Application of Text Analysis to Quality Control of Human Resources Documents

Thor D. Osborn

SYSTEMS ANALYSIS & DECISION SUPPORT
INSIGHTFUL. OBJECTIVE. INDEPENDENT.

Info.sandia.gov/sysanalysis/

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND2016-8120 C

Motivation for Quality Control of Human Resources Documents

SYSTEMS ANALYSIS & DECISION SUPPORT

- The job description set (JDS) of most large organizations sits at the nexus of many strategic, operational, and individual decisions
- “Equal pay for equal work” is protected by law in the U.S. and elsewhere
- Confounded jobs with differing pay scales may be contested as equivalent, incurring substantial risks and liabilities
- Recent legal and policy shifts have led to increasing accountability for organizational improprieties among top leadership, regardless of their direct involvement
- Continuous improvement of the JDS mitigates risk over time and signals positive intent

Analysis Outline

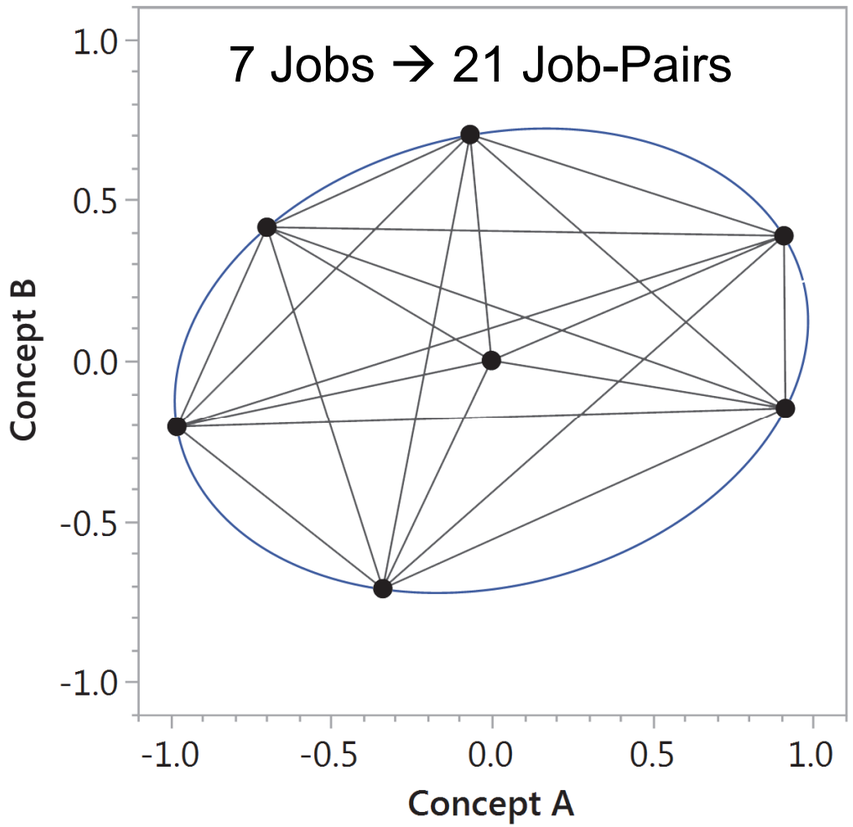
- Establish objective criteria for:
  - Document set differentiability
  - Flagging confounded (close) job pairs for examination
- Determine appropriate term weighting method
- Demonstrate
  - Impact of adding poorly differentiated content to the JDS
  - JDS adjustment to improve job family classification performance
  - JDS adjustment to improve overall differentiability

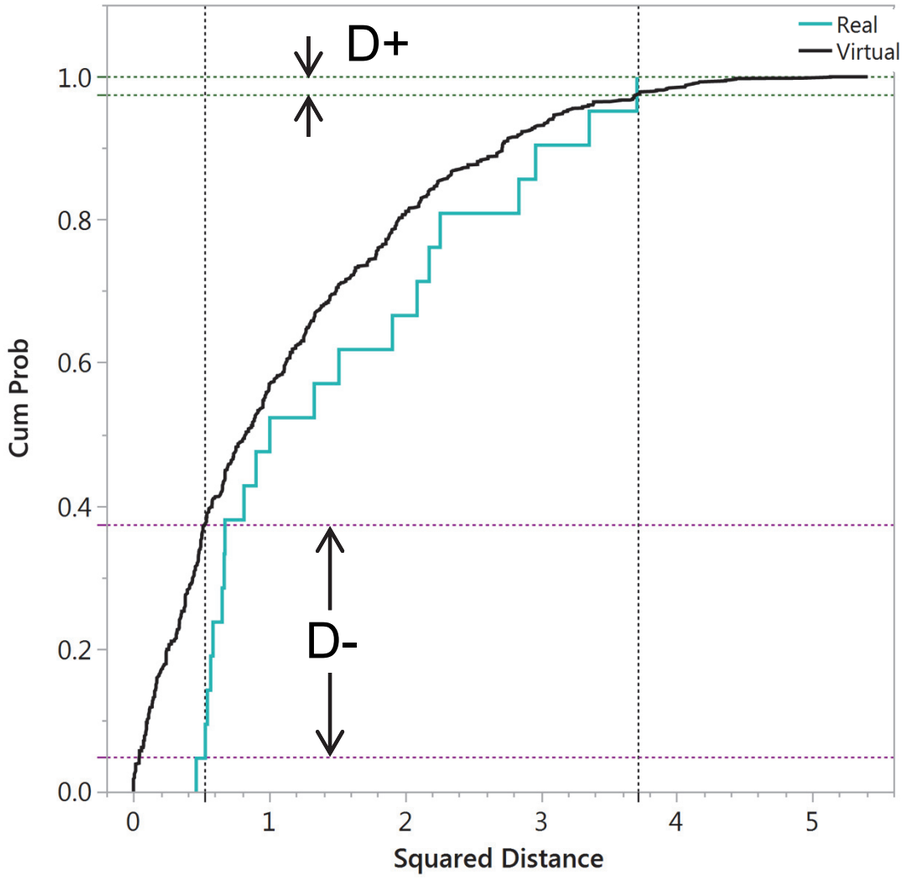
Experimental Job Description Set

- Notional hospital system (NHS)
- Represents the job description set necessary to operate a regional hospital system
- Not modeled after any specific real organization
- 250 job descriptions
- 15 SOC major groups

Major Group	Description	N
11	Management Occupations	32
13	Business and Financial Operations Occupations	24
15	Computer and Mathematical Occupations	7
17	Architecture and Engineering Occupations	1
19	Life, Physical, and Social Science Occupations	9
21	Community and Social Service Occupations	5
23	Legal Occupations	2
25	Education, Training, and Library Occupations	1
27	Arts, Design, Entertainment, Sports, and Media Occupations	5
29	Healthcare Practitioners and Technical Occupations	97
31	Healthcare Support Occupations	9
33	Protective Service Occupations	1
35	Food Preparation and Serving Related Occupations	11
37	Building and Grounds Cleaning and Maintenance Occupations	6
43	Office and Administrative Support Occupations	31
49	Installation, Maintenance, and Repair Occupations	3
51	Production Occupations	4
53	Transportation and Material Moving Occupations	2

Example of Job Differentiation in the Design Concept Space





Kolmogorov-Smirnov Asymptotic Test

KS	KSa	D=max|F1-F2|	Prob > D	D+=max(F1-F2)	Prob > D+	D-=max(F2-F1)	Prob > D-
0.0211254	1.4894543	0.3257143	0.0237*	0.0248485	0.9745	0.3257143	0.0118*

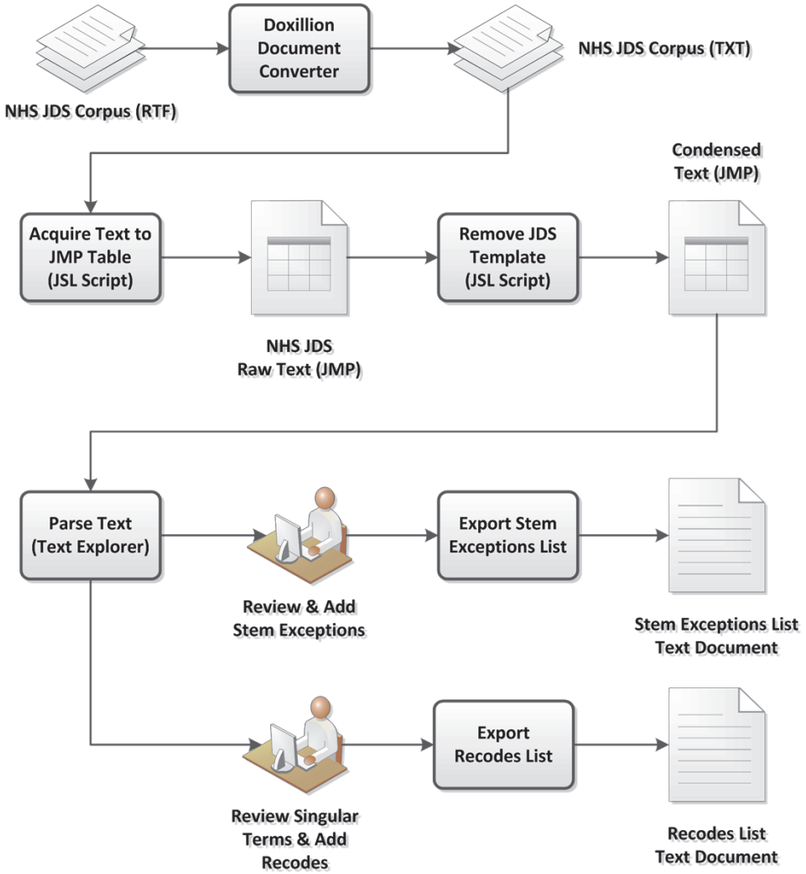
- Random virtual jobs produced by sampling Concept vectors with replacement
- Virtual job-pair distribution represents random "design" within the Design Concept Space (DCS)
- Real job-pair distribution expected to exhibit significantly better separation
- Two-sample KS test indicates real and virtual job-pair samples drawn from different distributions

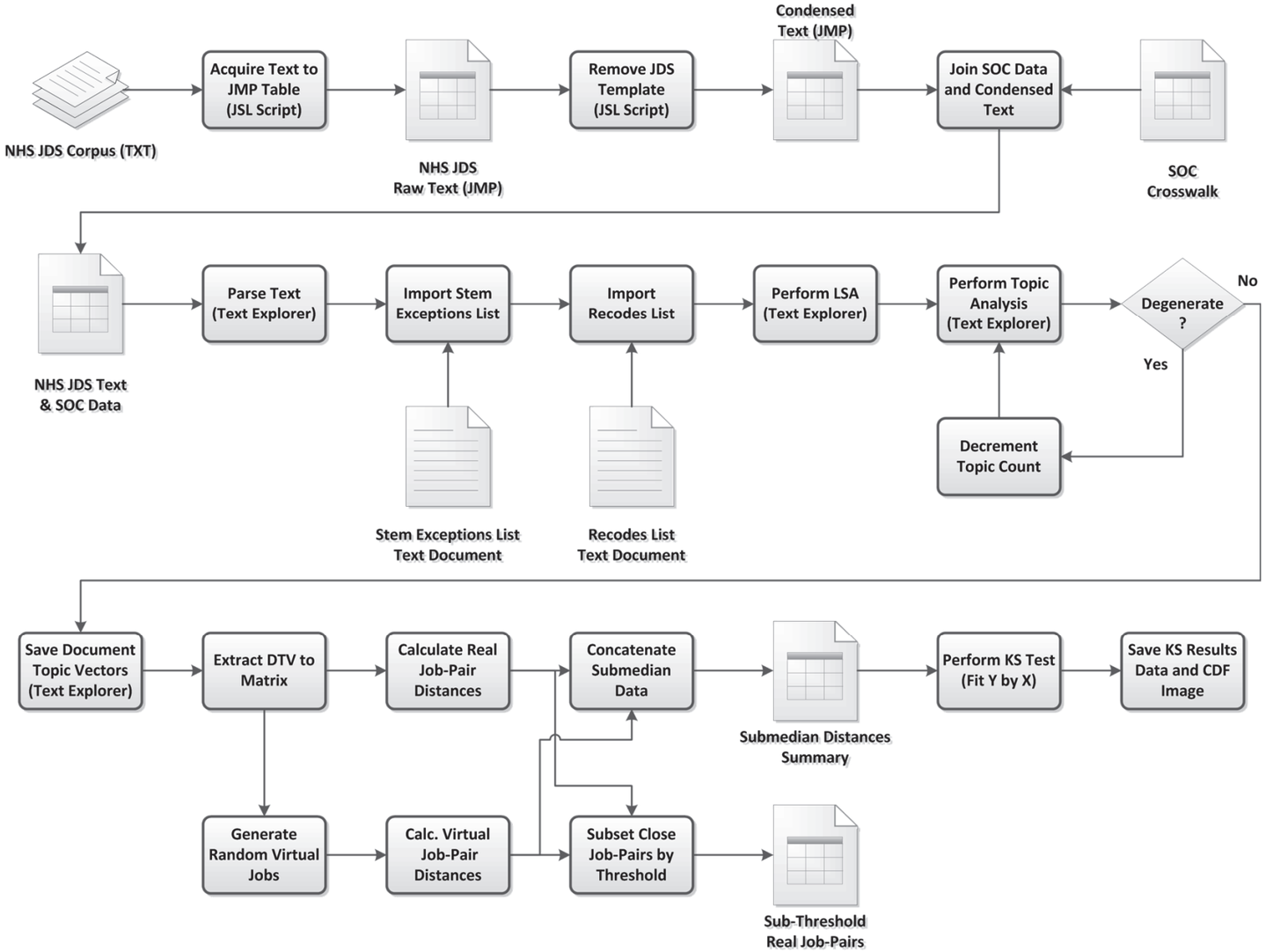
Evaluation Process Steps

SYSTEMS ANALYSIS & DECISION SUPPORT

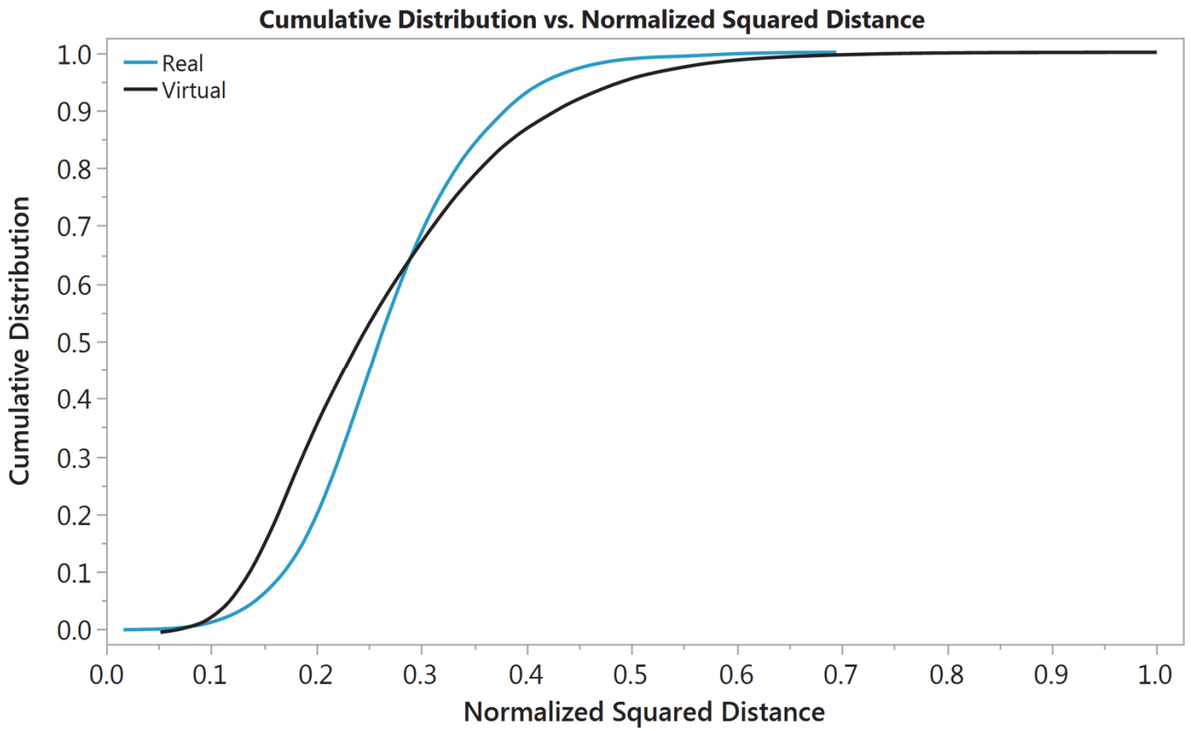
Main (Base Pool) Process →

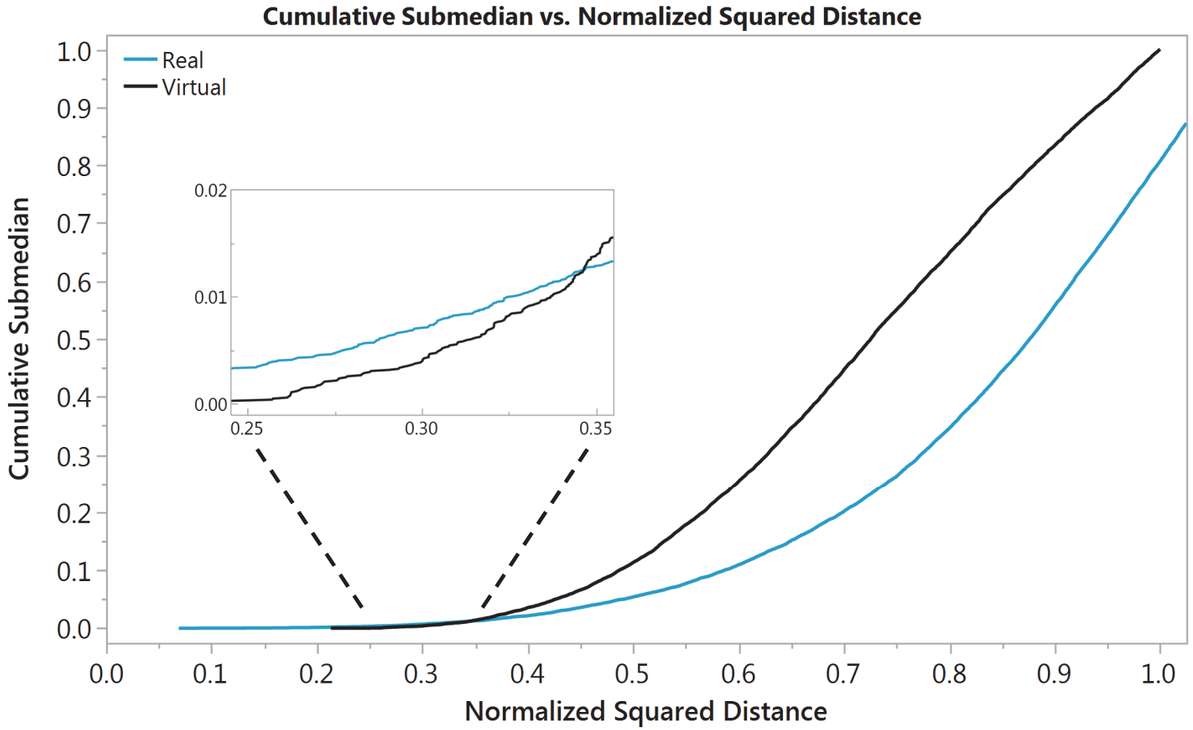
Initial Data Preparation ↓





Evaluation Focuses on Submedian Separation Distribution





- Complete NHS JDS used – 250 jobs, 31125 job pairs
- Real job-pairs feature tighter distribution – more evenly separated within Design Concept Space

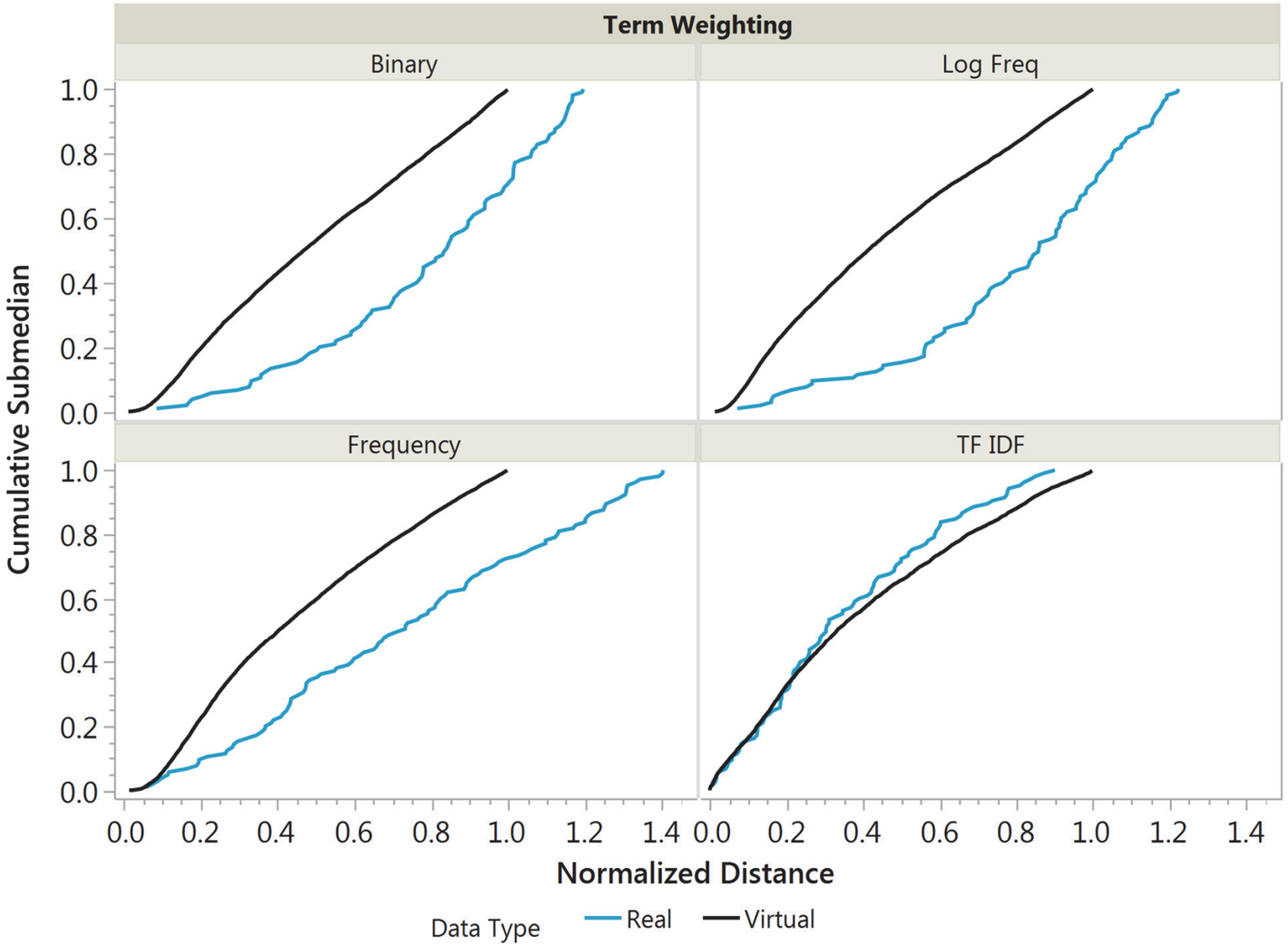
- Differentiability governed by separation of close (submedian) job-pairs
- Elevated ‘tail’ of real distribution responsible for Kolmogorov-Smirnov one-sided D+ > 0

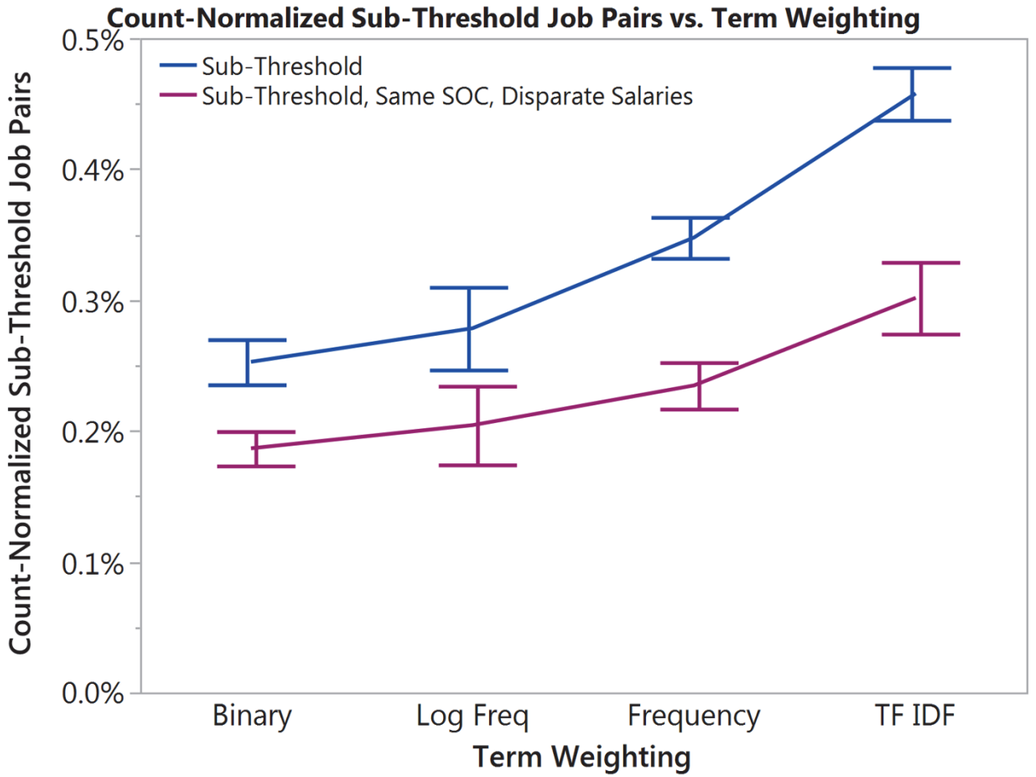
Live Demonstration 1

SYSTEMS ANALYSIS & DECISION SUPPORT

- Estimate variability of warning threshold
  - Show submedian virtual job-pair squared distance distribution for JDS corpus
  - Show bootstrap estimate of the 0.1% quantile (warning threshold stability)
  - Flags between 56 and 63 job descriptions (95% confidence interval)

Term Weighting Method Selection

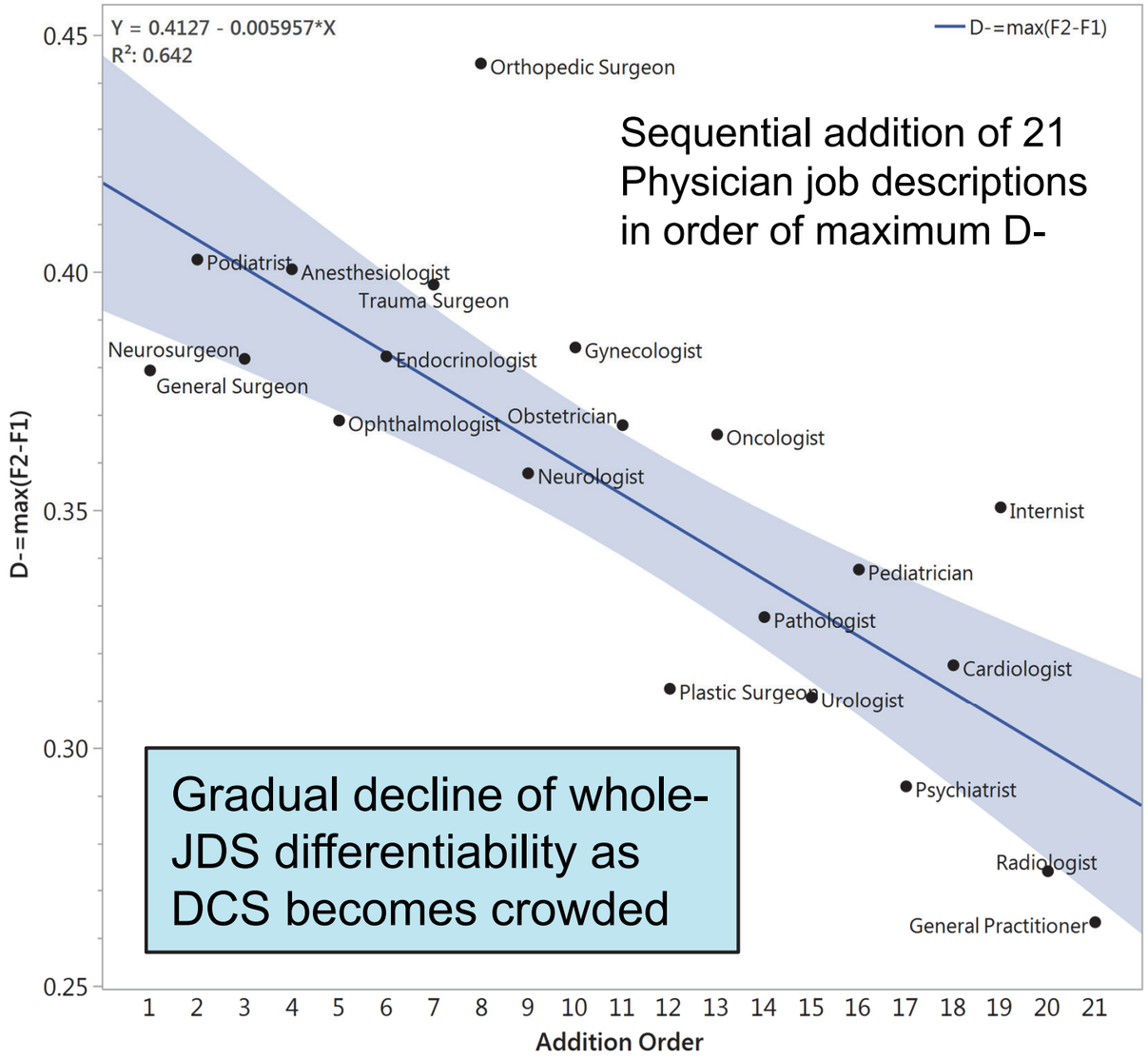




- TF IDF inappropriate and unstable
- Frequency inappropriate
- Log Freq plausible but least consistent

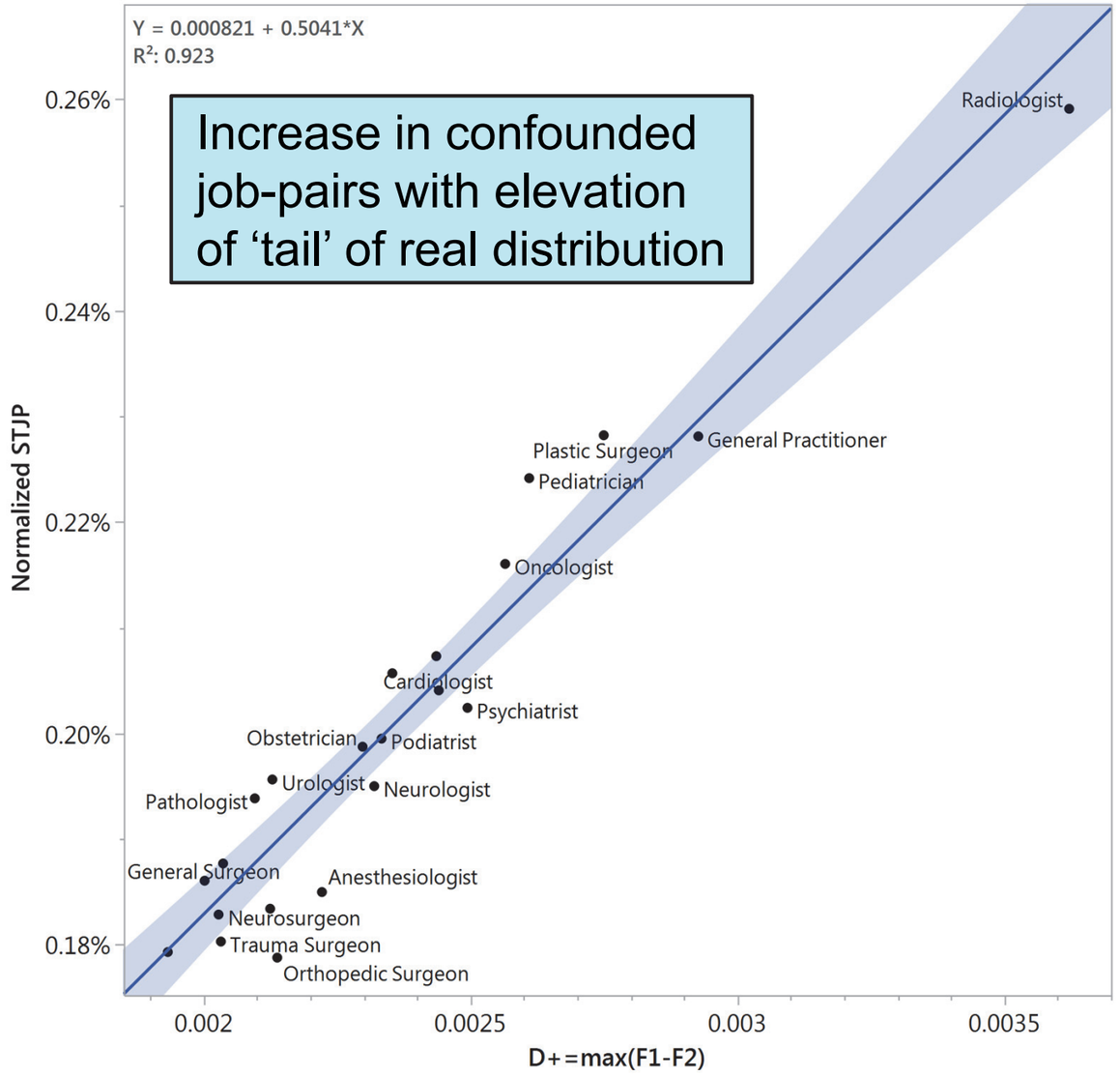
Binary method chosen

Addition of Poorly Differentiated Jobs Confounds JDS



Sequential addition of 21 Physician job descriptions in order of maximum D-

Gradual decline of whole-JDS differentiability as DCS becomes crowded



Increase in confounded job-pairs with elevation of 'tail' of real distribution

Job Family Classification

SYSTEMS ANALYSIS & DECISION SUPPORT

Discriminant platform "interesting rows":

Row	Actual	SqDist(Actual)	Prob(Actual)	-Log(Prob)		Predicted	Prob(Pred)	Others
98	29	126.065	0.0001	9.417	*	31	0.9999	
109	29	84.638	0.4614	0.773	*	31	0.5386	
159	11	131.358	0.4115	0.888	*	13	0.5885	
194	27	90.006	0.0874	2.437	*	13	0.9126	

Restated in terms of SOC major group:

Job Title	Actual SOC	Code	Predicted SOC	Code
Orthopedic Assistant Pharmacy Technician	Healthcare Practitioners and Technical Occupations	29	Healthcare Support Occupations	31
Compliance Director	Management Occupations	11	Business and Financial Operations Occupations	13
Grant Writer	Arts, Design, Entertainment, Sports, and Media Occupations	27		

- Four (4) job family classification errors
- Compliance Director had been improperly assigned to SOC Major Group 11
- Other job descriptions augmented with additional detail
- Second classification analysis: four errors at left corrected, but Janitor reclassified to Healthcare Support Occupations

The DCS will evolve with every change to the JDS content, potentially altering classification of jobs other than those updated

Live Demonstration 2

- Discriminant analysis of JDS corpus augmented with document-topic vectors to show classification of jobs by SOC major group (job family)

Repairing Job Family Classification

SYSTEMS ANALYSIS & DECISION SUPPORT

- Janitor is properly found in SOC Major Group 37: Building and Grounds Cleaning and Maintenance Occupations
- Greatest topical divergences of Janitor position from SOC Major Group 37 shown at right
- Adjusted Janitor job description:
  - Added references to "wastebaskets, "waxes," and "polishes"
  - Added phrase "such as replacing light bulbs"
  - Added sentence "Must be comfortable using ordinary hand tools to make minor repairs and adjustments to building infrastructure and equipment."
- Zero classification errors after adjustment

Topic56 (+2.3)	
Term	**Score**
walk· the employe·	0.15605
climb or balance	0.13598
replac·	0.13078
mechanical part· fumes	0.13029
hall·	0.13001
expos· to move·	0.12412
distanc· vision	0.12026
equipment polishes metalwork	0.11440
polishes floor· clean·	0.11440
sweeps scrubs waxes	0.11440
rugs carpets upholstered	0.11440
door· panel·	0.11440
empti· and clean·	0.11440
sills empti· wastebaskets	0.11440
ashtrays transport· trash	0.11440
dusts furniture	0.11440
woodwork wash· window·	0.11440
furniture and draperies	0.11440

Topic32 (+1.8)	
Term	**Score**
need·	0.13776
mechanical part·	0.11547
design	0.11233
schedul·	0.10940
follow· duti· personally	0.10864
expos· to move·	0.10393
subordin· supervisor·	0.10311
worker·	0.10242
steam	0.10184
designed	0.10109
plan· develops	0.09961
fund·	0.09950
establishment by perform·	0.09887
fabric·	0.09824
tool·	0.09781
review·	0.09764
exist·	0.09759
superintendent	0.09727

Topic20 (+1.4)	
Term	**Score**
equipment	0.1513
pounds frequent· lift·	0.1373
direct·	-0.1314
must regular· lift·	0.1229
damag·	0.1154
maintain·	0.1052
water·	0.1011
balance and stoop	0.0994
kneel crouch	0.0993
prevent·	0.0979
improv·	0.0977
remov·	0.0973
beds	0.0938
equipment high school·	0.0905
diploma or general·	0.0905
educ· degree	0.0905

Topical divergence of a job description from target category may suggest appropriate adjustments for improving classification

General Improvement of JDS Differentiation

SYSTEMS ANALYSIS & DECISION SUPPORT

Closest six job descriptions before revisions:

Distance	Job Title A	Job Title B	Salary Disparity	Same SOC
10.0	Ophthalmic Technician	Orthoptist	1.17	Yes
10.2	Cafeteria Attendant	Counter Supply Worker	1.00	Yes
15.0	Cytotechnologist	Histotechnologist	1.00	Yes
16.5	Cardiopulmonary Technologist	Electromyographic Technician	1.33	Yes
18.9	Ophthalmic Technician	Optometric Assistant	1.00	Yes
19.4	Optometric Assistant	Orthoptist	1.17	Yes

Closest six job descriptions after revisions:

Distance	Job Title A	Job Title B	Salary Disparity	Same SOC
20.1	Hospital Chaplain	Corporate Lawyer	2.62	No
21.4	Chief Cardiopulmonary Technologist	Infection Control Nurse	1.23	Yes
21.5	Assistant Director of Development	Director of Corporate Relations	1.07	Yes
21.6	Dentist	Oral & Maxillofacial Surgeon	1.94	Yes
22.8	Chief Cardiopulmonary Technologist	Chief Dietitian	1.06	Yes
23.0	Geneticist	Microbiologist	1.11	Yes

- Nine most confounded job descriptions chosen for improvement example
- Before revisions:
  - 87 job-pairs below threshold separation
  - 63 job descriptions involved
- After revisions:
  - 63 job-pairs below threshold separation
  - 50 job descriptions involved
- Revisions made using O*NET job-related task information

Impacts of improvement efforts are readily monitored and visualized

Concluding Remarks

SYSTEMS ANALYSIS & DECISION SUPPORT

- Text analysis approach shows promise as a tool for improving management of organizational job description sets
- Approach guides improvement efforts by:
  - Exposing prior classification errors
  - Showing topical discrepancies suggesting possible gaps in descriptive content
- Further development will require analysis and improvement efforts using job description sets for sizeable real organizations
- User-friendliness enhancements to the interface may be essential for obtaining HR subject matter expert engagement
- Approach may reduce risk exposures, improve clarity of job roles, and reduce proliferation of substantially identical jobs across large organizations