Improving Bilingual Projections via Sparse Covariance Matrices

Jagadeesh Jagarlamudi, Raghavendra Udupa, Hal Daumé III, Abhijit Bhole

Introduction

- Interlingual space can help bridge the language barrier.
- Covariance matrices are used to find the mappings (A & B)

E.g. Canonical Correlation Analysis

\[
\begin{bmatrix}
0 & C_{xy} \\
C_{yx} & 0
\end{bmatrix}
\begin{bmatrix}
a \\
b
\end{bmatrix}
= (1-\lambda)\begin{bmatrix}
C_{xx} & 0 \\
0 & C_{yy}
\end{bmatrix}
\begin{bmatrix}
a \\
b
\end{bmatrix}
\]

where \(C_{xx} = XX^T\), \(C_{yy} = YY^T\) and \(C_{xy} = XY^T\)

- But, covariance matrices tend to be very dense
  - Due to the noisy co-occurrence pairs

Approach

- Recover the sparseness of the co-variance matrices and thus learn better mappings (or projection directions A & B)

Main Idea

- Let \(I_{xy}\) be a binary matrix representing the desired sparsity matrix, then:
  - \(C_{xy} \leftarrow C_{xy} \odot I_{xy}\) (Element-wise matrix multiplication)
  - i.e. \(I_{xy}\) is used for feature selection.

- Similarly, we update monolingual covariance matrices as well
  - \(C_{xx} \leftarrow C_{xx} \odot I_{xx}\) and \(C_{yy} \leftarrow C_{yy} \odot I_{yy}\)

- Updated covariance matrices are used to find the projection directions (A & B)

Synthetic Data

\[
z \sim N(0, I_d)
\]
\[
x|z \sim (W_1 z + \mu_1) + \eta N(0, I_d)
\]
\[
y|z \sim (W_2 z + \mu_2) + \eta N(0, I_d)
\]

The true feature correspondence are known \((I_d)\), thus exact sparsity is known.

Covariance Selection

1. Word Pair Association
2. Selection Strategies

1. Covariance
2. Mutual Information
3. Yule’s \(\omega\)
   \[
   \omega = \frac{n_{00}n_{11} - n_{01}n_{10}}{n_{00}n_{11} + n_{01}n_{10}}
   \]
4. Bilingual Dictionary

Experimental Results

Conclusions

1. Sparsifying all the covariance matrices is very useful.
2. Selection criteria is important.
3. External resources such as bilingual dictionaries gave better results.