Bridging Languages Through Images with Deep Partial Canonical Correlation Analysis

GUY ROTMAN ¹, IVAN VULIĆ ² & ROI REICHART ¹

¹ Faculty of Industrial Engineering and Management, Technion, IIT
² Language Technology Lab, University of Cambridge
Motivation
Motivation

- A visual scene can be described in any language

- Imagine that you are sitting in a restaurant in a foreign country and you need a spoon...
Goal

- Find a shared space for textual inputs from several languages
- Utilize mutual images to bridge between the textual inputs

**English**
A man is sitting at a table holding a spoon

**Spanish**
Un hombre está sentado en una mesa sujetando una cuchara
Technical Details
Multilingual Word Embeddings

- Vectors in different languages are in different spaces
Multilingual Word Embeddings

- Vectors in different languages are in different spaces
Mapping Two Views To a Shared Space: Canonical Correlation Analysis (CCA)

- CCA (Hotelling, 1936) is a statistical technique for finding linear projections of two random matrices such that their projected columns are maximally correlated.
Mapping Two Views To a Shared Space: Canonical Correlation Analysis (CCA)

• **Objective in matrix form:**

\[
\begin{align*}
\min_{\theta = \{W,V\}} & \quad \frac{1}{N-1} ||W^TX - V^TY||_F^2 \\
\text{Subject to} & \quad W^T\hat{\Sigma}_{xx}W = V^T\hat{\Sigma}_{yy}V = I
\end{align*}
\]

- \(\hat{\Sigma}_{xy} = \frac{1}{N-1}XY^T\), \(\hat{\Sigma}_{xx} = \frac{1}{N-1}XX^T\), \(\hat{\Sigma}_{yy} = \frac{1}{N-1}YY^T\)

- \(X, Y \) have zero mean
Limitations of CCA

- Projection is linear

- Inapplicable for large datasets due to whitening constraints:
  - Hard to compute stochastic estimations of the covariance matrices
  - Objective does not decompose over samples

- Cannot benefit from an additional view (such as images)
Partial CCA (PCCA)

- PCCA (Rao, 1969) is a statistical technique for finding linear maximal correlated projections of two random matrices conditioned on a third variable

\[ \max_{\theta = \{W, V\}} \text{Corr} (W^T (X|Z), V^T (Y|Z)) \]

- Z (a visual input) is a mutual variable of X and Y (textual inputs)
- PCCA was not used before in the multilingual multimodal setup
New model - Deep Partial CCA (DPCCA)

- CCA has a deep variant – Deep CCA (Andrew et al., 2013)
New model - Deep Partial CCA (DPCCA)

- CCA has a deep variant — Deep CCA (Andrew et al., 2013)

- Can we develop a deep variant for Partial CCA?

  - Partial CCA suffers from similar limitations to those of CCA

  - A new stochastic optimization algorithm is required
The DPCCA Model
Architecture of Deep Partial CCA (DPCCA) - Variant A

A man is sitting at a table holding a spoon

Un hombre está sentado en una mesa sujetando una cuchara
Architecture of Deep Partial CCA (DPCCA) - Variant B

A man is sitting at a table holding a spoon

Un hombre está sentado en una mesa sujetando una cuchara
Deep Partial CCA (DPCCA)

• (1) learn non-linear representations of $X$ and $Y$:

$$F(X) = W^T f(X), \quad G(Y) = V^T g(Y)$$

• $f$ and $g$ are two deep neural networks
• $W$ and $V$ are the final projection matrices
Deep Partial CCA (DPCCA)

• (2) perform multivariate linear multiple regressions for \( F(X) \) and \( G(Y) \) on a shared variable \( Z \):

\[
F(X) = \underbrace{AZ}_{\text{explained}} + \underbrace{F(X|Z)}_{\text{residual}}
\]

\[
G(Y) = \underbrace{BZ}_{\text{explained}} + \underbrace{G(Y|Z)}_{\text{residual}}
\]

\[
\min_A \frac{1}{N-1} \left| \left| F(X) - AZ \right| \right|_F^2
\]

\[
\min_B \frac{1}{N-1} \left| \left| G(Y) - BZ \right| \right|_F^2
\]
Deep Partial CCA (DPCCA)

• (2) perform multivariate linear multiple regressions for \( F(X) \) and \( G(Y) \) on a shared variable \( Z \):

\[
F(X) = AZ + F(X|Z)
\]

\[
G(Y) = BZ + G(Y|Z)
\]

\[
\begin{align*}
\min_A & \frac{1}{N-1} ||F(X) - AZ||_F^2 \\
\min_B & \frac{1}{N-1} ||G(Y) - BZ||_F^2
\end{align*}
\]

• (3) compute the residual matrices and their covariances w.r.t. the optimal solutions:

\[
F(X|Z) = F(X) - \hat{A}Z
\]

\[
\hat{\Sigma}_{FF|Z} = \frac{1}{N-1} F(X|Z)F(X|Z)^T
\]

\[
G(Y|Z) = G(Y) - \hat{B}Z
\]

\[
\hat{\Sigma}_{GG|Z} = \frac{1}{N-1} G(Y|Z)G(Y|Z)^T
\]
Deep Partial CCA (DPCCA)

- (4) perform CCA on the residuals:

\[
\min_{\theta = \{W_f, W, V_f, V\}} \frac{1}{N - 1} \| F(X|Z) - G(Y|Z) \|_F^2
\]

Subject to \( \hat{\Sigma}_{FF|Z} = \hat{\Sigma}_{GG|Z} = I \)
Deep Partial CCA (DPCCA) – Optimization

- Optimization is not trivial
Deep Partial CCA (DPCCA) – Optimization

- *Optimization is not trivial*

- *We introduce new stochastic optimization algorithms for our DPCCA variants*

- *Full Pseudocode is given in the paper*
Deep Partial CCA (DPCCA) – Optimization

• *Optimization is not trivial*

• *We introduce new stochastic optimization algorithms for our DPCCA variants*

• *We adopt some key techniques from the Nonlinear Orthogonal Iteration (NOI) algorithm which was suggested for Deep CCA (Wang et al., 2015)*

• *Full Pseudocode is given in the paper*
Experiments and Results
Experimental Setup – Tasks and Datasets

• **First Task:** Cross-lingual image description retrieval

  - **English**
    
    A man is sitting at a table holding a spoon

  - **Spanish**
    
    Un hombre está sentado en una mesa sujetando un tenedor
    
    Un hombre está sentado en una mesa sujetando una cuchara
    
    Un hombre está sentado en un balcón sujetando una cuchara

• **Dataset:** Multi30k (*Elliott et al.*, 2016)
Experimental Setup – Tasks and Datasets

- **First Task:** Cross-lingual image description retrieval

  **English**
  
  A man is sitting at a table holding a spoon

  **Spanish**
  
  Un hombre está sentado en una mesa sujetando un tenedor

  **Un hombre está sentado en una mesa sujetando una cuchara**

  **Un hombre está sentado en un balcón sujetando una cuchara**

- **Dataset:** *Multi30k* (Elliott et al., 2016)
### Experimental Setup – Tasks and Datasets

- **Second Task: Multilingual Word Similarity**

<table>
<thead>
<tr>
<th>English</th>
<th>German</th>
<th>Italian</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>inspect-examine</td>
<td>prüfen-überprüfen</td>
<td>inspezionare-esaminare</td>
<td>осматривать-изучать</td>
</tr>
<tr>
<td>easy-flexible</td>
<td>leicht-flexibel</td>
<td>facile-flessibile</td>
<td>покладистый-гибкий</td>
</tr>
<tr>
<td>plane-airport</td>
<td>flugzeug-flughafen</td>
<td>aereo-aeroporto</td>
<td>самолет-аэропорт</td>
</tr>
</tbody>
</table>

- **Dataset: Multilingual Simlex-999 (Leviant and Reichart., 2015)**
New Dataset – Word Image Word (WIW)

- Word pairs in different languages with mutual images

- The new dataset is available at: github.com/rotmanguy/DPCCA
Experimental Setup - Baselines

- *Linear and deep CCA-based models:*
  - *Probabilistic Partial CCA (PPCCA) (Mukuta, 2014) – T*
  - *Nonparametric CCA (NCCA) (Michaeli et al., 2016) - T*
  - *Generalized CCA (GCCA) (Horst, 1961) – TI*
  - *Deep CCA (DCCA) with various optimization algorithms – T*
  - *Deep CCA Autoencoder (DCCAE) (Wang et al., 2015) – T*

*Text – T,  Text + Images – TI*
Experimental Setup - Baselines

- **Linear and deep CCA-based models:**
  - Probabilistic Partial CCA (PPCCA) (Mukuta, 2014) – T
  - Nonparametric CCA (NCCA) (Michaeli et al., 2016) - T
  - Generalized CCA (GCCA) (Horst, 1961) – TI
  - Deep CCA (DCCA) with various optimization algorithms – T
  - Deep CCA Autoencoder (DCCAE) (Wang et al., 2015) – T

- **Other related works:**
  - Bridge Correlational Networks (BCN) (Rajendran et al., 2016) – TI
  - Image Pivoting (Gella et al., 2017) – TI

Text – T, Text + Images – TI
Main Results

• *PCCA* gets very good results, outperforming NN based methods and linear methods (including CCA, Image Pivoting, BCN ...)

• *DPCCA* is the best model, outperforming all baseline

• *Training with images improves performance on words that are more abstract, such as adjectives and verbs*
Cross-lingual Image Description Retrieval

<table>
<thead>
<tr>
<th>Model</th>
<th>English to German</th>
<th>German to English</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPCCA Variant A</td>
<td>83.6%</td>
<td>82.7%</td>
</tr>
<tr>
<td>DPCCA Variant B</td>
<td>84.8%</td>
<td><strong>83.9%</strong></td>
</tr>
<tr>
<td>DPCCA Variant B + DCCA NOI (Concatenation)</td>
<td><strong>86.3%</strong></td>
<td>83.7%</td>
</tr>
<tr>
<td>DCCA NOI</td>
<td>84.9%</td>
<td>83.0%</td>
</tr>
<tr>
<td>IMG PIVOTING</td>
<td>78.9%</td>
<td>78.1%</td>
</tr>
<tr>
<td>BCN</td>
<td>62.8%</td>
<td>62.9%</td>
</tr>
<tr>
<td>PCCA</td>
<td><strong>82.4%</strong></td>
<td><strong>78.7%</strong></td>
</tr>
<tr>
<td>CCA</td>
<td>80.3%</td>
<td>75.4%</td>
</tr>
<tr>
<td>GCCA</td>
<td>74.2%</td>
<td>74.3%</td>
</tr>
</tbody>
</table>

- Results are reported on BLEU + 1
## Multilingual Word Similarity

<table>
<thead>
<tr>
<th>Model</th>
<th>EN - ADJ</th>
<th>EN - Verbs</th>
<th>EN - Nouns</th>
<th>DE - ADJ</th>
<th>DE - Verbs</th>
<th>DE - Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPCCA Variant A</td>
<td>64.0%</td>
<td>31.1%</td>
<td>36.9%</td>
<td>43.0%</td>
<td>32.1%</td>
<td>40.4%</td>
</tr>
<tr>
<td>DPCCA Variant B</td>
<td>62.6%</td>
<td><strong>31.6%</strong></td>
<td><strong>38.2%</strong></td>
<td><strong>46.2%</strong></td>
<td>31.9%</td>
<td>39.9%</td>
</tr>
<tr>
<td>DCCA NOI</td>
<td>61.1%</td>
<td>30.8%</td>
<td>36.1%</td>
<td>44.1%</td>
<td>29.7%</td>
<td>39.8%</td>
</tr>
<tr>
<td>PCCA</td>
<td>61.4%</td>
<td>29.6%</td>
<td>34.0%</td>
<td>30.5%</td>
<td>14.3%</td>
<td>34.0%</td>
</tr>
<tr>
<td>CCA</td>
<td>55.7%</td>
<td>29.7%</td>
<td>32.1%</td>
<td>28.4%</td>
<td>15.7%</td>
<td>34.6%</td>
</tr>
<tr>
<td>GCCA</td>
<td>63.6%</td>
<td>28.0%</td>
<td>37.8%</td>
<td>44.6%</td>
<td>27.7%</td>
<td>39.8%</td>
</tr>
</tbody>
</table>

- Results are reported on Spearman's correlation coefficient
Summary

• Goal: Learning a shared bilingual space for textual inputs
Summary

• **Goal:** Learning a shared bilingual space for textual inputs

• **Our Contributions:**
  • **Method:** Adding mutual visual information to the learning process
  • **Model:** Applying PCCA to our settings, and introducing its deep variants
  • **Optimization:** New optimization algorithm for DPCCA
  • **Results:** Improvements over previous work
  • **New Dataset:** Word Image Word (WIW)
Future Work

- Expanding DPCCA to support more than two languages
- Exploiting the internal structure of images and sentences
Thank you!

- Code and data are available at:

  github.com/rotmanguy/DPCCA