



PROJECT RECOMMENDATION SYSTEM

Sudiksha Aapan (MT19018), Priyanka Boral (MT19127), Akanksha Dewangan (MT19049)

Indraprastha Institute of Information and Technology, New Delhi

Motivation

We are working on the project recommendation system which suggests projects to the students according to their interest and domain knowledge. Many topics has already made projects. So its better to have information about already implimented projects so student can make a better projects as per their pre-knowledge and interest and thats what out project does it provide platform to the students to get recommondation of projects as per their knowlege and specialization with exploration of untouched area.

Work Done

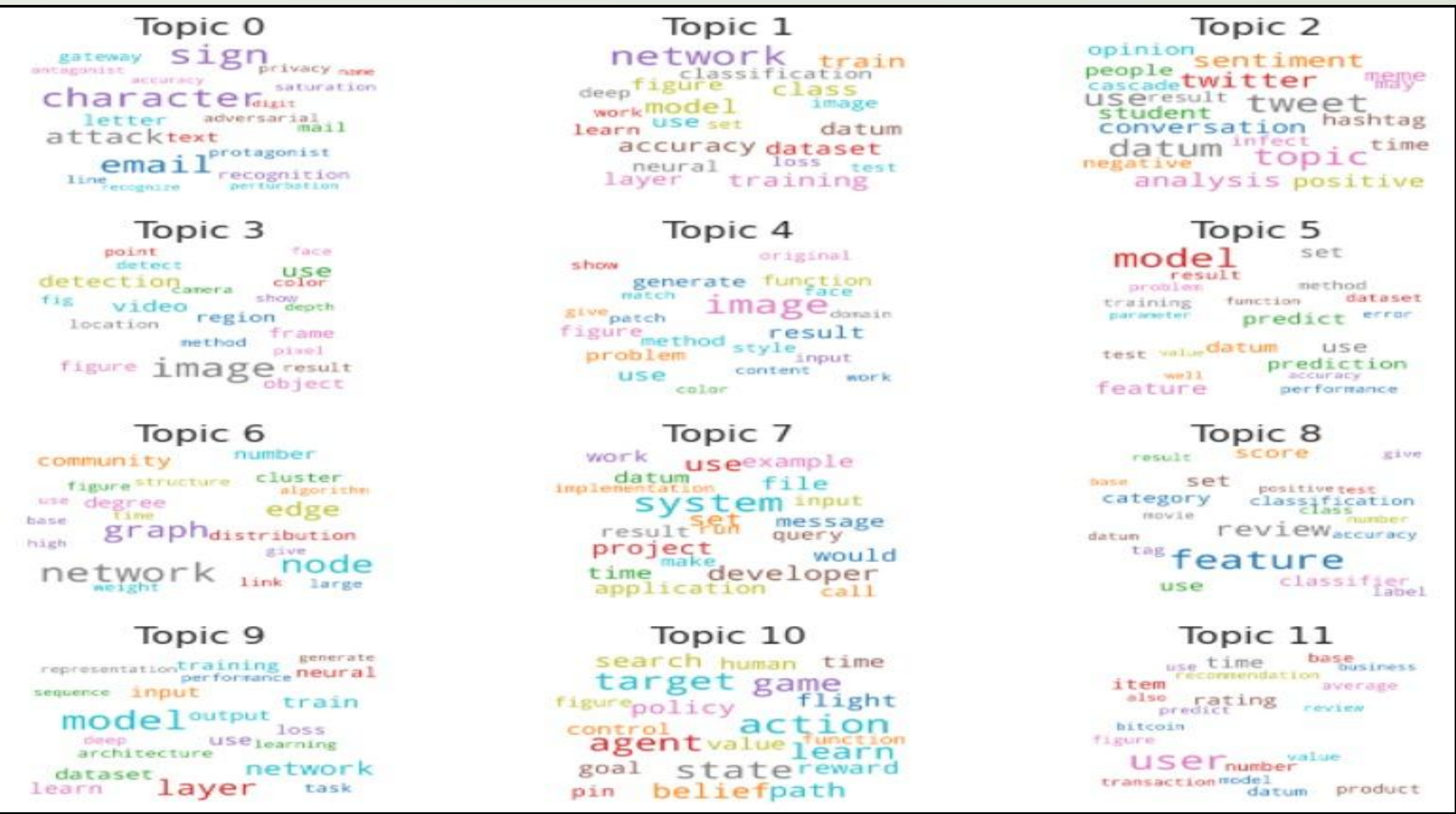
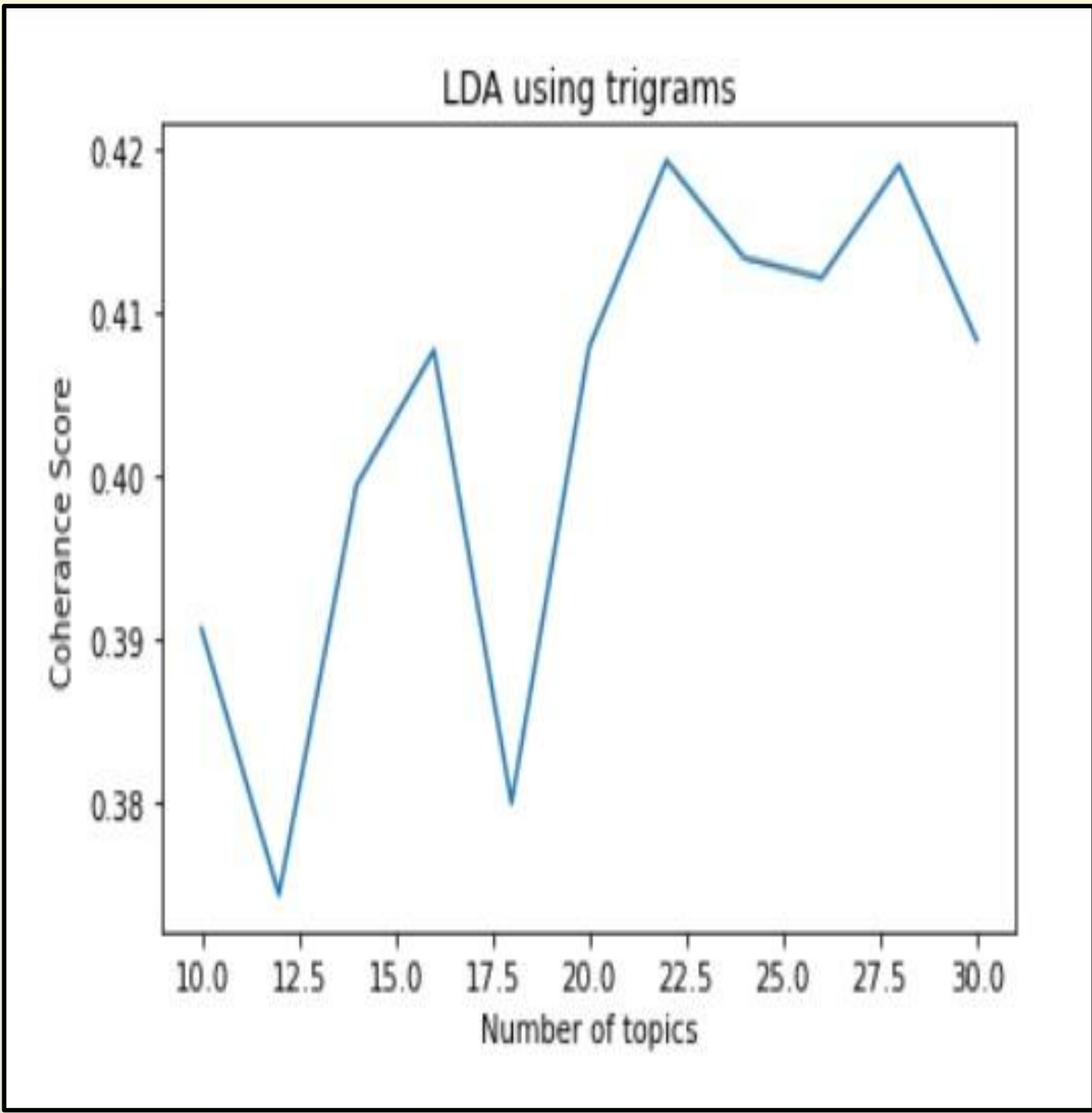
For recommending the project we implement LDA model. LDA is a form of unsupervised learning that views document as a bag of words.LDA works by first making a key assumption: the way document was generated was by picking a set of topics and then for each topic picking a set of words.And that various other topics modeling techniques like LIS, HDP, Tf-Idf Modeling are also applied. We have also applied K-means clustering on document vector of all the report text in our data-set. The graph besides shows the coherence score verses number of topics for LDA using trigrams.

Dataset

Dataset consist of 3266 projects having information such as university name, links of document which are links to the project reports made by student of other university, subjects corresponding to the area of project reports and document name which is project title. Data is scrated from website of different foreign universities.

Analysis

In topic modeling we are allocating topics to all the projects which are present in the dataset. LDA assigen topics to the bag of words or tf-idf words of documents.In figure below topics are assigned.



Methodology

Here we match the keywords of the topic with the users query and recommend the projets to the user that match with that topic . For matching project user get list of all the areas as a keyword in which it want to select and then list out all projects to the user.

Result

Models are judged on the basis of their coherance score as it should be between 0.3 to 0.6 other then this values will not sematically closer to the assigned topic words.

MODEL	SCORE
(Bag of Words)	
1. LDA	37%
2.LSI	36%
3.HDP	61%

Document Name	links
Diagnosing TMJ Arthritis	http://web.stanford.edu/class/cs341/project/Lo...
Predicting Risk of Breast Cancer Relapse from ...	http://cs229.stanford.edu/proj2019spr/report/5...
Classifying Leukemia Using Logistic Regression...	http://cs229.stanford.edu/proj2019spr/report/6...
Multi-omics Factorization Illustrates the Adde...	http://cs229.stanford.edu/proj2019spr/report/6...
Cardiovascular Disease Risk Prediction using EHRs	http://cs229.stanford.edu/proj2019spr/report/6...
Predicting Microculture Results for Optimized ...	http://cs229.stanford.edu/proj2019spr/report/8...
Predicting the Survivability of Breast Cancer ...	http://cs229.stanford.edu/proj2018/report/155.pdf
Painless Prognosis of Myasthenia Gravis using ...	http://cs229.stanford.edu/proj2018/report/166.pdf
Characterizing Data-Driven Disease Phenotypes ...	http://cs229.stanford.edu/proj2017/final-repor...
Predicting Diabetes Readmittance	http://cs229.stanford.edu/proj2017/final-repor...
Automated Semantic Segmentation of Volumetric ...	http://cs229.stanford.edu/proj2017/final-repor...
Cardiovascular disease prediction: a novel ris...	http://cs229.stanford.edu/proj2017/final-repor...
Optum: Investigating Links between the Immune ...	http://cs229.stanford.edu/proj2017/final-repor...
AKI Prediction	http://cs229.stanford.edu/proj2017/final-repor...

Conclusion

Among LDA, HDP, LSI models. LDA assigen topics correctly to the projects with an average coherence score with having bag of words as a document.