Frameworks

Trish Goedecke, Preston Provins, Jerry Duncan

Quick Recap

We know you forgot.

Choosing a JavaScript frameworks for a new project can be a challenging process.

- There seems to be a new handful gaining popularity every year or so
- Previous ones don't seem to lose steam either.

Some of the most popular frameworks:

- React.js (maintained by Facebook)
- Angular.js (supported by Google)
- Ember.js (opinionated best practices)
- Backbone.js
- jQuery
- VueJS

Objectives

1) Collect framework related resources
- Identify Projects that use the frameworks.
- Collect dependencies of the frameworks.
- Attain questions and answers of frameworks on StackOverflow.
- Retrieve Issues and Responses of each framework.

2) Build a website to display the data that we had collected.

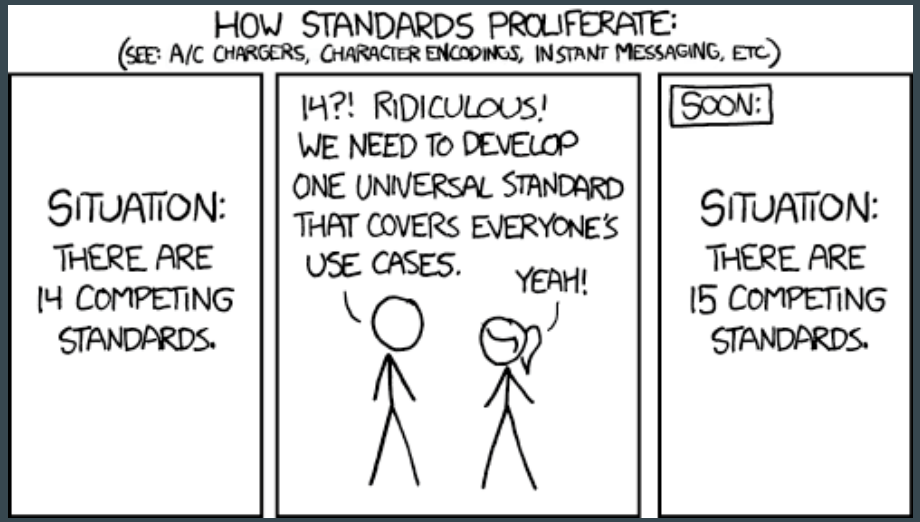
3) Derive predictors on collected data and train a model to make predictions.

SeedsJS – A Sails Ember Ember-Data Semantic-UI new full stack JavaScript framework in beta (github.com)

submitted 3 years ago by henk53

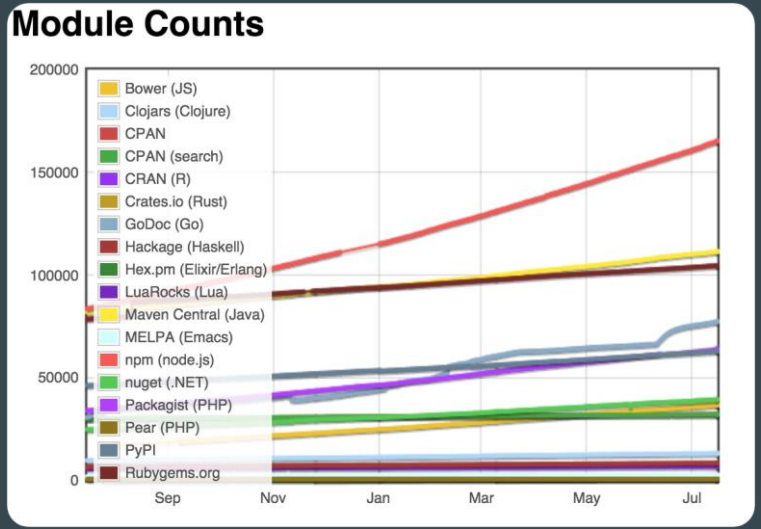
9 comments share save hide report hide all child comments

Motivations.



Motivations Cont.

Jokes aside, frameworks are developed and used in projects haphazardly. With so many frameworks being created and maintained it is hard to tell which frameworks will be relevant years down the road.

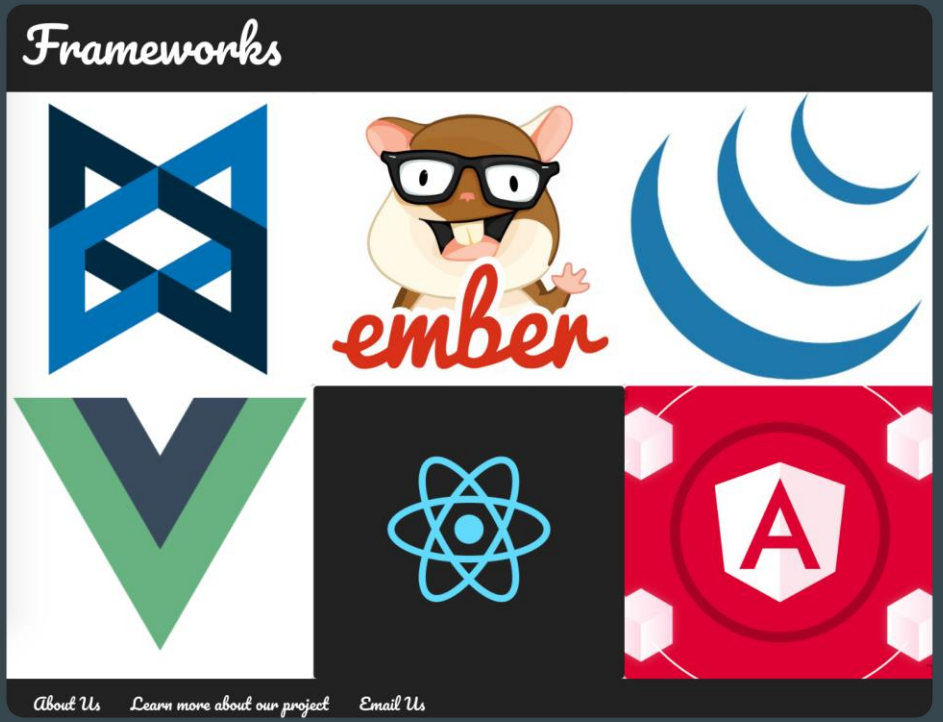


Check us out on the Web!

The data that we collected is easily accessible on the web.

If you would like to see the visuals (and amazing web design) please follow us as we go to:

https://fdac-frameworks.herokuapp.com/



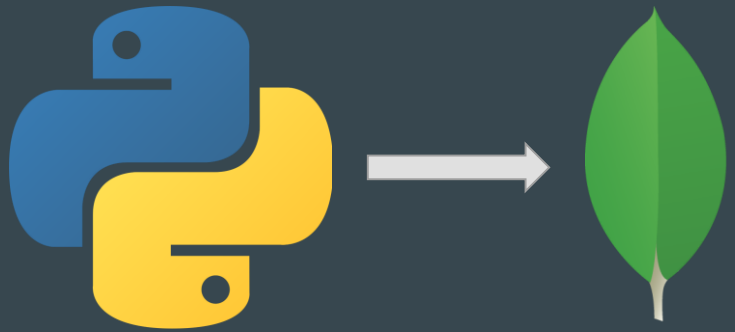
Collecting the Data

Data attained from GitHub and StackOverflow:

- GitHub Issues required stripping information from repository pages.
  - Stored in .json files, not kept in database.
- StackOverflow had a poorly designed SQL Database, but made dumps that were .xml formatted
  - A python script was made to take XML data and place inside MongoDB
  - I should probably ask StackOverflow to pay me for my troubles.

Collection Tools:

- Python was used solely.
- MongoDB was our best option for Database
  - Most data came in incomplete.
  - StackOverflow was SQL, but poorly structured.



LDA hypothetical document generation

Representation $\sim$

1. A corpus consists of D documents
2. A document consists of N words from vocabulary of size V
3. Each word w_n is chosen from one of K topics.

LDA hypothetical document generation

1. Number of words (N) for document selected from Poisson, parameter ξ.
2. Multinomial probability vector θ length K selected from Dirichlet, parameter α
3. The nth word is chosen from the following process:

   - Choose a topic, Z_n, from $Mult(1, \theta)$
   - Choose a $V \times 1$ vector ϕ from $Dir(\beta_{Z_n})$
   - Choose a word, w_n from $Mult(1, \phi)$

JS Frameworks: Growth Rates and Topics

Growth rates in downloads of JavaScript frameworks, September 2017-September 2018

With 4 most highly correlated topics, from 20-topic LDA topic model

Top 5 terms of each topic shown

JS Framework	Growth	Topic			
		1	10	12	17
angular	-0.05	0.03	0.05	0.02	0.02
backbone	-0.08	0.11	0.04	0.02	0.06
ember	0.54	0.02	0.10	0.01	0.01
jquery	0.09	0.04	0.01	0.01	0.05
react	0.57	0.02	0.01	0.10	0.02
vue	0.88	0.02	0.18	0.12	0.01
Correlation		-0.66	0.71	0.79	-0.76

Topics are generated from GitHub Issue documents.

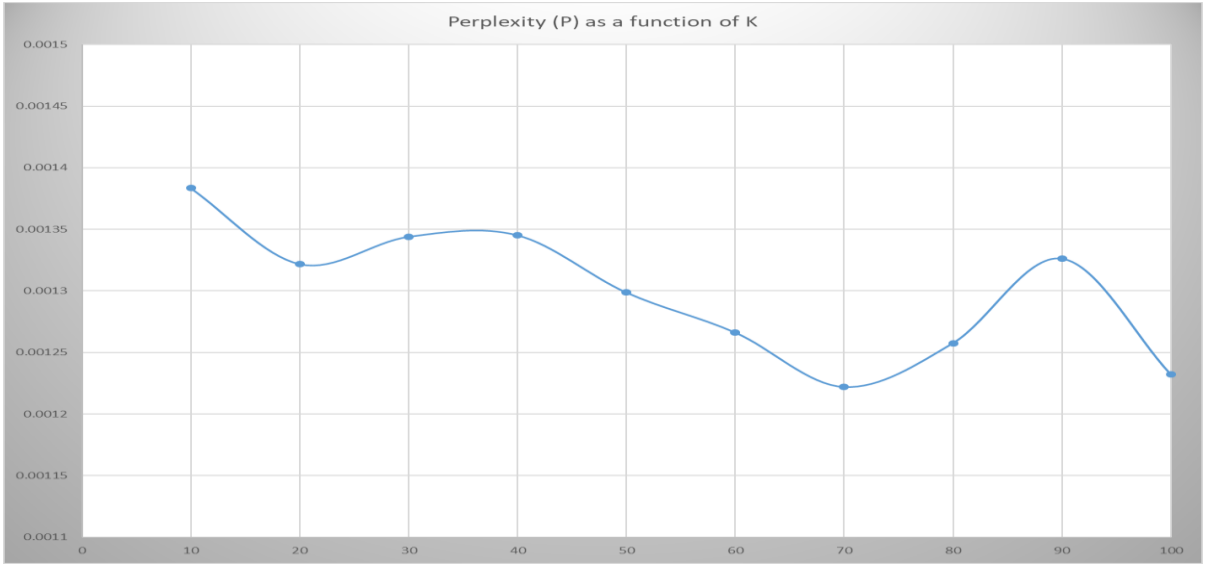
backbone	data	component	file
jquery	date	function	server
would	property	state	patch
like	reproduction	child	data
support	expected	data	error

Future Work

Taking this Idea to the Next Semester

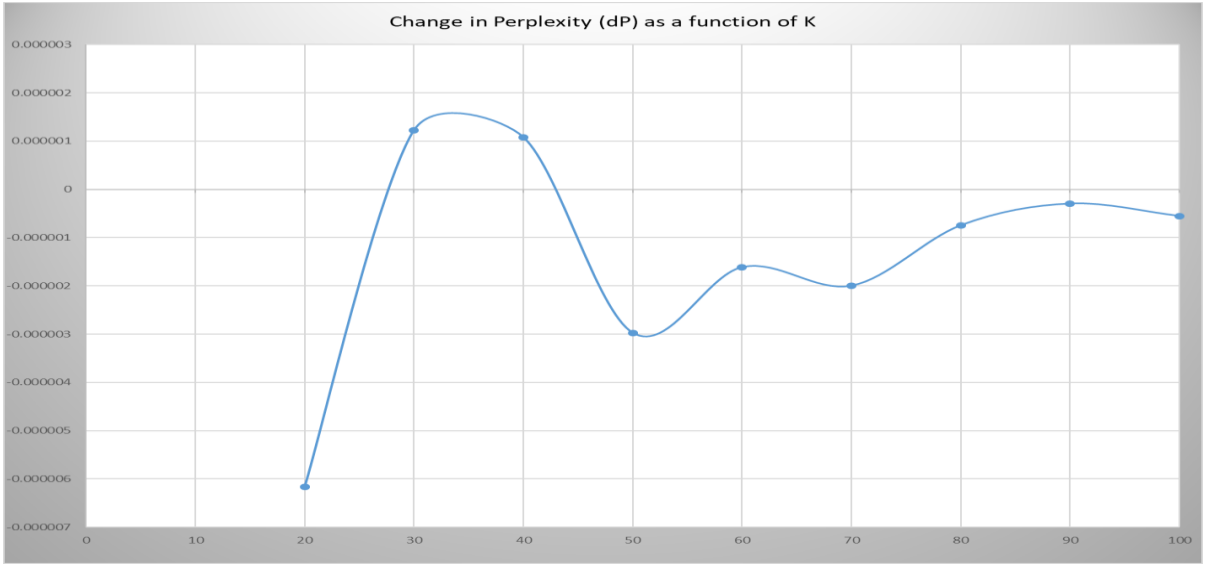
Unfortunately, we did not have time to get to the prediction model; however, the data has all been collected and set up so that next semester in Dr. Mockus' Evidence Engineering we can continue working on the model.

Perplexity (P)



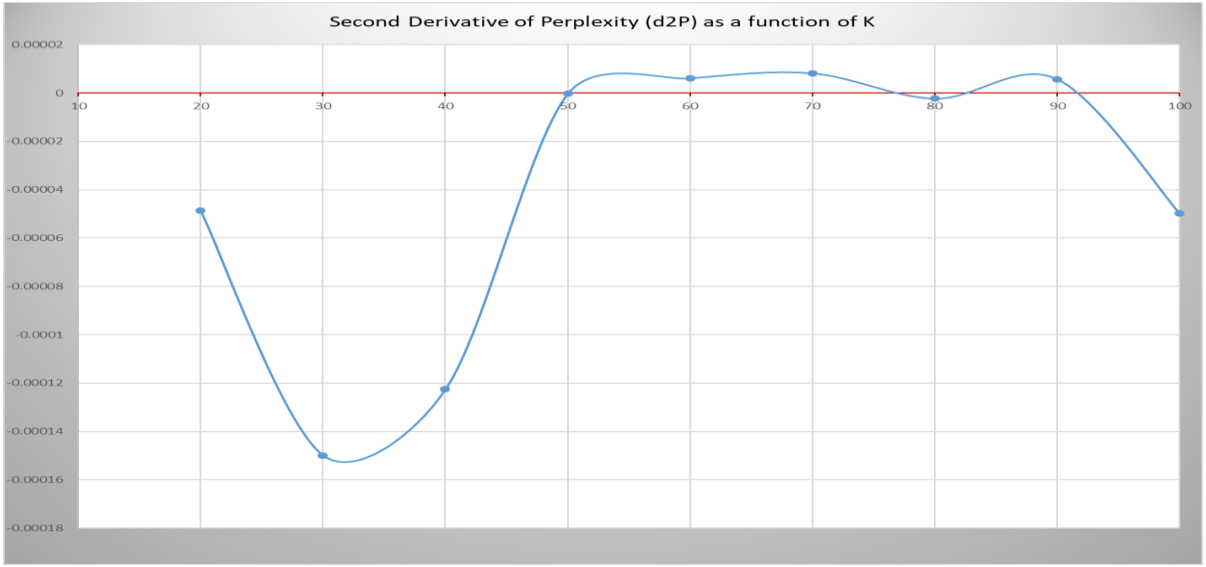
[1] Blei et al., 2003

Change in perplexity (dP)

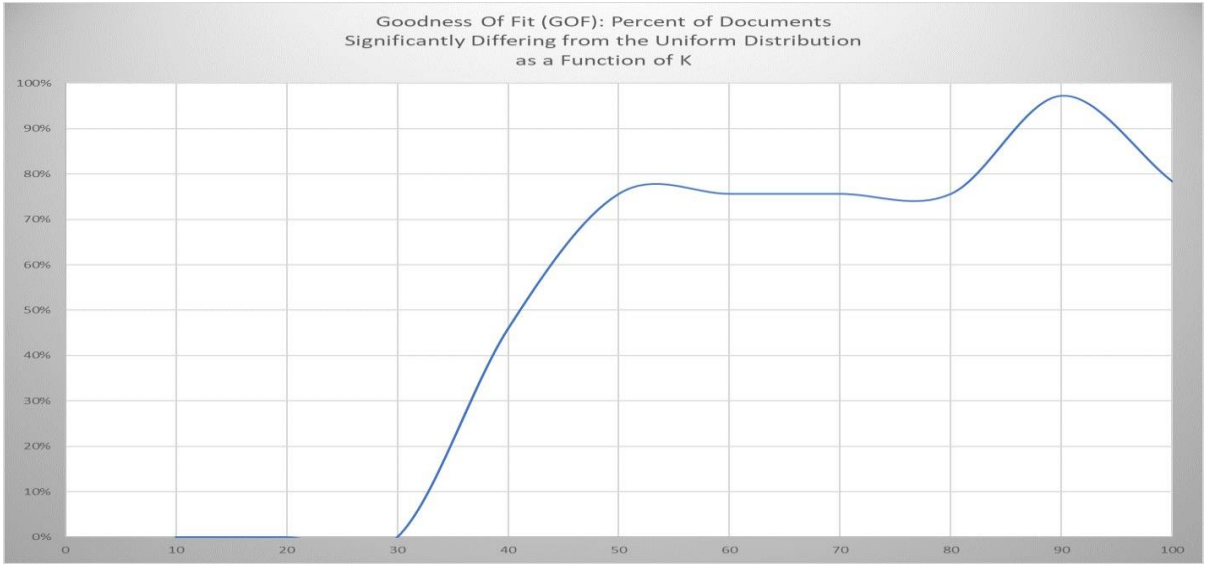


[1] Zhao et al., 2015

2nd derivative of perplexity (d2P)



Goodness Of Fit (GOF)



[1] Bowman, Chen and George (in press)

Pointwise Mutual Information (PMI)

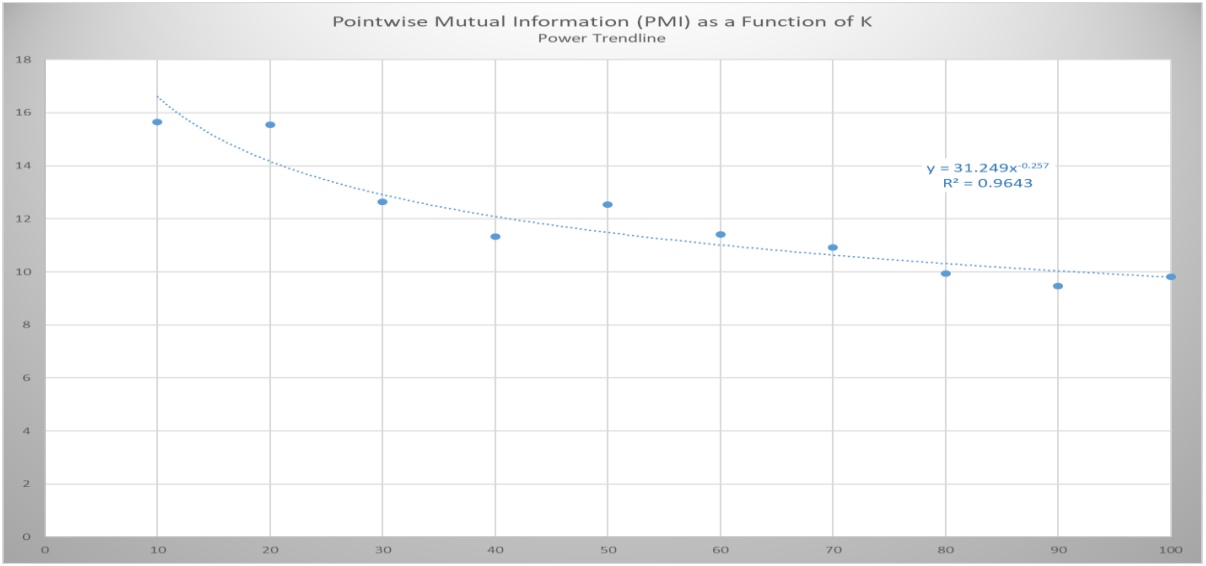
$$\mathrm{PMI}(x, y) = \log \frac{p(x, y)}{p(x)p(y)}$$

Newman et al. (2010) Extrinsic to modeling process

Corroborated by human judgment

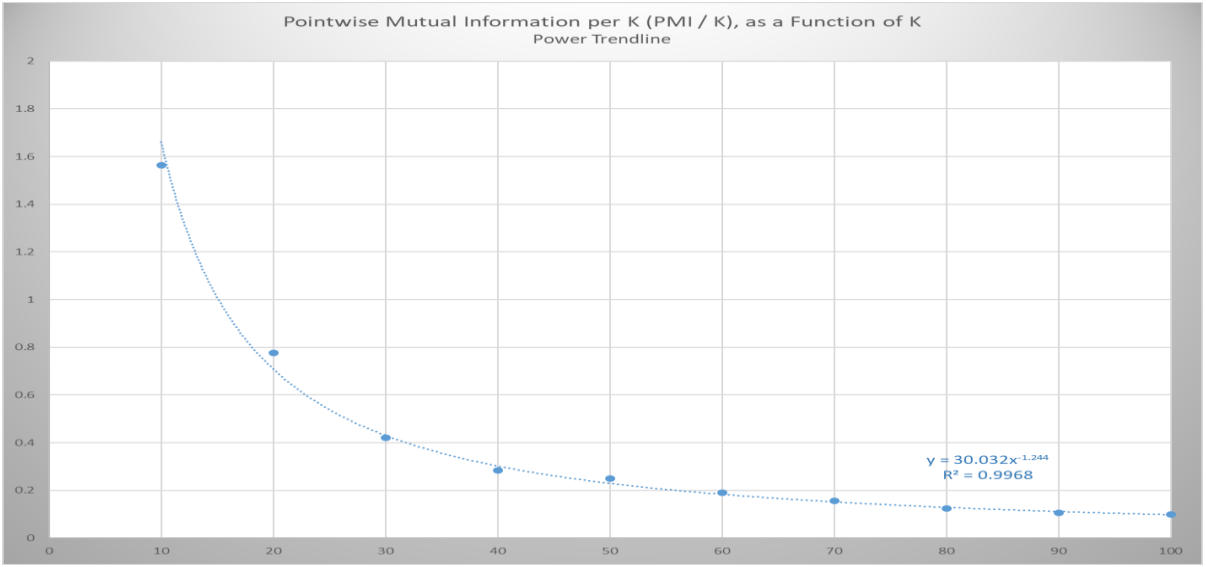
[1] Image credit: http://lintool.github.io/UMD-courses.html

Pointwise Mutual Information (PMI)

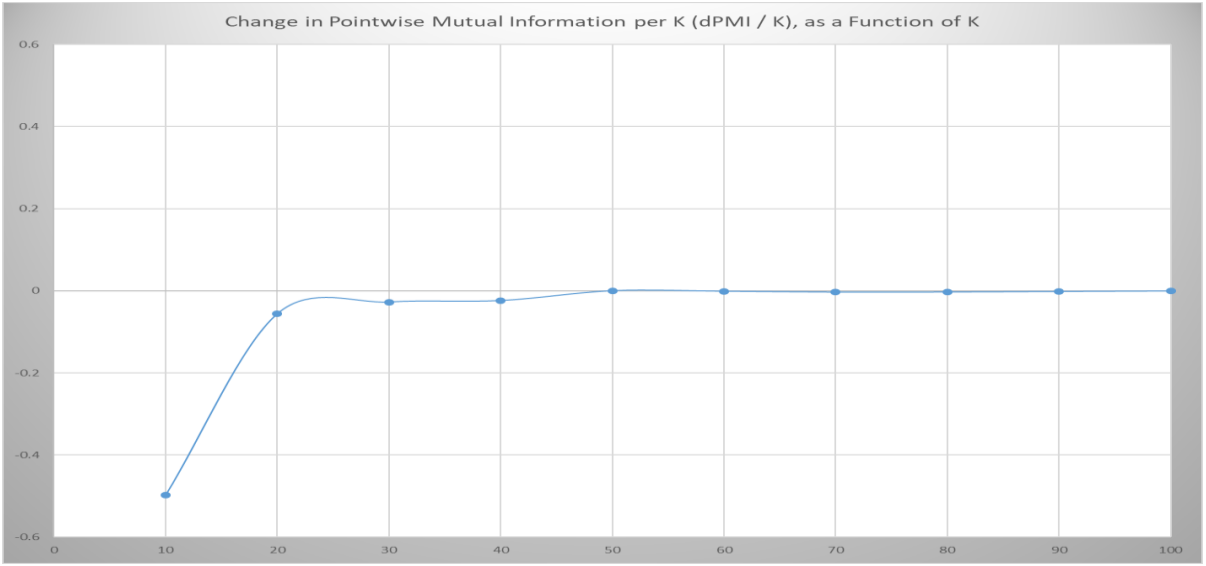


[1] Naraula et al., 2013

Pointwise Mutual Information per K (PMI / K)



Change in Pointwise Mutual Information per K (PMI / K)



Conclusion: Optimization method comparisons

- Methods of optimizing K do not converge
- Different optimization methods will be appropriate for differing purposes
- Judgment is called for

Questions?

Thank you!