2019 PilotCity Mathematics Clinic

Extracting Educational Priorities for Workforce Innovation

Stakeholders

PilotCity

- Startup in San Leandro, CA
- Connects classrooms to work-based learning opportunities and internships with global tech companies

ies INSTITUTE OF EDUCATION SCIENCES

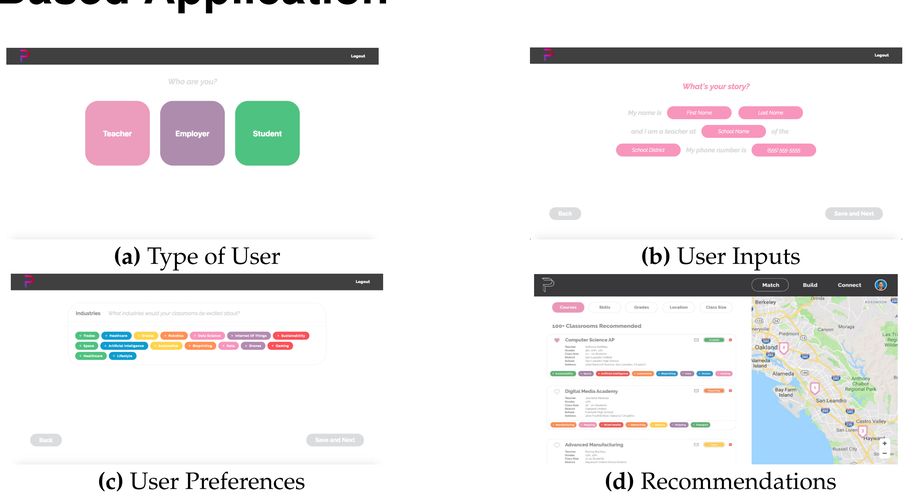
- R&D branch of the U.S. DoED
- Extracts data-driven educational insight to inform educators, policy makers, and the public

Problem Statement

We were tasked with extracting **educational priorities** from **uncurated** data sources and developing a **recommender system** to **automate** PilotCity's program delivery. Further responsibilities included building a **web-based application** for PilotCity users and a **topic modeling engine** to gain insight from unstructured data.

Deliverables

Web-Based Application



(a) Type of User **(b)** User Inputs

(c) User Preferences **(d)** Recommendations

Figure 1: Onboarding web application designed based on user interviews

Recommender System

- Scores compatibility between every employer and classroom

Inputs to the recommender system

Classroom	*Employer*
Course name, technologies, skills, industry preference	Industry, service, product, vision

- Semantic Word Embedder: The GloVe Model
  - Unsupervised learning algorithm producing vector representations of words, pre-trained on 2014 Wikipedia data
  - Cosine distance between vectors relates to semantic similarity

Topic Modeling Engine

- Built a pipeline that takes in a large corpus of texts/PDFs and returns summaries and visualizations of representative topics
- Experimented with algorithms such as LDA and NMF
- Applied these algorithms on 2000+ course syllabi scraped from Las Positas College in Livermore, CA

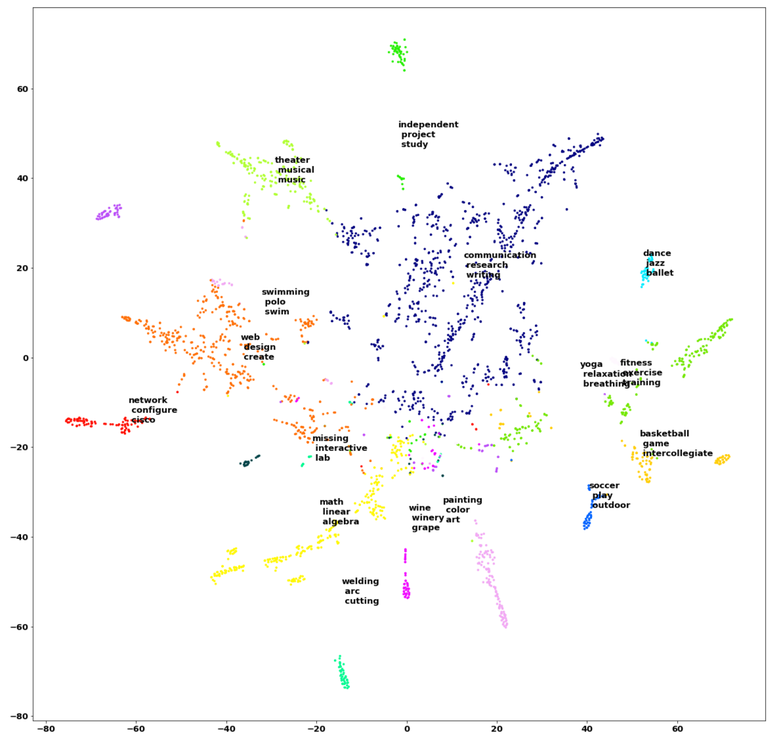


Figure 2: Visualization of All Las Positas Syllabi with 16 Topics using NMF

Conclusions

- We can extract educational priorities using a variety of NLP techniques.
- Our work can help institutions make data-driven decisions.
- Topic modeling can help improve PilotCity recommendations.

For Further Information

- Contact us at `mathclinic-pilotcity-l@g.hmc.edu`.
- View our code at: `www.github.com/hmcmathclinic/18-19-PilotCity-Code`.

References

- Greene, Derek. Parameter Selection for NMF. GitHub, 27 Sept. 2017.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation." 2014.
- Las Positas College. Las Positas College CurricUNET. 2019.
- Pedregosa et al. Scikit-learn: Machine Learning in Python, JMLR 12, 2011.
- Radim Rehurek and Petr Sojka. "Software Framework for Topic Modelling with Large Corpora." 2010.
- W, Shuai. Topic Modeling and t-SNE Visualization. Shuai's AI Data Blog, 2016.

Acknowledgments

We greatly appreciate Derick Lee, the PilotCity local team, Dr. Talithia Williams, Dr. Weiqing Gu, and IES for supporting our project.

Team Members

Madison Hobbs (PM), Aanya Alwani, Dominique Macias, Jean Adedze, Xuming Liang

Faculty Advisor Dr. Talithia Williams
Liaison Derick Lee, PilotCity