







Partnership for International Research and Education A Global Living Laboratory for Cyberinfrastructure Application Enablement

CRL Healthcare Viz Project

Student: Ohannes Ohannessian, BSCE Student, Florida Atlantic University **Research Advisor:** Dr. Shihong Huang, Florida Atlantic University **CI-PIRE Partner Advisor:** Dr. Yangqiu Song, IBM – China Research Lab, Beijing, China



. Research Overview and Outcome

Problem Statement

Expansion of TIARA's visual analytic system dimensions

- Addition of geographical topology to its time-height algorithm

- Change the existing software to make it smart and able to analyze the large collection of texts by its geographical distribution

Topic General Overview

Data mining and clustering is becoming smarter and faster. Information retrieval is requesting the software world to develop exploratory textual display. Clustering is an essential machine learning method. There are supervised and unsupervised clustering algorithms. Unsupervised algorithms are still raw in a sense that more has to be done on and mostly focused on too. IBM's TIARA provides topic analysis with unsupervised topic modeling to cluster it and semantically latent it using LDA (Latent Dirichlet Allocation). And the text visualization that uses it is designed to visually analyze text analytics that previously were collected in clusters and injected by KDD (Knowledge Discovery of Data).

Research Results

geoTIARA

Existing Model



TIARA

TIARA's Multiple Corpora Time-Varying

Cluster mining evolution is taking different solutions for text analytics. As the title proposes, the evolutionary hierarchical Dirichlet processes (evoHDP) is an analysis of set of correlated timevarying data where it discovers cluster patterns that evolve within a corpus or with other corpora, emerge and disappears. evoHDP extends HDP. Time is involved in scaling the dependencies to build HDP for multiple corpora into adjacent epochs.

Dirichlet Process (DP) \rightarrow Hierarchical DP (HDP) \rightarrow Evolutionary HDP (evoHDP)

Future Works

•To make KDD (Knowledge Discovery in Data) smarter To find space field theories for clusters

References

1)TIARA: A Visual Exploratory Text Analytic System (Furu Wei, Shixia Liu, Yangqiu Song, Shimei Pan, Michelle X. Zhou, Weihong Qian, Lei Shi, Li Tan and Qiang Zhang – IBM) 2)Evelotionary Hierarchical Dirichlet Processes for Multiple Correlated Time-varying **Corpora** (Jianwen Zhang, Yangqiu Song, Changshui Zhang, Shixia Liu – Tsighua University &



Proposed geoTIARA's Multiple Corpora Time-Space-Varying

p-evoHDP (Post evoHDP) will expand the architecture that will generate a cluster of geographic indexed topic. Visualization will be coded to display it afterwards into a 3D $DP \rightarrow HDP \rightarrow evoHDP \rightarrow p-evoHDP$ representation.







II. International Experience



07.05.2010 '

Located 50 miles north of Beijing. It is one of the most beautiful sites of nature in China. Going in between the mountains its water is reserved. It's a unforgettable.

4.2010 17:35

nature, Chinese tradition and ambition to achieve were key enjoyment that I had while there. Seven weeks were not even enough to take a sip from years of culture that it exist. I enjoyed communicating with people, sometimes not even and educational growths with PIRE Program. My old mathematical solutions for clustering architecture.

<u>Tsinghua</u>

One of the oldest beautiful campus & was built in 1919.





III. Acknowledgement

The material presented in this poster is based upon the work supported by the National Science Foundation under Grant No. OISE-0730065. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.