A Subspace distance for Semantic Spaces

Guido Zuccon

11th November 2008

guido@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/~guido

Information Retrieval Group
University of Glasgow
A Subspace distance for Semantic Spaces

- The initial part of this investigation has been conducted in collaboration with Dawei and sponsored by grant numbers EP/E002145/1 and EP/F014708/1.
- Now this work is carried with the collaboration of Leif and Keith.
How to derive a Semantic Spaces representation

- By HAL: sliding window collects co-occurrences of words over the corpus
- By Co-occurrence matrix: multiply a term-document matrix for its transpose
- By dimensionality reduction: using LSI or RI, associations between words are underlined

We focus on the first methodology (HAL): the intuition is that co-occurrences of words into a window depict semantical relationships providing a key for the interpretation of words meaning.
Semantic Spaces (by means of HAL):

- used as representation model of semantic memory [Burgess et al., 1998]
- are compatible with human reasoning [Landauer et al., 1998]
- have been used for modeling information inference for query expansion [Song and Bruza, 2001].
Motivation

Why comparing Semantic Subspaces (SSS)? Why a distance between them?

► comparing the SSS of documents would provide a measurement of the semantic similarity between documents (and the similarity between the concepts expressed in the documents)

► if the cluster hypothesis holds for the SSS distance, then documents related to the same concepts will be close to each other w.r.t. the SSS distance;

► if a SSS can be coupled with an information need and a user’s context (i.e. knowledge and context of user are known), then it’s possible to rank documents w.r.t. SSS distance from the "relevance SSS".

A distance between word representations in SSS can be calculated for example by means of Minkowski distance. However, no global distance has been investigated to compare SSS
Comparing Semantic Spaces – Local

How do we compare two Semantic Spaces?

- Local comparison: consider how each word is represented, e.g.:

1. a word $w$ is represented by
   \[ w_1 = [0.1678, 0.4196, 0, 0.5874, 0.6713] \] in $H_1$ and by
   \[ w_2 = [0, 0.8660, 0.2887, 0.2887, 0.2887] \] in $H_2$

2. we compare the two word representations with a local distance, for example Euclidean distance:
   \[
   \text{dist}(w_1, w_2) = \sqrt{\sum_{i=1}^{n} (w_1(i) - w_2(i))^2} = 0.7393
   \]

3. this procedure is applied to each pair of the same word in the two spaces: we are comparing the word representations

- Global comparison: consider the geometry of the subspaces
How do we compare two Semantic Spaces?

- Local comparison: consider how each word is represented
- Global comparison: consider the geometry of the subspaces
  1. we don’t compare the word representation word-by-word
  2. we compare an entire subspace representation against another
  3. takes into account the whole geometry of the subspace built by a set of documents, not the geometry of a word
  4. we take into account the concepts expressed in a set of documents by getting the basis of the SSS and then measure the distance between two SSS, representing the distance between the concepts/contexts expressed in two sets of documents.
Desired properties of a SSS distance:

- invariant to the choice of the basis
- symmetric (since the chosen SSS representation – HAL – is symmetric)
- non-negative and \( \text{distance}(H_{d_1}, H_{d_2}) = 0 \iff H_{d_1} = H_{d_2} \)
- upper boundedness
- triangularity inequality (thus is a metric)
- take into account difference in the subspaces dimensions
A possible candidate is the Subspace distance:

\[
\text{distance}(H_{d_1}, H_{d_2}) = \sqrt{\max(m, n) - \sum_{i=1}^{n} \sum_{j=1}^{m} (u_i^T v_j)^2}
\]

- \( \text{rank}(H_{d_1}) = m \) and \( \text{rank}(H_{d_2}) = n \)
- \( u_i \), i-th vector of the orthonormal basis for \( H_{d_1} \)
- \( v_j \), j-th vector of the orthonormal basis for \( H_{d_2} \)
A possible candidate is the Subspace distance:

$$\text{distance}(H_{d_1}, H_{d_2}) = \sqrt{\max(m, n) - tr(Q_2^T Q_1 Q_1^T Q_2)}$$

where:

- $\text{rank}(H_{d_1}) = m$ and $\text{rank}(H_{d_2}) = n$
- $Q_1$ orthonormal (orthogonal and normalized) basis for $H_{d_1}$
- $Q_2$ orthonormal basis for $H_{d_2}$
The Subspace distance

- The distance has been originally proposed in its inner product form by [Wang et al., 2006]
- it has been proposed and used in the area of face recognition
- has a strong relationship with the chordal distance
  \[ d_c(S_a, S_b) = \sqrt{m - \text{tr}(P_aP_b)} \]
- chordal distance has an important role in the detection of MUB, i.e. bases which spans planes totally orthogonal between them [Bengtsson et al., 2007]
The subspace distance – properties

- the distance is invariant to the choice of the orthonormal basis (hint: use Parseval’s theorem for demonstration)
- it is symmetric
- non-negative \( \text{distance}(H_{d1}, H_{d2}) \geq 0 \) and \( \text{distance}(H_{d1}, H_{d2}) = 0 \iff H_{d1} = H_{d2} \)
- upper boundedness: \( \text{distance}(H_{d1}, H_{d2}) \leq \sqrt{\max(m, n)} \), and \( \text{distance}(H_{d1}, H_{d2}) = \sqrt{\max(m, n)} \iff H_{d1} \perp H_{d2} \)
- the triangularity inequality holds
The proposed SSS distance assumes orthonormal basis. What does it happen if we relax such condition?

- concepts are not exclusive (orthonormal) between them (this is often the case in the real world)
- the SSS distance as it has been defined is invariant to the choice of the orthonormal basis
- relaxing the orthonormality condition makes losing the invariant property
- thus the distance depends by the choice of the basis, once the orthonormality is not required
- a uniform way of choosing the basis is then needed
- NMF could be a solution: each vector of the factorized basis can be associated to a sense
Open questions

- does the SSS distance cluster documents according to the cluster hypothesis?
- how to choice a basis if the orthonormality condition is relaxed?
- is a ranking w.r.t. the SSS distance effective in addressing an information need?


