tải dữ liệu quan tâm từ kho lưu trữ trung tâm về máy tính cục bộ của họ để phân tích. Điều này không thể hoạt động trong các ứng dụng dữ liệu lớn vì kích thước dữ liệu tăng mạnh không thể cho phép hệ thống hiện tại cung cấp dữ liệu cho người dùng để tính toán cục bộ. Đặc biệt, nếu một trường hợp khẩn cấp như một trận động đất xảy ra, một lượng lớn dữ liệu sẽ được nhận để phân tích dữ liệu trong một thời gian rất ngắn, vì ít giây có thể cứu sống nhiều người bằng những cảnh báo kịp thời. Đây là một thách thức lớn khác đối với những người sở hữu dữ liệu lớn, vì dữ liệu cực kỳ đa dạng và nhiều chiều cần được xử lý và phân tích trong một khoảng thời gian ngắn do thuộc tính khối lượng và vận tốc của dữ liệu lớn. Do đó, một nền phát triển nền tảng phân tích dữ liệu lớn theo thời gian thực để xử lý dữ liệu viễn thám trực tuyến cùng với dữ liệu ngoại tuyến trong trung tâm dữ liệu cục bộ hoặc từ dữ liệu phân tán trung tâm cho một ứng dụng thời gian thực, chẳng hạn như dự báo thời tiết, cảnh báo nguy hiểm, v.v.

3) Triển khai dữ liệu: Như đã thảo luận trong Phần III-B1, một thách thức quan trọng là xác định nguồn dữ liệu thích hợp để đạt được một mục tiêu cụ thể khó có thể hoàn thành nếu không có dữ liệu lớn. Một thách thức khác của dữ liệu lớn là làm thế nào để triển khai dữ liệu cho các ứng dụng thực tế. Trong giai đoạn ứng dụng dữ liệu lớn, việc triển khai dữ liệu lớn bao gồm chuẩn bị dữ liệu, công nghệ quản lý dữ liệu, phương pháp dữ liệu và kỹ thuật, và như vậy. Đó là, làm thế nào để có được dữ liệu, cách lưu trữ dữ liệu trong môi trường máy tính, và cách xây dựng các mô hình để hiểu sâu hơn về dữ liệu lớn nên được thiết kế cẩn thận trong bước triển khai dữ liệu lớn. Do các đặc tính về khối lượng và vận tốc của dữ liệu lớn, các phương pháp truyền thống không thể được sử dụng cho mục đích triển khai. Theo đó, các công nghệ mới nên được áp dụng tính đến, chẳng hạn như công nghệ quản lý dữ liệu phân tán, mô hình dữ liệu không có lược đồ, trực quan hóa tích cực kỹ thuật, v.v., để có được những hiểu biết có ý nghĩa về dữ liệu lớn

4) Biểu diễn dữ liệu: Các nguồn dữ liệu viễn thám khác nhau có độ phân giải không gian và quang phổ khác nhau và thường được thu thập vào các ngày khác nhau [33].

Ví dụ, trong dữ liệu quang học, chữ ký quang phổ của mọi vật liệu là duy nhất trong phép đo phòng thí nghiệm. Tuy nhiên, chữ ký phổ của dữ liệu trường có thể thay đổi do sự thay đổi của vật liệu, tác động môi trường, chất gây ô nhiễm bề mặt, tác động lân cận của các vật thể gần đó, thay đổi theo mùa, v.v. [34]. Điều này có thể dẫn đến hiện tượng các chữ ký giống nhau có thể biểu thị các đối tượng khác nhau, trong khi các chữ ký khác nhau có thể biểu thị cùng một đối tượng. Hiện tượng này tương tự như “khoảng cách ngữ nghĩa” được quan sát thấy trong thị giác máy tính, tức là, sự khác biệt giữa thông tin đến từ dữ liệu và khía cạnh tri thức được người dùng diễn giải [35]. Trong những năm gần đây, mạng lưới thần kinh sâu (DNNs) [36] đã giải quyết thành công vấn đề phân loại trong thị giác máy tính để lấp đầy khoảng trống ngữ nghĩa thông qua trích xuất tính năng tự động một cách sâu sắc [37]. Mặc dù DNN đã được thông qua

đối với các nhiệm vụ lựa chọn và phân loại đối tượng khi phân tích ảnh viễn thám, trong hầu hết các công trình được thực hiện cho đến nay, chỉ các đối tượng phổ hoặc các đối tượng phổ đã biến đổi (ví dụ: các thành phần chính) được sử dụng làm đầu vào cho DNN để tạo ra một “hiệu quả tốt hơn”. ” biểu diễn ảnh viễn thám [38]–[40]. trong ấn bản làm việc, thông thường các thành phần từ các vectơ đặc trưng

ban đầu tương ứng với hệ số phản xạ phổ của các mục tiêu che phủ mặt đất, có nghĩa là các thành phần đặc trưng có ý nghĩa vật lý rõ ràng. Về vấn đề này, ảnh viễn thám có thể được hiểu là dữ liệu có cấu trúc. Mặc dù độ chính xác và độ bền của phân loại có thể được cải thiện một chút bằng cách kết hợp các mẫu không được gắn nhãn để học tính năng, nhưng không rõ hiệu suất phân loại có thể được hưởng lợi bao nhiêu từ biểu diễn tính năng “tốt hơn” do thiếu lượng lớn dữ liệu huấn luyện và số lượng lớp hạn chế có thể được sử dụng trong thực tế khi triển khai mạng nơ-ron. Chẳng hạn, độ chính xác phân loại không thể được cải thiện đáng kể với hơn năm lớp như được mô tả về số lớp được thông qua trong [40].

Ngoài ra, mặc dù nhiều loại dữ liệu viễn thám khác nhau (được thu thập bởi các cảm biến khác nhau, từ các vị trí khác nhau vào các ngày khác nhau) được thu thập và khai thác để xử lý một vấn đề ứng dụng đầy thách thức cùng với dữ liệu viễn thám, nhưng các phương pháp dữ liệu hiện có không thể thao tác với các dữ liệu đó. data để lấy giá trị của những dữ liệu đó. Trong khi đó, dữ liệu viễn thám bao gồm các kích thước và độ phân giải không gian khác nhau, chẳng hạn như hình ảnh độ phân giải vừa phải đa bán cầu trong không gian quang phổ kế (MODIS) trong 36 dải quang phổ với không gian mặt đất 250 m (dải 1–2), 500 m (dải 3–7), và 1000 m (dải 8–36)⁵ và siêu quang phổ trong không khí máy quang phổ hình ảnh hệ thống quang học phản xạ (ROSIS) với độ phân giải mặt đất nhỏ hơn 1 m trong 115 băng tần)⁶ Hơn nữa, các biểu diễn của dữ liệu ngoài viễn thám có thể không có cấu trúc (ví dụ: các ảnh riêng lẻ), khác biệt đáng kể so với dữ liệu viễn thám quang học hoặc vi sóng. Do đó, việc biểu diễn dữ liệu khác nhau trở thành trở ngại lớn cho việc khai thác dữ liệu viễn thám lớn

5) Kết hợp dữ liệu: Do thách thức về biểu diễn dữ liệu được thảo luận trong Phần III-B4, một thách thức tiếp theo là làm thế nào để tích hợp dữ liệu từ nhiều nguồn khác nhau, trong đó các đặc điểm dữ liệu khác nhau đáng kể (ví dụ: chữ ký quang phổ trong viễn thám quang học dữ liệu, bức xạ điện từ trong dữ liệu vi sóng, đặc điểm cấu trúc của văn bản, đặc điểm phi cấu trúc của hình ảnh bằng máy ảnh kỹ thuật số, v.v.).

Theo truyền thống, hợp nhất dữ liệu có thể được thực hiện dưới dạng hợp nhất cấp độ pixel, hợp nhất cấp độ tính năng và hợp nhất cấp độ quyết định [41]. Tuy nhiên, dữ liệu lớn trong viễn thám thường bao gồm các quy mô và/hoặc định dạng khác nhau. Do đó, các phương pháp truyền thống không thể được sử dụng để tích

hợp thông tin cho hợp nhất dữ liệu. Do đó, các phương pháp mới nên được phát triển để giải quyết sự hợp nhất của dữ liệu lớn trong viễn thám. Ví dụ: trong các ứng dụng đô thị, mỗi pixel có thể được chú thích bằng các bức ảnh được chụp bởi các cá nhân từ mạng xã hội ở cùng một vị trí [3] bằng kỹ thuật cung cấp dịch vụ cộng đồng [30]. Việc đo lường mối tương quan giữa các nguồn dữ liệu khác nhau cũng trở thành một thách thức bổ sung khi có sự trợ giúp của trí tuệ nhân tạo, khai thác dữ liệu, học máy hoặc thống kê.

IV. CƠ HỘI LỚN

Bất chấp những thách thức lớn nói trên, tiềm năng giá trị của dữ liệu viễn thám lớn là rất ấn tượng. Thực ra, kỹ thuật viễn thám đã được sử dụng thành công cho các ứng dụng khác nhau, chẳng hạn như các ứng dụng nông nghiệp (ví dụ: giám sát an ninh lương thực, giám sát đồng cỏ), các ứng dụng đại dương (ví dụ: phát hiện tàu, phát hiện tràn dầu), quy hoạch đô thị, giám sát đô thị, định cư của con người (cả thành thị và nông thôn), giám sát an ninh lương thực, giám sát chất lượng nước, đánh giá năng lượng, dân số bệnh tật, đánh giá hệ sinh thái, nóng lên toàn cầu, toàn cầu thay đổi, đánh giá tài nguyên rừng toàn cầu, cảnh cổ xưa khám phá (khảo cổ học), v.v.

Kết hợp với các hoạt động của con người và dữ liệu từ xã hội khoa học, kỹ thuật viễn thám được liệt kê ở trên đã trở thành công cụ mạnh mẽ hơn nhiều để cải thiện đáng kể hiệu quả của sản xuất và hoạt động cho con người phúc lợi. Bằng cách này, dữ liệu viễn thám lớn cung cấp khả năng hoàn thành các mục tiêu khó hoặc không thể đạt được theo cách truyền thống. Ví dụ, một di tích ẩn có thể được tìm thấy bằng điều khiển từ xa có độ phân giải cao cảm nhận dữ liệu trong một khu rừng rậm không có cơ sở hạ tầng hiện đại, đó là một rào cản đáng kinh ngạc cho các nhà khảo cổ học thực địa để thâm nhập. Một ứng dụng thành công là trên Maya nghiên cứu ở vùng Peten phía bắc Guatemala [5].

Trong các ứng dụng quy hoạch đô thị, phép đo mặt đất cũng như các hình ảnh viễn thám trong không gian và trên không được tích hợp để mang lại kết quả tốt hơn và kịp thời cho đô thị lập kế hoạch, quản lý và tính bền vững [12], [14], [15]. Trong ngữ cảnh này, dữ liệu viễn thám có thể được thu thập qua một khu vực rộng lớn trong một chuỗi ở độ phân giải rất cao (1 E., dưới 1 m/pixel) sử dụng kỹ thuật viễn thám tiên tiến. Các dự án liên quan bao gồm dự án 100 thành phố cho đặc tính môi trường đô thị, giám sát và ra quyết định của chính phủ, và dấu ấn đô

thị toàn cầu sử dụng độ phân giải không gian rất cao với tổng số 180000 Cảnh TerraSAR-X và TanDEM-X cho toàn cầu lập bản đồ các khu định cư. Kết hợp với số liệu điều tra dân số, dữ liệu viễn thám được tích hợp để hiểu sự phát triển đất đai, sử dụng đất và phát triển đô thị ở Puerto Rico [11]. Cùng với các biến kinh tế xã hội, hình ảnh vệ tinh có độ phân giải cao đã được sử dụng để phân tích sự gia tăng dân số đô thị có liên quan chặt chẽ đến tăng trưởng kinh tế [13].

Từ đầu những năm 1990, dữ liệu viễn thám đã được sử dụng cho các ứng dụng nông nghiệp với sự thay đổi hiện tượng khu vực và các yếu tố khí tượng liên quan [6]. Trong nông nghiệp chính xác, vai trò của dữ liệu viễn thám ngày càng trở nên quan trọng hơn về mặt phát triển bền vững nông nghiệp, bao gồm cả an ninh lương thực [8], hoặc đánh giá cây trồng dự báo điều kiện và năng suất [7]. Ngoài ra, trung bình khoảng cách năng suất lớn giữa các quốc gia đối với các loại cây ngũ cốc chính, ngô, lúa mì và gạo, vv. Thông thường, thâm canh nông nghiệp có thể làm giảm đáng kể khoảng cách năng suất này [42]. trong này trường hợp, viễn thám đã được chứng minh là giúp ích rất nhiều cho giám sát cây trồng trong một khu vực rộng lớn, chẳng hạn như lập bản đồ tiềm năng năng lượng sinh học của cây ngô [43] bằng cách kết hợp ảnh hưởng của khí hậu và đất đai đến năng suất [44].

Đặc biệt, an ninh lương thực là yếu tố then chốt của hệ thống nông nghiệp thông minh và chỉ có viễn thám từ Vệ tinh quan sát trái đất (ví dụ: Landsat, Resourcesat, MODIS) có thể cung cấp nhất quán, lặp đi lặp lại, và dữ liệu chất lượng cao để mô tả đặc điểm và lập bản đồ đất trồng trọt chính các tham số để ước tính đất trồng trọt toàn cầu và phân tích an ninh lương thực kết hợp với số liệu thống kê quốc gia, dữ liệu fieldplot và dữ liệu thứ cấp [dài (50–100 năm) ghi lại lượng mưa và nhiệt độ, loại đất, địa giới hành chính] [31], [45]. Cùng với dữ liệu khảo sát nhân khẩu học và sức khỏe, nhiều ứng dụng có thể được hưởng lợi từ việc phân tích dữ liệu viễn thám [9], [10], và tiếp tục mối quan hệ giữa sức khỏe con người và thay đổi môi trường có thể được mô hình hóa chính xác [10].

Ngoài ra, phân tích dữ liệu viễn thám có thể được sử dụng để thị trường bảo hiểm toàn cầu, chẳng hạn như thiệt hại mùa màng và lũ lụt và đánh giá rủi ro hỏa hoạn [46]. chẳng hạn như lập bản đồ mật độ cây trên toàn cầu quy mô [16], mà còn về khoa học xã hội, chẳng hạn như nghiên cứu đô thị, nhân khẩu học, khảo cổ học, nghiên cứu chiến tranh và xung đột, v.v. [17].

V. NGHIÊN CỨU TÌNH HUỐNG

Trong phần này, hai trường hợp nghiên cứu chứng minh tính hiệu quả của dữ liệu lớn trong các ứng dụng viễn thám: mô tả. Trong cả hai trường hợp, hãy sử dụng một bộ xử lý dữ liệu mới. Các phương pháp và kiến trúc máy tính mạnh mẽ được. Cần thiết, nhu cầu. Vấn đề được giải quyết là tràn dầu tự động phát hiện và truy xuất thông tin dựa trên nội dung từ một kho lưu trữ lớn dữ liệu viễn thám đa phổ và siêu phổ và dữ liệu liên quan từ các miền khác, tương ứng.

A. Dữ liệu lớn để phát hiện sự cố tràn dầu

Trong các ứng dụng phân loại viễn thám truyền thống, các mẫu được dán nhãn được lấy trên mặt đất. khảo sát, giải thích hình ảnh, hoặc kết hợp

chiến lược nêu trên . Khảo sát mặt đất tại chỗ có thể dẫn đến độ chính xác cao của việc ghi nhãn, nhưng những kỹ thuật tốn kém và mất nhiều thời gian. Giải đoán ảnh ảnh nhanh và rẻ, nhưng không thể đảm bảo

chất lượng ghi nhãn cao. Mặc dù các giải pháp lai có thể mất

lợi thế của khảo sát mặt đất và giải đoán ảnh trong hầu hết các vấn đề viễn thám, vẫn còn khó khăn để dán nhãn sự cố tràn dầu trên biển bằng cách sử dụng giải pháp hỗn hợp trong điều khoản của dữ liệu viễn thám được cung cấp bởi không khí/vũ trụ dụng cụ do trôi dầu và khuếch tán. Vì vậy,việc dán nhãn cho sự cố tràn dầu trên biển mang đến một thách thức lớn đến nhiệm vụ phát hiện tràn dầu. Trong trường hợp nghiên cứu này, đầu tiên chúng tôi xác định dữ liệu thích hợp bao gồm viễn thám lớn

dữ liệu và sau đó giải quyết thách thức ghi nhãn bằng một dữ liệu mới

phương pháp, tức là, bằng cách tích hợp các phương tiện truyền thông xã hội dữ liệu với sự hỗ trợ của dịch vụ cộng đồng và học tập tích cực

kỹ thuật hóa.

Cụ thể, chúng tôi chọn sự kiện tràn dầu lớn xảy ra ở Vịnh Mexico (Mỹ) năm 2010 làm đối tượng nghiên cứu. Trường hợp có nhiều dữ liệu truyền thông xã hội và các dạng dữ liệu khác có thể thu được để phân tích dữ liệu lớn. điều khiển từ xa quang học dữ liệu cảm biến được sử dụng chứa đa thời gian và đa nguồn hình ảnh, tức là dữ liệu từ hình ảnh có độ phân giải trung bình máy quang phổ (MERIS), do European Space vận hành

Cơ quan (ESA) và hình ảnh có độ phân giải vừa phải quang phổ kế (MODIS), do NASA vận hành. Các chứng từ khác dữ liệu trong bối cảnh này bao gồm dữ liệu truyền thông xã hội, nghĩa là, các hình ảnh từ phương tiện truyền thông xã hội và mô tả văn bản.

Ví dụ, ảnh từ Panoramio, 9 một trang web chia sẻ ảnh định hướng theo định vị địa lý, có thể dễ dàng lấy và thường được gắn thẻ địa lý, ở dạng tọa độ chính xác của vị trí mà những hình ảnh này đã được

được chụp, cũng như gắn thẻ văn bản [xem Hình 4(a)]. Cũng thế,

dữ liệu trong không khí trong khu vực bị ô nhiễm có thể được sử dụng để dán nhãn hình ảnh viễn thám, chẳng hạn như sự cố tràn dầu được phát hiện bởi cảm biến trên không từ một viện chính thức [xem Hình 4(b)].

Những dạng dữ liệu lớn khác nhau đó có thể được sử dụng để cải thiện độ chính xác phát hiện tràn dầu trong bối cảnh cụ thể này.

Cần lưu ý rằng nó tốn nhiều thời gian cho quy trình ghi nhãn để kết hợp ý tưởng về nguồn cung ứng cộng đồng và dữ liệu bên ngoài không thể bao gồm tất cả các pixel trong ảnh viễn thám. Theo đó, điều quan trọng là lựa chọn thông minh một số lượng thông tin mẫu để ghi nhãn để đảm bảo độ chính xác của nhiệm vụ phân loại. Ở đây, quá trình dán nhãn có

được thực hiện thông qua học tập tích cực theo cách lặp đi lặp lại.

Sau khi xóa dữ liệu bị hư hỏng nặng bởi

mây, ảnh viễn thám đa bán cầu từ các niên đại khác nhau (nghĩa là ảnh đa thời gian) và ảnh từ các cảm biến khác nhau (tức là hình ảnh đa nguồn) có

được khai thác để phát hiện sự cố tràn dầu bằng thuật toán học máy. Ở đây, chúng tôi đã sử dụng các bộ phân loại phổ biến như máy vectơ hỗ trợ (SVM) [50]–[52], lan truyền ngược mạng thần kinh và phân loại. Trong các thí nghiệm của chúng tôi, SVM đã cho độ chính xác phân loại tốt nhất và cũng mạnh mẽ nhất. Bản đồ phân loại SVM thu được cho sự cố tràn dầu đang được xem xét ở Vịnh Mexico được đưa ra trong

Hình 5, cho thấy sự cố tràn dầu lan rộng xung quanh vị trí giàn khoan dầu nước sâu.

Vẫn còn nhiều vấn đề mở đối với việc ghi nhãn mẫu khi kết hợp ảnh viễn thám và dữ liệu xã hội dữ liệu truyền thông. Ví dụ, một chiến lược hiệu quả nên được phát triển để có được dữ liệu bên ngoài phù hợp nhất

cho một nhiệm vụ cụ thể. Đồng thời, những dữ liệu bên ngoài chẳng hạn như hình ảnh và thông tin văn bản được tự động liên kết với các mẫu tương ứng.

B. Truy xuất hình ảnh dựa trên nội dung từ kho dữ liệu siêu phổ

Trong nghiên cứu trường hợp thứ hai này, chúng tôi giải quyết một trường hợp cụ thể nghiên cứu về truy xuất ảnh dựa trên nội dung (CBIR) áp dụng cho dữ liệu siêu phổ được cảm nhận từ xa, được đặc trưng bởi độ cao của nó trong miền phổ. Hệ thống, được giới thiệu trong [1], được xác thực bằng cách sử dụng một cơ sở dữ liệu hình ảnh siêu âm phức tạp và triển khai trên cụm Beowulf tại Chuyến bay Không gian Goddard của Trung tâm NASA. Trong bối cảnh này, thách thức chính của

trường hợp này khi nghiên cứu của chúng tôi đang giải quyết thách thức lớn về dữ liệu viễn thám lớn, trong thử nghiệm của chúng tôi bao gồm một bộ sưu tập 154 bộ dữ liệu siêu âm độ phân giải cao (hơn 20 TB dữ liệu) do NASA thu thập qua khu vực Trung tâm Thương mại Thế giới (WTC) ở Thành phố New York trong hai tuần cuối tháng 9 năm 2001, chỉ có một số ít vài ngày sau vụ tấn công khủng bố đã hạ gục cả hai tòa tháp chính và các tòa nhà khác trong khu phức hợp WTC. Độ phân giải không gian của dữ liệu là 3,7 m/pixel, và độ phân giải phổ là 224 dải phổ hẹp giữa 0,4 và 2,5m. Hình 6 cho thấy một hỗn hợp màu sai của một trong những hình ảnh như vậy, với 614 512 pixel và 224 ban nhạc. Bố cục màu sai đã được hình thành sử dụng các kênh 1682-, 1107- và 655-nm, được hiển thị

lần lượt là đỏ, lục và lam. Các khu vực thực vật có màu xanh lê trong Hình 6, trong khi các khu vực bị đốt cháy có màu tối

xám. Khói bốc ra từ khu WTC hiện ra sáng rực màu xanh lam do độ phản xạ quang phổ cao ở bước sóng 655-nm. Khu vực được sử dụng làm truy vấn đầu vào trong thử nghiệm của chúng tôi được hiển thị trong một hình chữ nhật màu đỏ và được căn giữa tại khu vực nơi các tòa tháp sụp đổ.

Sử dụng khu vực tìm kiếm ở phần ngoài cùng bên phải của Hình 6

như truy vấn đầu vào, hệ thống CBIR song song được đề xuất đã truy xuất thành công tất cả các trường hợp hình ảnh có chứa WTC phức tạp trên cơ sở dữ liệu, không có tính giả phát hiện. Đối với mục đích minh họa, Hình 7 cho thấy bảy đường bay được chụp ảnh đầy đủ trong cơ sở dữ liệu AVIRIS đã được xem xét có chứa khu vực tìm kiếm tập trung tại

khu phức hợp WTC.

Để điều tra các thuộc tính song song của đề xuất hệ thống CBIR, chúng tôi đã đánh giá hiệu suất của nó khi được triển khai trên cụm Thunderhead Beowulf của

NASA, một hệ thống bao gồm 256 nút Intel Xeon 2,4 GHz kép, mỗi bộ nhớ 1 GB và bộ nhớ chính 80 GB,

được kết nối với Myrinet sợi quang 2 GHz. sử dụng 256 bộ xử lý trên

Thunderhead, hệ thống đã có thể tìm kiếm những cảnh tương tự nhất trên cơ sở dữ liệu đầy đủ của 154 hình ảnh (với siêu dữ liệu được tính toán trước) chỉ trong 4 giây, dẫn đến tổng thời gian xử lý khoảng 10 giây để ghi danh mục và mô tả đầy đủ một mục mới trong cơ sở dữ liệu. Nơi đây

đại diện cho một cải tiến đáng kể so với việc thực hiện cùng một quy trình CBIR trên một Thunderhead duy nhất bộ xử lý, mất hơn 1 giờ tính toán cho thao tác tương tự.

VI. PHÂN KẾT LUẬN

Trong bài báo này, ý nghĩa của dữ liệu viễn thám lớn đã được thảo luận. Dữ liệu lớn trong viễn thám có thể chứa nhiều loại dữ liệu viễn thám từ các nguồn khác nhau: phản xạ quang phổ, độ phân giải không gian mặt đất khác nhau, và các địa điểm khác nhau (chẳng hạn như quang học, radar, vi sóng, v.v.), cũng như dữ liệu từ các lĩnh vực khác, chẳng hạn như khảo cổ học, nhân khẩu học, kinh tế học (trong đó đề cập đến “sự đa dạng” của ba thuộc tính V của dữ liệu lớn). Như một kết quả là dữ liệu viễn thám lớn có cùng ba V có đặc điểm là dữ liệu lớn nói chung [1] với khối lượng dữ liệu viễn thám được tích lũy ngày càng nhiều từ TB đến PB và thậm chí đến thang EB. Với khối lượng dữ liệu khổng lồ, một mặt, các nhiệm vụ khó bị tấn công có thể đạt được trong thời gian hợp lý (nghĩa là “tốc độ” của ba chữ V và dẫn đến các cơ hội lớn); mặt khác, dữ liệu viễn thám lớn với bất kỳ 2V hoặc 3V nào đều mang lại những thách thức lớn cho những người sở hữu dữ liệu lớn, phân tích dữ liệu lớn và sử dụng dữ liệu lớn, tương ứng.

Sau đó, một khuôn khổ bộ ba để hiểu dữ liệu lớn trong viễn thám đã được đề xuất cho những người sở hữu dữ liệu lớn, những người có thể cung cấp phương pháp dữ liệu và những người cần khai thác dữ liệu lớn để giải quyết các vấn đề trong thế giới thực. Về khuôn khổ, những thách thức chung và riêng lẻ của dữ liệu lớn đã được thảo luận trong bối cảnh các ứng dụng viễn thám. Là một thách thức chung quan trọng, dữ liệu viễn thám lớn trước tiên phải được xác định để đối phó với các ứng dụng viễn thám thực tế. Sau đó, cần phải có khả năng tính toán hiệu quả cao để xử

lý dữ liệu khổng lồ. Ngoài ra, các phương pháp dữ liệu mới hoặc thậm chí các phương pháp dữ liệu hoàn toàn mới nên được phát triển để tấn công vào sự phức tạp của dữ liệu viễn thám lớn. Tất nhiên, vẫn tồn tại những thách thức chung khác khi xử lý dữ liệu viễn thám lớn, chẳng hạn như cách thao túng chất lượng dữ liệu từ các quan điểm khác nhau của từng nhà cung cấp dữ liệu và người nhận dữ liệu.

Ngoại trừ những thách thức chung của dữ liệu lớn trong viễn thám, những thách thức riêng lẻ cần được tính đến từ ba khía cạnh của bộ ba dữ liệu lớn. Đối với những người sở hữu dữ liệu viễn thám lớn, ba yếu tố quan trọng cần được thiết kế cẩn thận, đó là truyền dữ liệu từ hệ thống cảm biến trên không/vệ tinh đến trạm mặt đất, sau đó lưu trữ dữ liệu vào hệ thống và phân phối dữ liệu đến người dùng quan tâm. Đối với những người khai thác dữ liệu lớn cho các ứng dụng viễn thám, những thách thức chính là xác định đúng dữ liệu để đạt được nhiệm vụ nhất định, triển khai dữ liệu lớn để xử lý và phân tích dữ liệu sau này và giải thích kết quả do các phương pháp dữ liệu cung cấp. Đối với những người có khả năng phát triển các phương pháp dữ liệu mới và/hoặc phương pháp luận dữ liệu cho các ứng dụng viễn thám, trước tiên cần quản lý việc biểu diễn dữ liệu do tính đa dạng của dữ liệu viễn thám đa nguồn và đa thời gian cũng như dữ liệu từ các miền khác. Sau đó, dữ liệu được mô tả ở các định dạng thuộc tính khác nhau sẽ được tích hợp để phân tích và xử lý dữ liệu lớn tốt hơn. Sau đó, các kết quả được cung cấp bởi các kỹ thuật phân tích dữ liệu cần được trực quan hóa tốt để phân tích dữ liệu được cải thiện và

diễn giải dữ liệu.

Mặc dù tiềm năng của dữ liệu viễn thám lớn đã được dự đoán trước, điều quan trọng cần lưu ý là dữ liệu thường đến từ các nguồn không đồng nhất và yêu cầu những nỗ lực tính toán đáng kể về mặt diễn giải. Do đó, cơ hội lớn là tích hợp dữ liệu viễn thám cùng với các dữ liệu bên ngoài khác để biến những tiềm năng này thành hiện thực. Trong bối cảnh này, chúng ta có thể hưởng lợi từ quy mô lớn, nhất quán, lặp đi lặp lại và dữ liệu viễn thám lớn chất lượng cao để giải quyết các ứng dụng liên quan đến giám sát an ninh lương thực, tiến độ đô thị hóa, mật độ dân số, v.v. Điều này có thể được sử dụng thêm để giải quyết các ứng dụng phù hợp khác liên quan đến sức khỏe con người, thay đổi môi trường hoặc hoạt động của con người nói chung.

Để hưởng lợi từ viễn thám trong dữ liệu lớn, dữ liệu thích hợp từ các nguồn khác nhau trước tiên phải được xác định để giải quyết một ứng dụng cụ thể. Ngoài trừ dữ liệu viễn thám đa thời gian, đa độ phân giải, đa bức xạ, cách xác định dữ liệu ngoài miền bổ sung có liên quan và cách lấy các bộ dữ liệu đó là thách thức lớn nhất đối

với ứng dụng dữ liệu viễn thám lớn. Sau đó, một phương pháp dữ liệu mới nên được thiết kế cẩn thận để xử lý dữ liệu, hợp nhất dữ liệu, v.v. Mặc dù có nhiều ứng dụng kết hợp dữ liệu viễn thám và dữ liệu đến từ các lĩnh vực khác, nhưng hầu hết các công trình có sẵn đều dựa trên kỹ thuật lấy mẫu để ước tính, ngay cả đối với công trình gần đây ước tính dân số cây trên toàn cầu dựa trên 429775 phép đo có nguồn gốc từ mặt đất mật độ cây từ mọi lục địa trên Trái đất[16]. Làm thế nào chúng ta có thể sử dụng tất cả các dữ liệu có sẵn xứng đáng được nghiên cứu thêm cho nhiệm vụ dữ liệu viễn thám lớn. Cuối cùng nhưng không kém phần quan trọng, làm thế nào để đánh giá hiệu suất và

làm thế nào để đảm bảo chất lượng dữ liệu là những dòng nghiên cứu thú vị khác sẽ được khám phá thêm.