**Lê Ngọc Thiên Phúc – 22521117**

**BÀI THỰC HÀNH 4: THU THẬP VÀ TIỀN XỬ LÝ DỮ LIỆU TỔNG HỢP.**

Bộ dữ liệu: **Human Activity Recognition Using Smartphones Data Set**

Link dữ liệu: [https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI HAR Dataset.zip](https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip)

Đọc file README.md và cho biết các thông tin sau:

1. Chức năng (nội dung) của bộ dữ liệu là gì?

Bộ dữ liệu cung cấp thông tin thu thấp từ 30 người từ 19-48 tuổi, thực hiện 6 hành động(đi, đi lên cầu thang, đi xuống cầu thang, ngồi, đứng, nằm) đeo điện thoại thông minh (Samsung Galaxy S II) ở thắt lưng. Bằng cách sử dụng gia tốc kế và con quay hồi chuyển tích hợp, thu được gia tốc tuyến tính 3 trục và vận tốc góc 3 trục ở tốc độ không đổi 50Hz. Các thí nghiệm đã được quay video để dán nhãn dữ liệu theo cách thủ công. Tập dữ liệu thu được được phân chia ngẫu nhiên thành hai tập, trong đó 70% tình nguyện viên được chọn để tạo dữ liệu huấn luyện và 30% dữ liệu thử nghiệm.

Các tín hiệu cảm biến (gia tốc kế và con quay hồi chuyển) đã được xử lý trước bằng cách áp dụng các bộ lọc nhiễu và sau đó được lấy mẫu trong các cửa sổ trượt có chiều rộng cố định là 2,56 giây và chồng chéo 50% (128 lần đọc/cửa sổ). Tín hiệu gia tốc của cảm biến, bao gồm các thành phần trọng lực và chuyển động của cơ thể, được tách bằng bộ lọc thông thấp Butterworth thành gia tốc cơ thể và trọng lực. Lực hấp dẫn được giả định chỉ có các thành phần tần số thấp, do đó bộ lọc có tần số cắt 0,3 Hz đã được sử dụng. Từ mỗi cửa sổ, một vectơ đặc trưng thu được bằng cách tính toán các biến từ miền thời gian và tần số. Xem 'features\_info.txt' để biết thêm chi tiết.

1. Các tác giả của bộ dữ liệu?

Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto.

1. Cấu trúc các tập tin của bộ dữ liệu? Ý nghĩa từng tập tin.

Bộ dữ liệu gồm các tâp tin:

- 'README.txt' : thông tin tổng quan về bộ dữ liệu

- 'features\_info.txt': thông tin về các thuộc tính

- 'features.txt': Danh sách các thuộc tính

- 'activity\_labels.txt': Links the class labels with their activity name.

- 'train/X\_train.txt': tập huấn luyện

- 'train/y\_train.txt': nhãn huấn luyện

- 'test/X\_test.txt': tập thử nghiệm

- 'test/y\_test.txt': nhãn thử nghiệm

Những files dưới đây cho dữ liệu train và test

- 'train/subject\_train.txt': Mỗi hàng xác định chủ thể đã thực hiện hoạt động cho từng mẫu cửa sổ. Phạm vi của nó là từ 1 đến 30.

- 'train/Inertial Signals/total\_acc\_x\_train.txt' Tín hiệu gia tốc từ trục X gia tốc kế của điện thoại thông minh theo đơn vị trọng lực tiêu chuẩn 'g'. Mỗi hàng hiển thị một vectơ 128 phần tử. Mô tả tương tự cũng áp dụng cho tệp 'total\_acc\_x\_train.txt' và 'total\_acc\_z\_train.txt' cho trục Y và Z.

- 'train/Inertial Signals/body\_acc\_x\_train.txt': Tín hiệu gia tốc cơ thể thu được bằng cách trừ đi trọng lực khỏi tổng gia tốc.

- 'train/Inertial Signals/body\_gyro\_x\_train.txt': Vectơ vận tốc góc được đo bằng con quay hồi chuyển cho từng mẫu cửa sổ. Đơn vị là radian/giây.

1. Các thông tin nào mà các record lưu trữ?

- Gia tốc ba trục từ gia tốc kế (gia tốc toàn phần) và gia tốc cơ thể ước tính.

- Vận tốc góc ba trục từ con quay hồi chuyển.

- Một vectơ đặc trưng 561 với các biến miền thời gian và tần số.

- Nhãn hoạt động của nó.

- Mã định danh đối tượng thực hiện thí nghiệm.