CSIC 5011 Mini-Project 1:

NIPS Paper Explained

ZHA Mengyue[1], Huang Chutian[2] {mzha, chuangat}@ust.hk
Department of Mathematics, HKUST

1. Introduction

This project finds 5 hottest topics of the NIPS paper from 1987-2015. Amazingly, these topics can be identified as Computer Vision, Matrix Computation, Reinforcement Learning, Bayesian Methods and Time Series. Visualization outcomes of manifold methods strongly support the justification of topic extraction.

2. Data

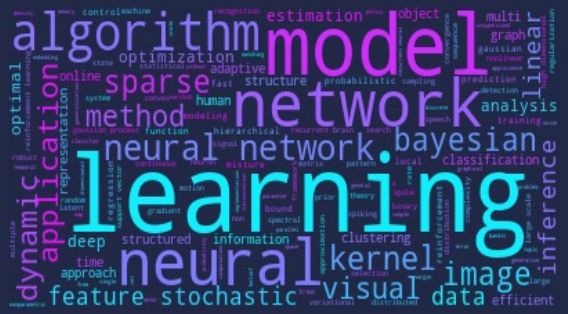


Fig1. Word Cloud by word frequencies

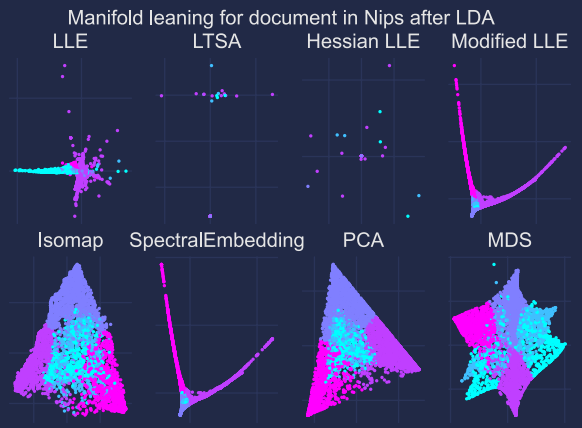
The dataset is a co-occurrence matrix for papers and words of size 11463 * 5804

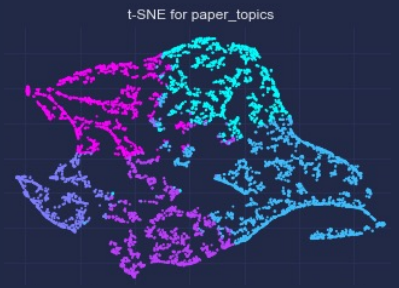
3. Aims & Methodology

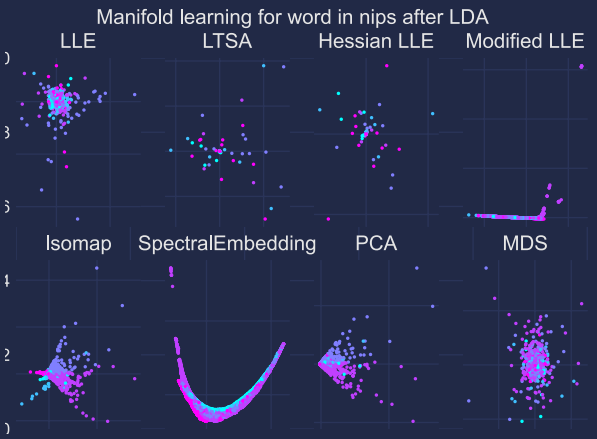
- Data Cleaning and Exploration. Showing Word Cloud, paper trends by year, topic trend by year.
- Latent Dirichlet Allocation(LDA). Extracting 5 hottest topic from the data. Build the relationships of topics to each paper and word.
- Manifold Learning. Visualizing papers and words with different topics.

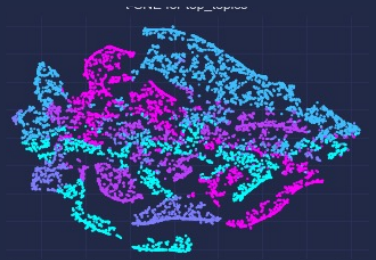
5. Reduction by LDA & Visualization by Manifold Learning

We visualize paper with the topic it belongs and word with topic it contributes to respectively via 7 manifold methods and classical PCA and MDS.









6. Print 5 Topics found by LDA

Topic0: training learning image set network features
Topic1: algorithm data function matrix learning problem
Topic2: learning state algorithm time policy function
Topic 3: model data models distribution inference gaussian
Topic 4: model time neural figure network neurons

4. Data Exploration

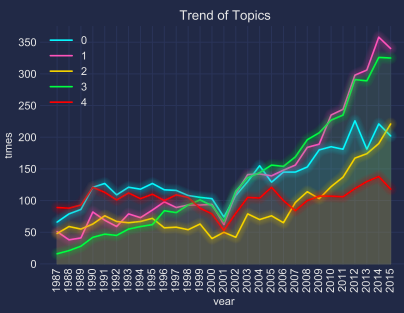


Fig2. Topic Trend by Year

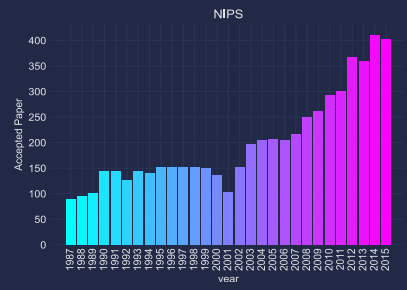


Fig3. Trend of the number of papers

7. Analysis & Conclusions

1. Same words appear in the Word Clouds and 5 topics given by LDA which preliminarily justifies our outcomes.
2. Readers can tell specific hot topics like Computer Vision, Matrix Computation, Reinforcement Learning, Bayesian Methods and Time Series from the topic found by LDA .
3. The 21st centuary saw a spike of paper quantity. Increasing trends of different topics are displayed above.
4. Visualization shows clear patterns of paper-topic, but not so good of word-topic, which further validates our results.

8. References

CSIC2011ADiveIntoNIPSWords Gu Hanlin, Huang Yifei, Sun Jiaze
Y.Yao, A Mathematical Introduction to Data Science
Ryanschaub,Kaggle,Hottest Topoics in Machine Learning

9. Contribution

ZHA Mengyue: Coding & Making the Poster
HUANG Chutian: Theory Support & Writing the Paper

💗💗💗 More explanations and outcomes, Welcome to our GitHub! 💗💗💗