Cross-lingual transfer learning: A PARAFAC2 approach

HPC Application

Signal Processing and Information Analysis (AIIA.SPINAL) group

Department of Informatics, Aristotle University of Thessaloniki, Greece

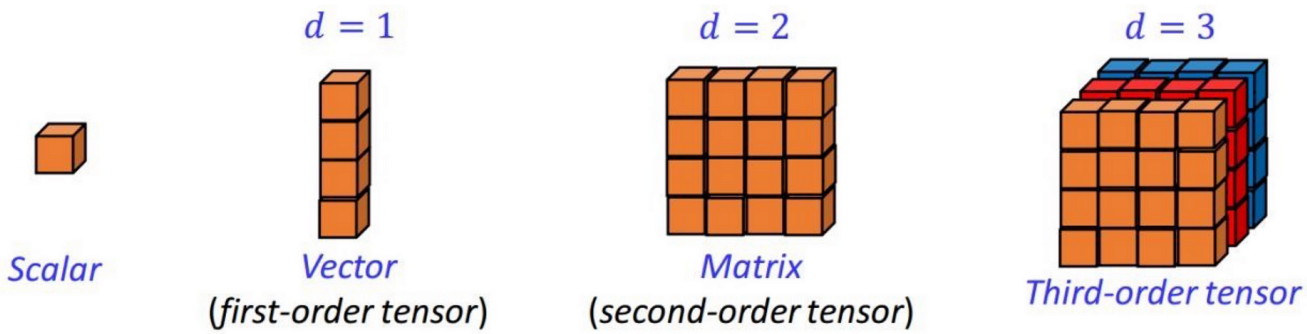
ioannis.tsingalis@gmail.com

October 16, 2023

Outline

Basic Notations

- Tensors are considered as the multidimensional equivalent of matrices (i.e., second-order tensors) and vectors (i.e., first-order tensors).



- Tensors are denoted by boldface Euler script calligraphic letters (e.g. $\mathcal{X}$), matrices are denoted by uppercase boldface letters (e.g., $\mathbf{U}$), vectors are denoted by lowercase boldface letters (e.g., $\mathbf{u}$), and scalars are denoted by lowercase letters (e.g., u).
- A slice of a tensor $\mathcal{X}$ is denoted by $\boldsymbol{X}^{(n)}$, where n is the number of the slice. We have the slices $\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$, and $\boldsymbol{X}^{(3)}$.

PARAFAC2 method

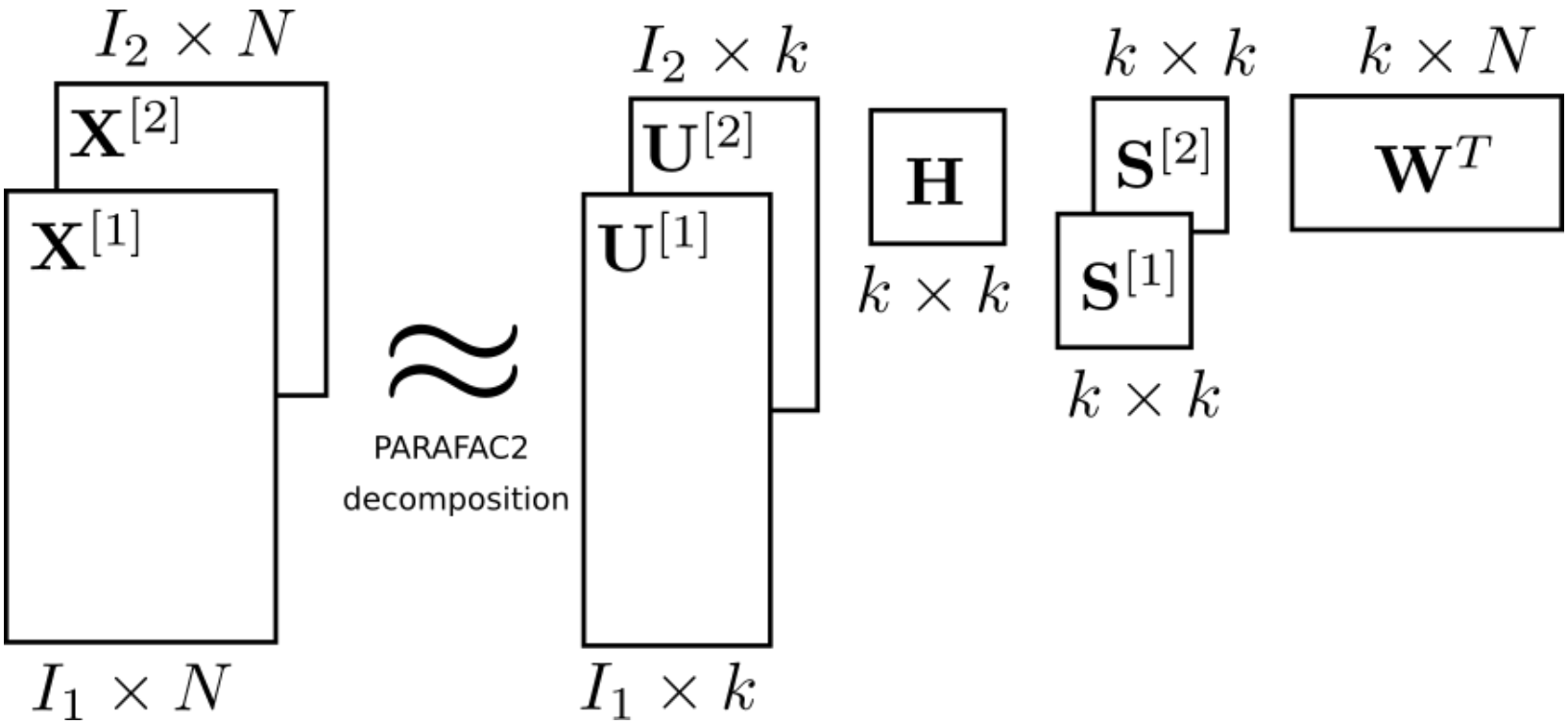
- In the PARAFAC2 (PARAllel FACtor analysis2)[1] model, we seek a decomposition of the form

$$\mathbf{X}^{(n)} \approx \mathbf{U}^{(n)}\,\mathbf{H}\,\mathbf{S}^{(n)}\,\mathbf{W}^{T}, \quad n = 1, 2, \ldots, L \tag{1}$$

  where
  - $\mathbf{X}^{(n)} \in \mathbb{R}^{I_n \times N}$, $n = 1, 2, \ldots, L$
  - $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times k}$, $n = 1, 2, \ldots, L$ is an orthonormal matrix
  - $\mathbf{H} \in \mathbb{R}^{k \times k}$ is a square matrix,
  - $\mathbf{S}^{(n)} \in \mathbb{R}^{k \times k}$ is a diagonal matrix of weights for the n-th slice of $\mathcal{X}$, and
  - $\mathbf{W} \in \mathbb{R}^{N \times k}$ is a coefficient matrix.

[1] H. A. Kiers, J. M. Ten Berge, and R. Bro, PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model. Journal of Chemometrics: A Journal of the Chemometrics Society, 13(3-4), 275-294, 1999.

PARAFAC2 method



PARAFAC2 method

$$\underset{\mathbf{U}^{(n)},\,\mathbf{H},\,\mathbf{S}^{(n)},\,\mathbf{W}}{\operatorname{argmin}} \sum_{n=1}^{L} \|\mathbf{X}^{(n)} - \mathbf{U}^{(n)}\,\mathbf{H}\,\mathbf{S}^{(n)}\,\mathbf{W}^{T}\|_F^2. \tag{2}$$

Algorithm 1 Computation of PARAFAC2

Input: latent dimension R, tolerance $\epsilon > 0$

Output: $\mathbf{U}^{[I]}$, $\mathbf{H}$, $\mathbf{S}^{[I]}$, and $\mathbf{W}$

1: Initialize $\mathbf{W} \sim \mathcal{U}_{[-1,1]}$ and $\mathbf{H}$, $\mathbf{S}^{[I]}$, $I = 1, 2$ as $\boldsymbol{I}_k$, a $k \times k$ identity matrix.

2: **while** $\texttt{fit}(\mathcal{X}, \hat{\mathcal{X}}) > \epsilon$ **do**

3: $\quad \mathbf{U}^{[I]} = \operatorname{argmin}_{\mathbf{U}^{[I]}} \operatorname{tr}\left(\mathbf{H}\,\mathbf{S}^{[I]}\,\mathbf{W}^{T}\,{\mathbf{X}^{[I]}}^{T}\,\mathbf{U}^{[I]}\right)$

4: $\quad \mathbf{H} = \left[\sum_{I=1}^{2} {\mathbf{U}^{[I]}}^{T}\,\mathbf{X}^{[I]}\,\mathbf{W}\,\mathbf{S}^{[I]}\right]\left[\sum_{I=1}^{2} \mathbf{S}^{[I]}\,\mathbf{W}^{T}\,\mathbf{W}\,\mathbf{S}^{[I]}\right]^{-1}$

5: $\quad \mathbf{S}^{[I]} = \texttt{diag}\left(\left[\left(\mathbf{W}^{T}\,\mathbf{W}\right) \circ \left(\mathbf{H}^{T}\,\mathbf{H}\right)\right]^{-1} \texttt{diag}\left(\mathbf{H}^{T}\,{\mathbf{U}^{[I]}}^{T}\,\mathbf{X}^{[I]}\,\mathbf{W}\right)\right)$

6: $\quad \mathbf{W} = \left[\sum_{I=1}^{2} {\mathbf{X}^{[I]}}^{T}\,\mathbf{U}^{[I]}\,\mathbf{H}\,\mathbf{S}^{[I]}\right]\left[\sum_{I=1}^{2} \mathbf{S}^{[I]}\,\mathbf{H}^{T}\,\mathbf{H}\,\mathbf{S}^{[I]}\right]^{-1}$

7: **end while**

Cross-Lingual Transfer Learning

Concept

- Cross-lingual transfer learning is a technique used in natural language processing and machine learning to apply knowledge learned from one language to another language.
- Here are the key benefits of cross-lingual transfer learning:
  1. Reduced Annotation Costs: Training models from scratch in a new language requires extensive annotated data. By transferring knowledge from a source language, you can reduce the need for costly data annotation in the target language.
  2. Multilingual applications: sentiment analysis, authorship attribution, and more.

Cross-Lingual Transfer Learning

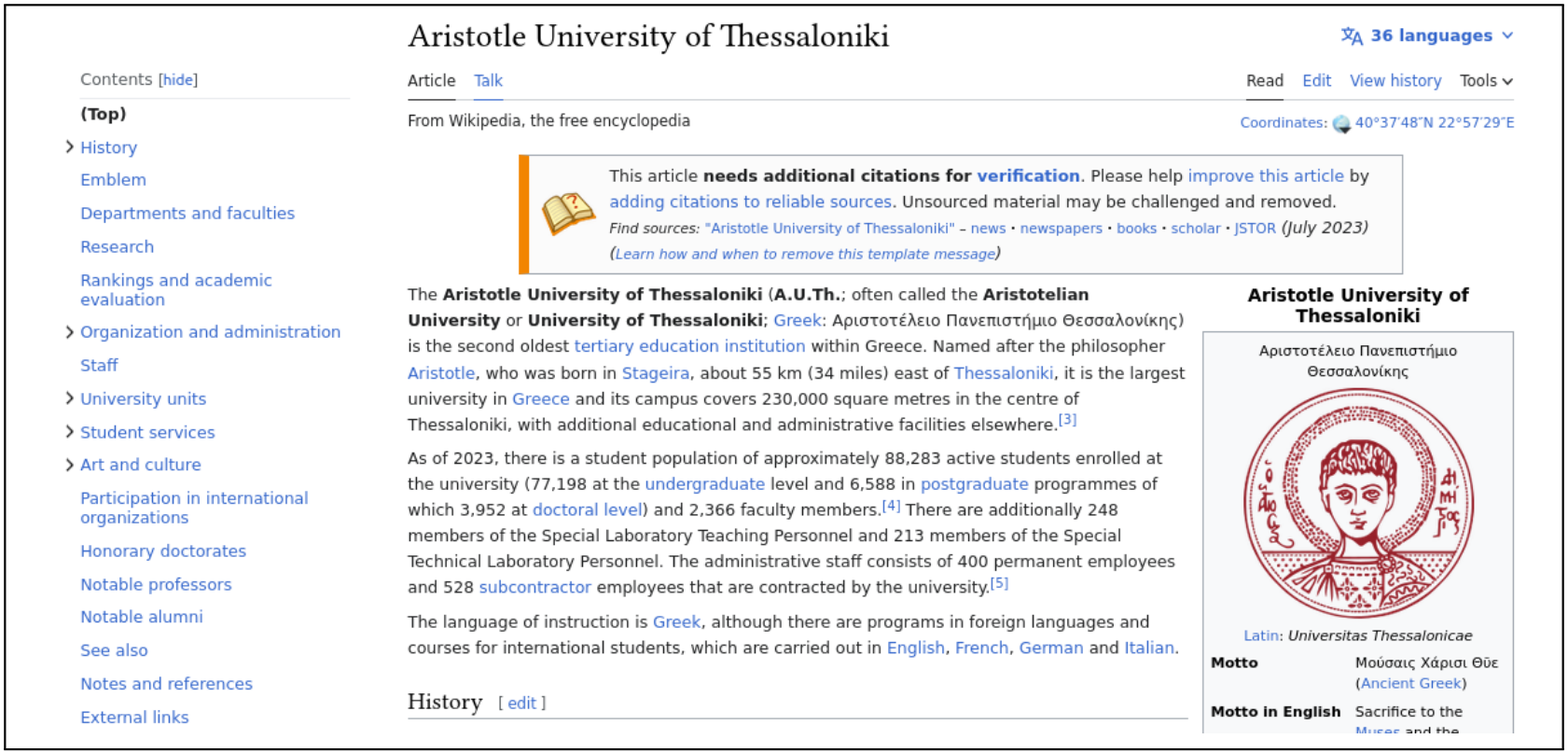
Steps

Basic steps in Cross-Lingual Transfer Learning:

1. Train a natural language processing model in parallel corpora.
   - Parallel corpora are large collections of texts where each document in one language corresponds to a similar or a translation of a document in another language. The trained model has learned conceptual connections between the languages.
2. Use the trained model to transfer knowledge from one language to another to solve a specific task, e.g., document classification.

Cross-Lingual Transfer Learning

English Parallel Data



Contents [hide]

(Top)

- History
- Emblem
- Departments and faculties
- Research
- Rankings and academic evaluation
- Organization and administration
- Staff
- University units
- Student services
- Art and culture
- Participation in international organizations
- Honorary doctorates
- Notable professors
- Notable alumni
- See also
- Notes and references
- External links

Aristotle University of Thessaloniki

36 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

Coordinates: 40°37′48″N 22°57′29″E

This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed.

Find sources: "Aristotle University of Thessaloniki" – news · newspapers · books · scholar · JSTOR *(July 2023)*

(Learn how and when to remove this template message)

The **Aristotle University of Thessaloniki** (**A.U.Th.**; often called the **Aristotelian University** or **University of Thessaloniki**; Greek: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης) is the second oldest tertiary education institution within Greece. Named after the philosopher Aristotle, who was born in Stageira, about 55 km (34 miles) east of Thessaloniki, it is the largest university in Greece and its campus covers 230,000 square metres in the centre of Thessaloniki, with additional educational and administrative facilities elsewhere.[3]

As of 2023, there is a student population of approximately 88,283 active students enrolled at the university (77,198 at the undergraduate level and 6,588 in postgraduate programmes of which 3,952 at doctoral level) and 2,366 faculty members.[4] There are additionally 248 members of the Special Laboratory Teaching Personnel and 213 members of the Special Technical Laboratory Personnel. The administrative staff consists of 400 permanent employees and 528 subcontractor employees that are contracted by the university.[5]

The language of instruction is Greek, although there are programs in foreign languages and courses for international students, which are carried out in English, French, German and Italian.

History [edit]

Aristotle University of Thessaloniki

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Latin: *Universitas Thessalonicae*

Motto	Μούσαις Χάρισι Θῦε (Ancient Greek)
Motto in English	Sacrifice to the Muses and the

Cross-Lingual Transfer Learning

Greek Parallel Data

Περιεχόμενα [απόκρυψη]

Αρχή
Ιστορία
Έμβλημα
Σχολές και τμήματα
Έρευνα
Ακαδημαϊκή Αξιολόγηση
Εγκαταστάσεις
Βιβλιοθήκη
Βιβλιοθήκες τμημάτων
Φοιτητική Λέσχη
Πανεπιστημιακό Γυμναστήριο
Φοιτητικές Εστίες
Εγκληματικότητα
Υπηρεσίες
Κέντρο Ηλεκτρονικής Διακυβέρνησης
Μονάδα Σημασιολογικού Ιστού

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

36 γλώσσες

Λήμμα Συζήτηση Ανάγνωση Επεξεργασία Επεξεργασία κώδικα Προβολή ιστορικού Εργαλεία

Από τη Βικιπαίδεια, την ελεύθερη εγκυκλοπαίδεια

Συντεταγμένες: 40°37′48″N 22°57′29″E

Το **Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης** (συντομογραφία: **ΑΠΘ**), γνωστό και ως **Αριστοτέλειο Πανεπιστήμιο** ή **Πανεπιστήμιο Θεσσαλονίκης**, είναι ανώτατο εκπαιδευτικό ίδρυμα της Ελλάδας, με έδρα την Θεσσαλονίκη και έτος ίδρυσης το 1925. Σήμερα στο πανεπιστήμιο λειτουργούν συνολικά περισσότερα από σαράντα τμήματα, τα οποία οργανώνονται σε έντεκα σχολές, καλύπτοντας ένα ευρύ φάσμα επιστημονικών πεδίων.[1]

Το Αριστοτέλειο Πανεπιστήμιο είναι το μεγαλύτερο σε αριθμό φοιτητών εκπαιδευτικό ίδρυμα της Ελλάδας[2] και είναι διεθνώς αναγνωρισμένο για το εκπαιδευτικό και ερευνητικό ακαδημαϊκό του έργο.[3][4][5][6][7]

Ιστορία [Επεξεργασία | επεξεργασία κώδικα]

Κύριο λήμμα: Ιστορία του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης

Ο ιδρυτικός νόμος του Πανεπιστημίου Θεσσαλονίκης (Ν. 3341/14-6-25) ψηφίστηκε στην Βουλή των Ελλήνων στις 5 Ιουνίου του 1925 και δημοσιεύτηκε στο φύλλο της Εφημερίδας της Κυβερνήσεως την 22α Ιουνίου 1925.[8] Η ίδρυση του δεύτερου ελληνικού Πανεπιστημίου συνοδεύτηκε στα πρώτα στάδια όμως με αρκετές δυσκολίες, απότοκες της πολιτικής ρευστότητας της εποχής.[9]

Ο καθηγητής του Πανεπιστημίου του Βερολίνου Κωνσταντίνος Καραθεοδωρή είχε προτείνει την ίδρυση ενός δεύτερου ελληνικού πανεπιστημίου καθώς μέχρι εκείνη την εποχή η Αθήνα

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Παλαιότερες ονομασίες	Πανεπιστήμιο Θεσσαλονίκης (22 Ιουνίου 1925 - 30 Δεκεμβρίου 1954)
Ρητό	Μούσαις Χάρισι Θῦε (Θυσίαζε στις Μούσες και στις Χάριτες)
Ίδρυση	22 Ιουνίου 1925, πριν 98 έτη

Cross-Lingual Transfer Learning

Framework

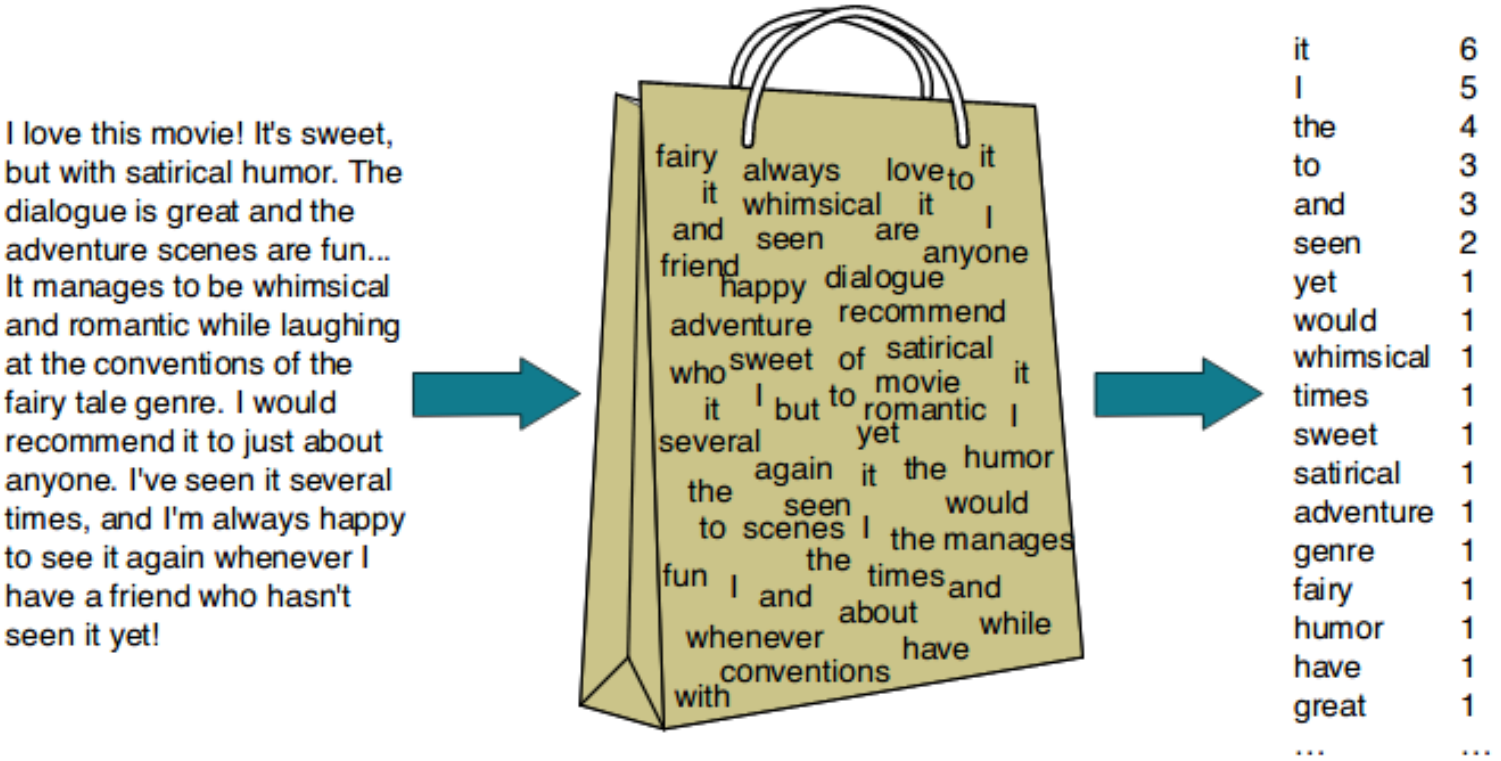


Figure 1: Source : Bag of words!

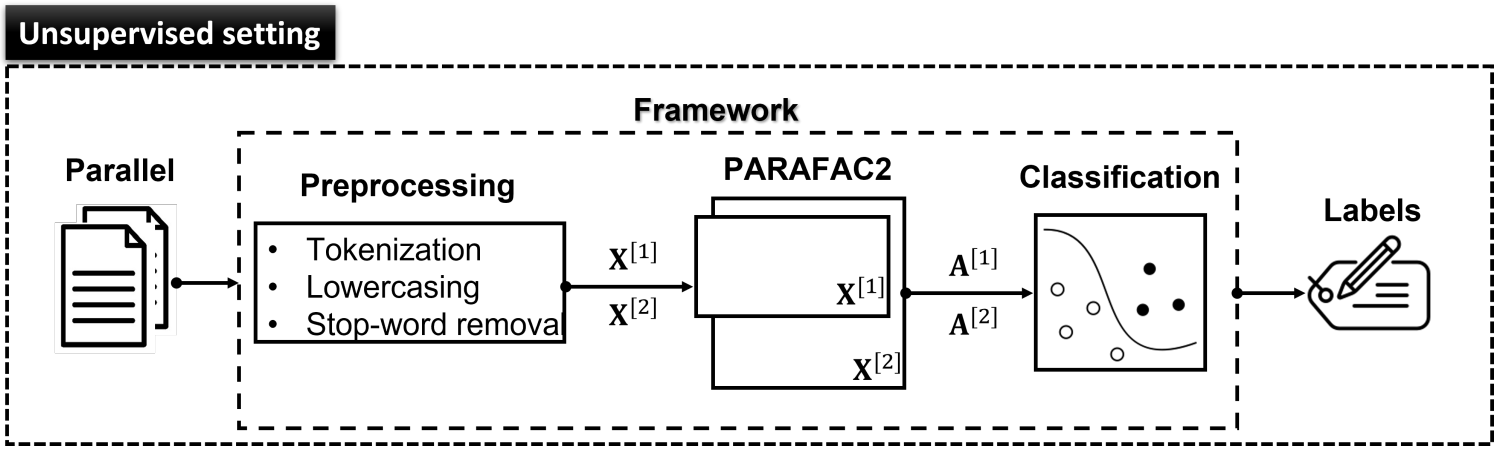
Dataset Description

MLDOC

MIDoc is an improved version of the Reuters benchmark dataset with balanced class priors for eight languages: English, German, Spanish, French, Italian, Russian, Japanese, and Chinese. It comprises 1,000 training and validation documents and 4,000 test documents for each language divided in 4 classes: Corporate/Industrial (CCAT), Economics (ECAT), Government/Social (GCAT), and Markets (MCAT).

Cross-Lingual Transfer Learning

Framework

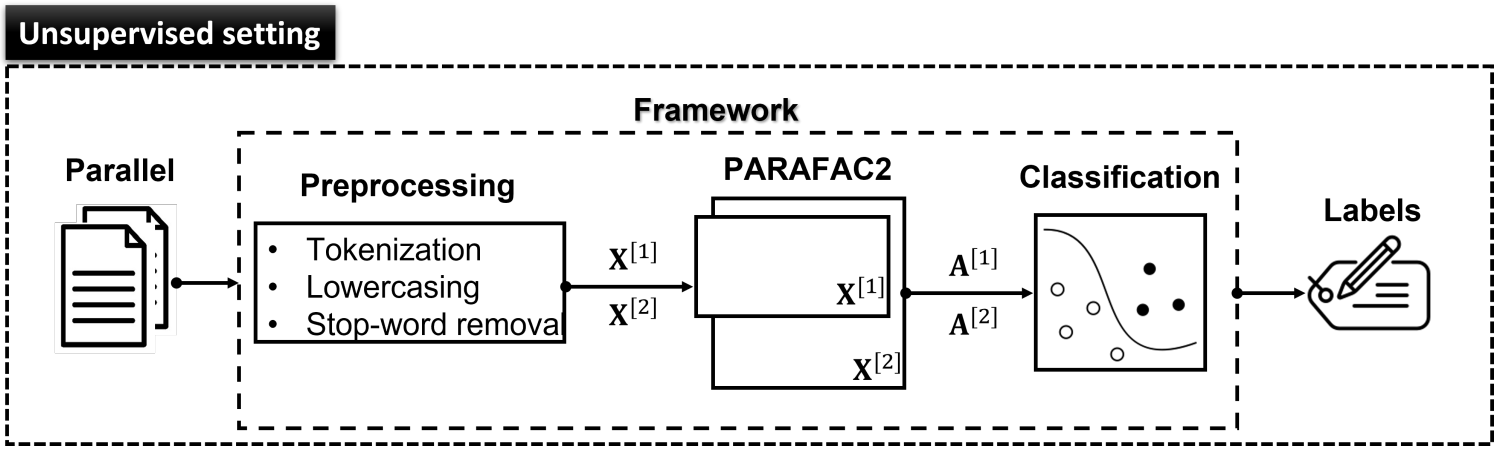


$$\underset{\mathbf{U}^{(n)},\,\mathbf{H},\,\mathbf{S}^{(n)},\,\mathbf{W}}{\operatorname{argmin}} \sum_{n=1}^{L} \|\mathbf{X}^{(n)} - \mathbf{U}^{(n)}\,\mathbf{H}\,\mathbf{S}^{(n)}\,\mathbf{W}^{T}\|_F^2.$$

- Wikimedia parallel dataset:
  1. English language is the source language represented by $\mathbf{X}^{[1]}$.
  2. Greek is the target language represented by $\mathbf{X}^{[2]}$.
  3. The goal is to create a general purpose cross-lingual language model using PARAFAC2.

Cross-Lingual Transfer Learning

Framework



- Multilingual Document Classification corpus (MIDoc) dataset:
  1. English language is the source language represented by $\mathbf{X}^{[1]}$. $\mathbf{X}^{[1]}$ is used to obtain $\mathbf{A}^{[1]} = {\mathbf{U}^{[1]}}^T \mathbf{X}^{[1]} \in \mathbb{R}^{k \times N}$ and train a classifier, e.g., Logistic Regression Classifier.
  2. Greek language is the target language represented by $\mathbf{X}^{[2]}$. $\mathbf{X}^{[2]}$ is used to obtain $\mathbf{A}^{[2]} = {\mathbf{U}^{[2]}}^T \mathbf{X}^{[2]} \in \mathbb{R}^{k \times N}$ and test the classifier.
  3. We achieve French document classification using only the labeled English documents.

Results

MLDOC

Table 1: Accuracies on the MIDoc zero-shot cross-lingual document classification task (test set) in 4 classes.

	EN → XX						
	DE	**ES**	**FR**	**IT**	**JA**	**RU**	**ZH**
PARAFAC2 (our)	88.20	**81.50**	84.55	75.40	**70.10**	68.27	73.7
Artetxe and Schwenk (2019)	84.78	77.33	77.95	69.43	60.30	67.78	71.93
Schwenk and Li (2018)	81.20	72.50	72.38	69.38	67.63	60.80	74.73
Wu and Dredze (2019)	80.2	72.6	72.6	68.9	56.5	**73.7**	76.9
Eisenschlos et al. (2019)	**91.62**	79.10	**89.42**	**76.02**	69.57	67.83	**82.48**
Siddhant et al. (2020)	77.4	73.0	77.2	64.2	69.0	68.9	73.4
Artetxe et al. (2020)	88.7	77.0	87.3	-	-	67.6	78.3

Bibliography

E. Pantraki, I. Tsingalis, and C. Kotropoulos, "Cross-lingual transfer learning: A PARAFAC2 approach". Pattern Recognition Letters, 159, 167-173, 2022.

Shell script

Python script to train the PARAFAC2 model using Algorithm 1.

- rep. `https://github.com/epantrak/CrossLingualPARAFAC2/`
- python script: `cuParafac2.py`

```
def dsmm(mat1: Tensor, mat2: Tensor) -> Tensor:
    return sdmm(mat2.t(), mat1.t()).t()
import torch
from torch import mm as ddmm
from torch.sparse import mm as sdmm
def __partial_fit_H(self, X):

    lhs = torch.stack([dsmm(ddmm(dsmm(self._U[k].T, X[k]), self._W), self._S[k])
                       for k in range(self.__n_languages)]).sum(0)

    rhs = torch.stack([dsmm(sdmm(self._S[k], ddmm(self._W.T, self._W)), self._S[k])
                       for k in range(self.__n_languages)]).sum(0)

    self._H = ddmm(lhs, torch.linalg.inv(rhs))
```

$$\mathbf{H} = \left[\sum_{l=1}^{2} \mathbf{U}^{[l]\,T}\, \mathbf{X}^{[l]}\, \mathbf{W}\, \mathbf{S}^{[l]}\right] \left[\sum_{l=1}^{2} \mathbf{S}^{[l]}\, \mathbf{W}^{T}\, \mathbf{W}\, \mathbf{S}^{[l]}\right]^{-1}$$

Shell script

Shell script (run.sh) to train the PARAFAC2 model on the Aristotle HPC

```
#!/bin/bash
#SBATCH --partition=ampere
#SBATCH --job-name=run_Parafac2
#SBATCH --nodes=1
#SBATCH --mem=10G
#SBATCH --cpus-per-task=10
#SBATCH --gres=gpu:1
#SBATCH --time=0-01:00:00     # Run time (days-hh:mm:ss) - (max 7days)

module load centos8 gcc/11.2.0-25clxrk cuda/11.6.1-klwuvft miniconda3
source $CONDA_PROFILE/conda.sh
conda activate nnParafac2

python run_tensorModels.py --task MlDoc --target_lang german
```

Run in terminal with:

```
$ sbatch run.sh
```

Questions?

Thank you for your support HPC team!

Any Questions?