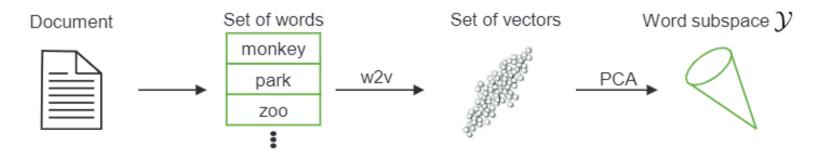




Subspace Representation for Natural Language Processing

Erica Kido Shimomoto*, François Portet+, Kazuhiro Fukui*

* University of Tsukuba, Japan + University of Grenoble Alpes, France



- Word Subspace: Compact, scalable and meaningful representation of the context of a set of words.
 - Low computational cost;
 - No limitation of number of words;
 - □ Interpretable: Basis vectors can be regarded as the main hidden topics of a text.

 $v_{elizabeth}$ v_{queen} v_{king} v_{Arthur} v_{legend}





Subspace Representation for Natural Language Processing

Erica Kido Shimomoto*, François Portet+, Kazuhiro Fukui*

* University of Tsukuba, Japan + University of Grenoble Alpes, France

- Text analysis based on the subspace representation.
- Word Importance Score: Measures how relevant a word w_k is with regards to a hidden topic ϕ_i .

$$I(\boldsymbol{x}_k, \boldsymbol{\phi}_i) = \boldsymbol{x}_k^{\mathsf{T}} \boldsymbol{\phi}_i$$

Basis vector 1	Basis vector 2	Basis vector 3	Basis vector 4
Thousands protesters protest public or protesters p	education education ahead second teachers mixt following science begins student Sunday Thursday education ahead second teachers mixt following science begins student S tudent S oz out	protestersdemonstrations Drote Cests Drote	Brousands later Duarte Of

Most important words with respect to the main hidden topics of a text about protests against cuts in education funding in Brazil.

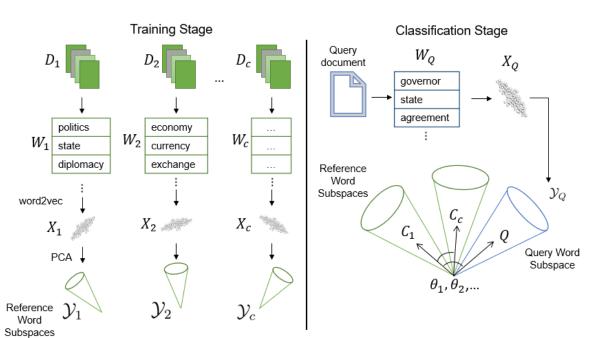




Subspace Representation for Natural Language Processing

Erica Kido Shimomoto*, François Portet+, Kazuhiro Fukui*

- * University of Tsukuba, Japan + University of Grenoble Alpes, France
- Geometrical interpretation of the word vectors -> Propose the best subspace-based method to solve each task.
- Topic Classification: MSM.



Sentiment Analysis: GOSM.

