TRƯỜNG ĐẠI HỌC PHENIKAA

KHOA CÔNG NGHỆ THÔNG TIN

BỘ MÔN KĨ THUẬT PHẦN MỀM

Logo, company name

Description automatically generated--❧•❧--

**BÀI TẬP GIỮA KÌ**

**ĐỀ TÀI:**

7.2.4 Use of Big Data for Software Analytics

7.2.5 Neural Network Approach to Bug Prediction and Cost Estimation

Giáo viên phụ trách: Vũ Quang Dũng

Học Phần: Kỹ thuật phần mềm\_1.2(14IT).4\_LT

Học Sinh: Phạm Tiến Thành Công

MSSV: 20010886

**7.2.4. Sử dụng Big Data cho phân tích phần mềm**

Trước hết để biết được ứng dụng của big data với phân tích phần mềm, ta cũng nên biết Phân tích Dữ liệu lớn (Big Data Analytics) là gì ?

Phân tích Dữ liệu lớn là một quy trình được sử dụng để trích xuất những thông tin chi tiết có ý nghĩa, chẳng hạn như các mẫu ẩn, mối tương quan chưa biết, xu hướng thị trường và sở thích của khách hàng. Phân tích Dữ liệu lớn cung cấp nhiều lợi ích khác nhau - nó có thể được sử dụng để đưa ra quyết định tốt hơn, ngăn chặn các hoạt động gian lận, trong số những thứ khác.

Quay lại với vấn đề sử dụng Big Data cho phân tích phần mềm. Cùng với sự phát triển của công nghệ, mạng máy tính thì lượng thông tin trên thế giới cũng sẽ gia tăng. Tăng khoảng 10 lần sau mỗi 5 năm. Thế giới cũng đã chứng kiến một sự cải tiến tương tự trong lưu trữ, tính toán và khả năng xử lí của các hệ thống máy tính. Cuộc cách mạng CNTT này đã dẫn tới sự phát triển thần kì của nhiều phương pháp dự đoán và phân tích dữ liệu.

Maryville đã đưa ra một mô tả chi tiết về những ứng dụng của big data trong lĩnh vực phần mềm. Với lượng dữ liệu thu thập từ những báo cáo về bug (lỗi kĩ thuật) và thống kê từ các bản vá phần mềm trước đó, chúng cho thấy tồn tại rất nhiều khả năng cho các ứng dụng từ big data thông qua SDLC (Chu kỳ phát triển phần mềm).

Các công ty như Netflix và Procter & Gamble sử dụng dữ liệu lớn để dự đoán nhu cầu của khách hàng. Họ xây dựng các mô hình dự đoán cho các sản phẩm và dịch vụ mới bằng cách phân loại các thuộc tính chính của các sản phẩm hoặc dịch vụ trong quá khứ và hiện tại và mô hình hóa mối quan hệ giữa các thuộc tính đó và thành công thương mại của các dịch vụ.

Dữ liệu lớn cho phép bạn thu thập dữ liệu từ phương tiện truyền thông xã hội, truy cập web, nhật ký cuộc gọi và các nguồn khác qua đó ta sẽ biết được trải nghiệm của khách hàng với phần mềm của mình như nào. Từ đó mà sẽ biết khách hàng cần gì, muốn gì mà nâng cao cải tiến chất lượng phần mềm hơn.

Ứng dụng của big data trong đời sống ngày nay rất nhiều không chỉ ở riêng lĩnh vực phân tích phần mềm, là thứ không thể thiếu. Tuy vậy vẫn có những mặt tiêu cực trong việc sử dụng big data cho phân tích phần mềm. Đó chính là vấn đề riêng tư của người sử dụng. Thông tin có khả năng định dạng người sử dụng có thể thu thập khi mà thu thập Big Data và điều này hoàn toàn không được sự cho phép của họ. Do đó mà đó cũng là một trong những thách thức của việc ứng dụng big data vào phân tích phần mềm cũng nhưng lĩnh vực khác. Dù vậy lợi ích mà nó mang lại là không thể bàn cãi và chúng ta sẽ phải học cách sử dụng chúng sao cho hợp lý nhất.

* + 1. **Phương pháp tiếp cận mạng neural để dự đoán lỗi và chi phí ước lượng**

Neural Network (Mạng neural) là một chuỗi những thuật toán được đưa ra để tìm kiếm các mối quan hệ cơ bản trong tập hợp các dữ liệu. Thông qua việc bắt bước cách thức hoạt động từ não bộ con người. Nói cách khác, mạng nơ ron nhân tạo được xem là hệ thống của các tế bào thần kinh nhân tạo.

Neural Network có khả năng thích ứng được với mọi thay đổi từ đầu vào. Do vậy, nó có thể đưa ra được mọi kết quả một cách tốt nhất có thể mà bạn không cần phải thiết kế lại những tiêu chí đầu ra. Khái niệm này có nguồn gốc từ [trí tuệ nhân tạo](https://itnavi.com.vn/blog/lap-trinh-ai/), đang nhanh chóng trở nên phổ biến hơn trong sự phát triển của những hệ thống giao dịch điện tử.

Để dự đoán lỗi có rất nhiều phương pháp và một trong số đó là sử dụng mạng neural. Một mạng nơron nhân tạo được cấu trúc cho một ứng dụng cụ thể (nhận dạng mẫu, phân loại dữ liệu,...) thông qua một quá trình huấn luyện-học (training) từ tập các mẫu huấn luyện. Dữ liệu đầu vào (input) sẽ chạy qua toàn bộ mạng nơ ron, sẽ được xử lý, tìm các mối liên hệ và tái tạo lại thành kết quả đầu ra (output). Các output này sẽ được so sánh với với các dữ liệu mục tiêu (target) mà hệ thống đã được học trước đó. Nếu có sự sai lệch giữa output và target thì nó sẽ xuất ra là lỗi. Đó là tổng quan về cách dự đoán lỗi của mạng neural.

Hiện nay khi lập báo cáo đầu tư cho các dự án dân dụng, chúng ta thường dùng phương pháp suất vốn đầu tư hoặc giá thành 1m2 sàn. Các phương pháp trên đều rất kém chính xác, đặc biệt trong điều kiện Việt Nam, khi mà cơ sở dữ liệu về chi phí còn rất sơ sài, không đồng nhất, thiếu cập nhật. Vướng mắc trên là một trong các nguyên nhân gây ra tình trạng vượt chi phí khi thực hiện dự án. Và mạng neural cho phép ước lượng chi phí của các dự án một cách dễ dàng trên cơ sở dữ liệu về chi phí của những dự án trước đây. Đầu tiên ta sẽ thu thập dữ liệu để chọn các biến đầu vào (input) và các xuất lượng (output). Sau đó sẽ thiết lập mô hình thuật toán và sử dụng các phần mềm phức tạp để huấn luyện mạng neural nhằm có bộ ma trận trọng số phục vụ cho bài toán ước lượng. Và sau đó ta sẽ ước lượng được chi phí của dự án.

Ứng dụng của mạng neural trong cuộc sống là rất nhiều đơn cử là 2 vấn đề trên là dự đoán lỗi và ước lượng chi phí. Việc ứng dụng nó sẽ giảm bớt sức lao động cho người làm cũng như làm việc với hiệu suất cao, chính xác hơn con người rất nhiều.