Tempo-Lexical Context driven Word Embedding for Cross-Session Search Task Extraction

Procheta Sen

PhD. Student,ADAPT, Dublin City University

March 26, 2018

Introduction

- "Tempo-Lexical Context driven Word Embedding for Cross-Session Search Task Extraction" - Accepted in NAACL 2018

Procheta Sen, DCU

Debasis Ganguly, IBM

Gareth Jones, DCU

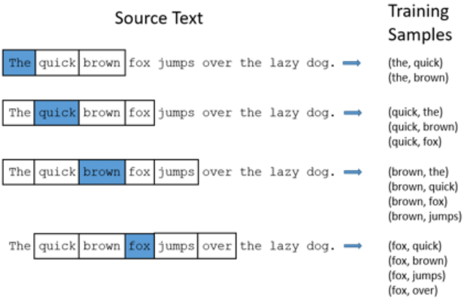
Overview

Outline

Word Embedding

- Words represented as vectors.
- Vectors are semantic representations of words.
- Allows seamless integration between different modalities.

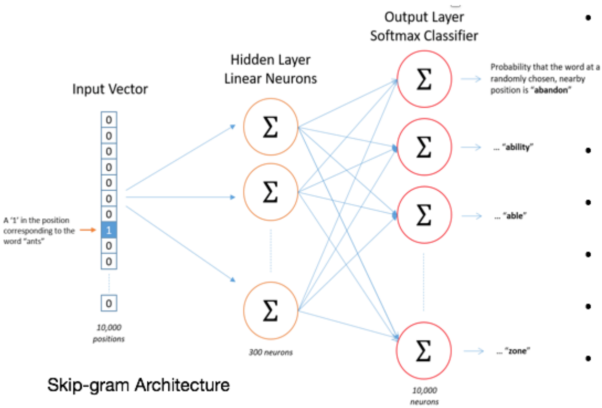
Word2Vec Objective Function



1. Unsupervised method.
2. Slide a moving window through *unlabeled* text.
3. Aim to make the current word *similar* to its context and *dissimilar* to other words outside this context.
4. Objective function maximized with SGD.

$$J(W) = \sum_{w_t, c_t \in D^+} \sum_{c \in c_t} p(D = 1 | \vec{w_t}, \vec{c}) - \sum_{w_t, c'_t \in D^-} \sum_{c \in c'_t} p(D = 1 | \vec{w_t}, \vec{c})$$

Word2Vec RNN Architecture



Skip-gram Architecture

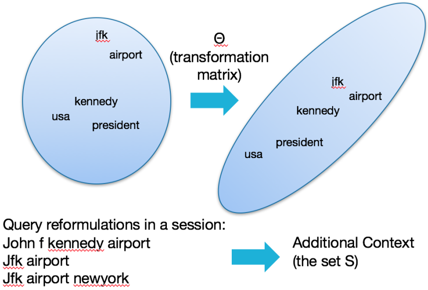
- Two models:
  - CBOW: Predict current word given context.
  - Skip-gram: Predict context given current word.
- Current word vector – One hot vector.
- Context: Concatenated one hot vectors.
- Latent layer – Dimension of word vectors.
- Output: Softmax layer of dimension V (vocab size).
- Output of the latent layer is used as embedded vectors.

Traditional Word Embedding Issues for Short Documents

- The word embedding algorithm respects document boundaries by not extending the context window across them.
- Short documents (e.g. query, tweets) often comprise of 2-3 words.
- Short contexts may result in an improper embedding.

Outline

Proposed Solution: Transforming Word Vectors



- $\vec{w'} = \theta \cdot \vec{w}$
- Transformation matrix changes the neighborhood of each word.
- Brings (externally specified) context (S) specific words closer in the transformed space.

Learning Transformation Matrix (θ)

- Given a set S of *contextually similar* word vector pairs,

$$\Phi(w) = \{v : (w, v) \in S\}$$

- Context can be defined based on the use case.
- Hinge loss function to learn θ

$$l(\vec{w}; \theta) = \sum_{\vec{v}: v \in \Phi(w)} \sum_{\vec{u}: u \notin \Phi(w)} \max\left(0, m - ((\theta \vec{w})^T \vec{v} - (\theta \vec{w})^T \vec{u})\right)$$

- m is the margin

Illustration With Use Case: Query Log

- Query corresponds to short documents.
- **Terminologies:**
  - A **session** is defined as a set of queries where the time gap between any two consecutive queries is no more than a threshold.
  - A **task** is a set of queries belonging to same search goal.

Defining Context ($\phi(W)$)

- **Temporal Semantic Context** Context comprises words of all the queries appearing within a session.
- **Tempo-Lexical Semantic Context** Context comprises words of lexically similar queries within a session.

Illustrating the Effectiveness of Transformed Word Vectors

- Queries will be presented as vectors using transformed word vectors.
- Query vectors are clustered to identify task related queries.

Dataset Details

Table: Dataset statistics of task annotated queries from AOL query log. Cross-session task labels are post-processed annotations of the dataset prepared by Lucchese et al.

Item Cross-session	Task label granularity
#Queries	1424
#Tasks annotated	224
#Sessions	307
#Sessions with cross-session tasks	239
#Query pairs across sessions judged in the same task	36768

Experiments and Results

Query Similarity	Parameters		Metrics		
	α	η	F-score	Prec	Recall
Qry vec skip-gram	0.7	0.8	0.524*	**0.465***	0.602
Qry vec with temporal context	1.0	0.7	0.536*†	0.461*	0.643*†
Qry vec with tempo-lexical context	0.6	0.7	**0.538***†	0.441*	**0.691***†‡

1. Learning transformation of word vectors on a query log further improves clustering effectiveness.

Questions?