

Identifying Subject Matter Experts

Extending Author Topic Modeling

Philip Robinson

Presented to OCIO
NASA - Jet Propulsion Lab

August 28, 2018



Introduction - Philip Robinson

Computer Science MSc at Oregon Health and Science

Thanks to my mentor Ian Colwell, from the OCIO (1762)



- probabilistic programming
- language processing
- image processing
- audio processing
- stem education
- environmental sciences
- ★ information retrieval



Presentation Overview

- 1 Introduction
- 2 Problem Description
- 3 Proposed Approach
- 4 Results
- 5 Conclusion



Problem Description

Our customers, Office of Safety and Mission Success (5x), are interested in identifying experts for resolving anomaly reports in the Problem Reporting System (PRS)

OCIO (17x) has been previously asked for subject matter expert identification systems and document similarity tools, so show particular interest in solutions that provide similarity metrics and can generalize to other teams and corpora.

- A-Team hierarchical frequent item set expert exploration tool
- TechConnect self reported skills host
- Gateway Profiles



Motivating Story

- Domain experts are lost between projects
- Domain experts are often coupled to a single project
- Very few candidates resolve the majority of tickets
- Expert discoverability
- Load balancing employees
- Identification of knowledge gaps



Objective

- Assign & Resolve anomalies quicker
- Support queries for expert discovery
- Find employees with similar domain expertise



Data Provided for Internship

- Problem Reporting System (Anomalies)
 - Problem Failure Report (PFR)
 - Developmental Problem Failure Report (DPFR)
 - Incident Surprise Anomaly (ISA)
-
- Free Text
 - Title
 - Description
 - Experts
 - Responsible Editor
 - Assignee



Approach

Author-Topic-Modeling

- Interpret doc as Bag-of-Words¹
- Model/Fit topics as mixture of words
- Author & document are projected into topic-space
- Measure distance from author to document

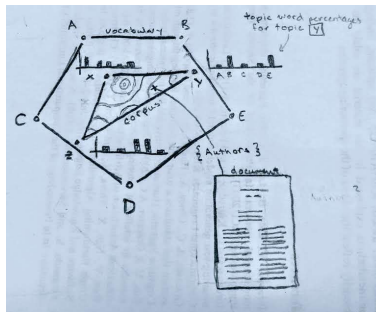


Figure: Latent Dirichlet Allocation

$T(x) = \text{Project } x \text{ into topic-space}$

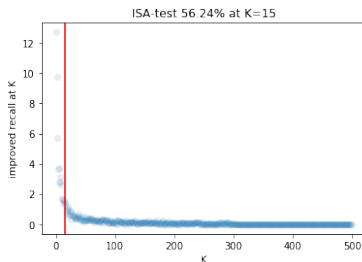
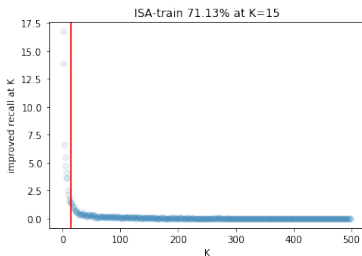
$$R_d = \underset{a \in A}{\text{argsort}} \{ \text{Distance}(T(a), T(d)) \}$$

¹equivalent to multinomial over vocabulary

Ranking

Seems to work

Given Authors/Experts in a “Topic Space” and a mapping from document to “Topic Space”, we can rank experts for a document.



K is cutoff for suggested candidates



Receiving Candidates

Results begin at 4 words

It is possible to get interesting results at a document length of 4 words, however it is hard to know why these results are interesting. This is an example of directly searching for experts.

'gimbal drive motor friction'

	Name	Title	Organization
0	Amanda Donner	Mission Assurance Manager	5150
1	John Trager	NaN	337C
2	Mathew Keuneke	Product Delivery Manager	397A
3	Jessica Bowles-Martinez	Systems Engineer	313G
4	NaN	NaN	NaN

'rtg temperature drive curiosity capacity'

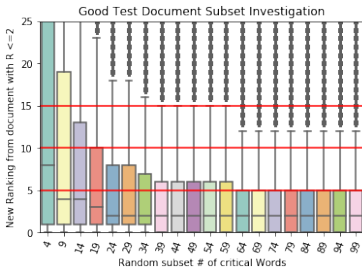
	Name	Title	Organization
0	John Rakiewicz	NaN	NaN
1	Angela Dorsey	Technologist	335S
2	Otfrid Liepack	Deputy System Manager	394G
3	Mohammad Shahabuddin	Flight Software Engineer	348D
4	Megan Lin	Delivery Manager	397S



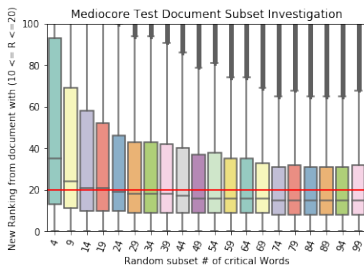
How does word count effect recall?

Best results at 30 words

We are interested in understanding how much text is required to inform our model prediction. For these plots, we randomly subset texts for known ticket-expert pairs and observe the expert's new ranking.



Expert found in top 2
24 critical words



expert found in 10-20 range
29 critical words



Conclusion

- This looks like it works
- Not computationally or socially prohibitive
- Motivating story to lift limitations on data access
- Further investigation is needed to insure best expert fields
- User interface is required for integration into workflows



Internship Notes

JPL is fun

I had never implemented a recommender system, nor used topic modeling in a project prior to this task. I will be leaving JPL with a holistic understanding of topic modeling, and a much better understanding of recommender systems.

Having multiple interns work with the same data helps. We were able to work through data difficulties together, and share results.



Thanks

I received a lot of input in for this project from my team. Their insight, especially, bridged the gap from academic to practical evaluation of models, and I sincerely appreciate their contributions.

- Ian Colwell
 - Valentinos Constantinou
 - Jerry Chen
 - Leslie Callum
 - Bruce Waggoner
 - Harald Schone
 - Chris Mattmann
- et. al.

