Summary of "Efficient Estimation of Word Representations in Vector Space"

Main Goal

The paper introduces two simple but powerful models to learn word vectors (word embeddings) from text:
- CBOW (Continuous Bag of Words)
- Skip-Gram
These models became famous under the name Word2Vec.

What Did They Do?

1. Created two models (CBOW & Skip-Gram):
   - CBOW: Predicts a word using the context around it.
   - Skip-Gram: Predicts the context using the word itself.

2. Trained these models on very large datasets (up to 1 trillion words) using Google's DistBelief distributed system.
3. Compared performance with older models like neural networks (RNNs and Feedforward), and found their models:
   - Are faster
   - Use less resources
   - Still achieve very high accuracy

Key Results

- Speed: Much faster than older neural models
- Simplicity: Very simple architecture
- Scalability: Can be trained on huge datasets (1 trillion words)
- Accuracy: Performs very well on semantic/syntactic tasks

Applications They Tested

- Sentiment Analysis
- Paraphrase Detection
- Machine Translation
- Knowledge Base Expansion
- Semantic Evaluation (like SemEval 2012 Task 2)

Conclusion

High-quality word vectors can be learned using simple architecture, and they can become essential building blocks for many future NLP tasks.