Name: Suhaila Ahmed Hassan
Track: AI
Branch: Alexandria

Efficient Estimation of Word Representations in Vector Space

Introduction

At the time of the paper's publication:
Most NLP techniques treated words as independent units with no similarities between them.
Reasons for this choice include simplicity, robustness and the observation that simple models trained on big data outperform complex models trained on small data.
However, this approach doesn't always produce good results.

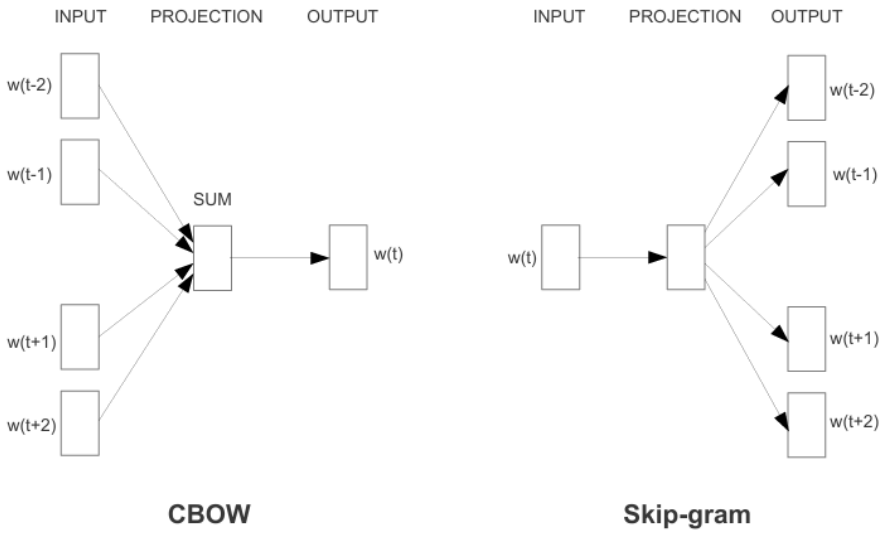
Goals of the Paper

Introduce techniques for learning high-quality word vectors from big data that captures multiple degrees of similarities (syntactic, semantic).

Model Architecture

Two new model architectures are proposed that forgo the complexity caused by the non-linear hidden layers in a traditional Neural network.

1. Continuous Bag-of-Words (CBOW): Predicts a target word based on its surrounding context words. Uses bag-of-words method and ignores word order.
2. Skip-Gram Model: Predicts surrounding context words given a target word..



Model Training

Word vectors are trained using Google News corpus. Corpus has about 6B tokens, but vocabulary size is restricted to 1 million most frequent words.
Increasing vector dimensionality and/or number of training words increases accuracy.
Goal is to maximize accuracy while minimizing computational complexity as much as possible.

Results

- CBOW and Skip-Gram outperform previous models on syntactic and semantic similarities.
- The CBOW model is more effective at capturing syntactic meaning.
- The Skip-Gram model is more effective at capturing semantic meaning.
- The Skip-Gram model has a better total performance on semantic and syntactic meanings.