Word2Vec Tutorial Part 2 - Negative Sampling

trong phần 2 của word2vec tutorial, tôi sẽ giới thiệu thêm một số thay đổi của mô hình skip-gram cơ bản điều rất quan trọng để train chính xác

khi bạn đọc hướng dẫn này trên skip-gram model cho Word2Vec, bạn có thể nhận ra vài điều nó là một mạng nơ-ron lớn.

trong ví dụ tôi gửi, chúng ta có 300 components và một từ vựng gồm 10,000 từ. Nhớ lại ràng neural network có hai ma trận weight – một lớp ẩn và một lớp đầu ra. Cả hai lớp sẽ có ma trận trọng luông với 300x10,000 = 3 triệu mỗi cái!

Chạy gradient descent trên neural network lớn có thể sẽ chậm. và tình hình còn tệ hơn nữa khi bạn cần một lướng lớn dữ liệu để huấn luyện để điều chỉnh và tránh quá tải. hàng triệu cân trọng lượng hàng tỉ mẫu huấn luyện có nghĩa là đào tạo mô hình này sẽ là một con quái vật.

Các tác giả của Word2Vec giải quyết các vấn đề này trong bài báo thứ hai của họ:

1. Giải quyết các cặp từ thông dụng như một từ đơn trong mô hình của họ
2. Lấy mẫu các từ thông dụng để giảm thiếu số lượng mẫu huấn luyện
3. Sử đội mục tiêu tối ưu hóa với một công nghệ gọi là “Negative Sampling”, cái là nguyên nhân của mỗi tập huấn luyện để cập nhất chỉ một phần nhỏ củ model’s weight.

Cần lưu ý ràng “Subsampling frequent words” và áp dụng “Negative Sampling” không chỉ giảm gánh nặng tính toán của quá trình đào tạo mà còn cải thiện chất lượng kết quả huấn luyện word vectors.