Clustering in Natural Language Processing & Text Mining

CpE 6330 Project Presentation

Stuart Miller
Master's Student in Computer Engineering



Introduction

- > IEEE Xplore provides a developer API
 - 176 Journals, Magazines, and Transactional Collections
 - Countless more conference and workshop proceedings





Questions/Goals

- > How similar are papers published in journals?
- > Can papers be neatly clustered by subject matter?
- > Do Journals and Transactions vary more or less than Conference and Workshop Proceedings?
- > Is this kind of clustering a useful tool for researchers?



Introduction

> Use Abstracts

- Assume representative of paper in general
- More manageable
- About 200 words each

The Internet of Things (IoT) shall be able to incorporate transparently and seamlessly a large number of different and heterogeneous end systems, while providing open access to selected subsets of data for the development of a plethora of digital services. Building a general architecture for the IoT is hence a very complex task, mainly because of the extremely large variety of devices, link layer technologies, and services that may be involved in such a system. In this paper, we focus specifically to an urban IoT system that, while still being quite a broad category, are characterized by their specific application domain. Urban IoTs, in fact, are designed to support the Smart City vision, which aims at exploiting the most advanced communication technologies to support added-value services for the administration of the city and for the citizens. This paper hence provides a comprehensive survey of the enabling technologies, protocols, and architecture for an urban IoT. Furthermore, the paper will present and discuss the technical solutions and best-practice guidelines adopted in the Padova Smart City project, a proof-of-concept deployment of an IoT island in the city of Padova, Italy, performed in collaboration with the city municipality.

A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, "Internet of Things for Smart Cities," in *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22-32, Feb. 2014. doi: 10.1109/JIOT.2014.2306328



Outline

- > Introduction
- > 2018 Papers Analysis by LDA and LSA
- > Model Applicability
- > TF-IDF Analysis
- > Perplexity & Insights by Publication
- > Topic Accuracy
- > Conclusions and Ending Points



First Data Set

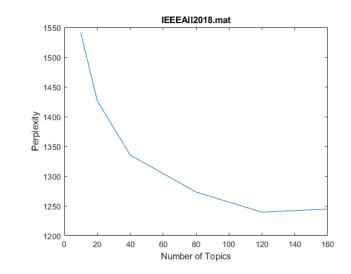
- > Consider all available papers with 2018 publishing dates
 - 21,985 papers
 - 30 MB of abstract and associated bookkeeping information



Latent Dirichlet Allocation Approach

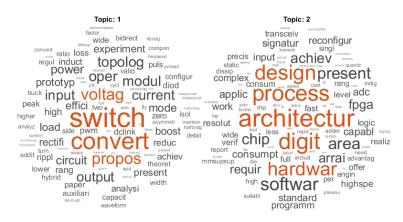
- > Perplexity
 - $perplexity(t) = e^{-\sum_{d} p(t,d) \cdot \log(p(t,d))}$
 - Where p(t, d) is the probability of topic t in document d, summed over all documents
 - Higher number mean even distribution (lower number means better fit)

- > Total summed perplexity for **IEEAll2018** data set
 - Choose 120 topics





IEEEAll2018 First 8 Topics





Topic: 3





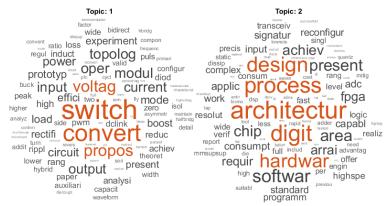








IEEEAll2018 First 8 Topics (of 120)



Topic: 5

regener sepeat capabl acquir into Stimul larger larger sideband

column signal liver sideband liver sideband successfulli span electrod physiolog coher dispersion inject a span for suffici needl minim disord promis tumor cerebrosphal inject a span for suffici needl minim disord promis tumor cerebrosphal span for suffici needl minim disord promis span for sufficient span for sufficient span for sufficient sufficient span for sufficient sufficient span for sufficient su

spinal spinal surger across features features features features features features features format features format features format features features

specific anim Snr was signific specific coeffici mulat signific specific mod singl

bitstream demonstr digit

perfus restor character

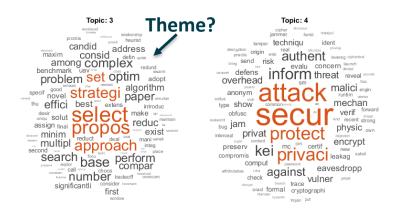


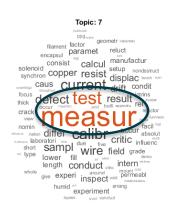
adopt standard successfulli

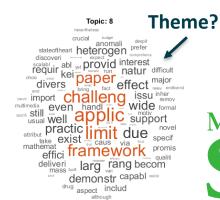
ensur togeth describ

consequ preliminar

although speedup

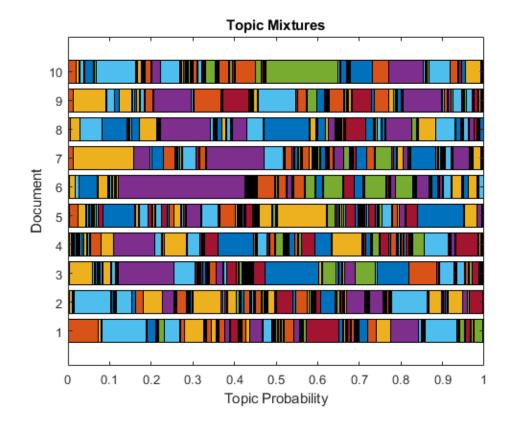








IEEEAll2018 First 10 Documents



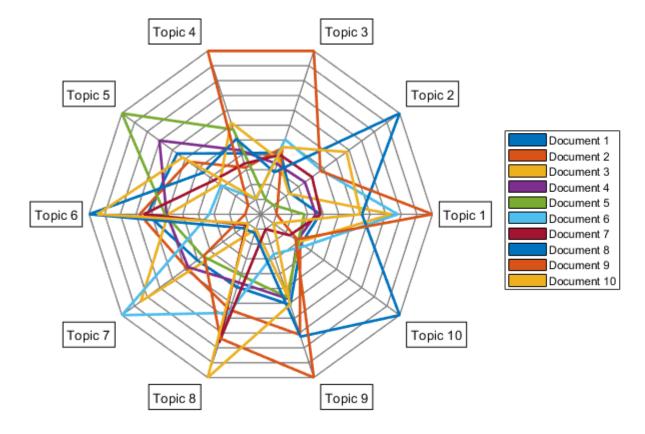


Latent Semantic Analysis Approach

- > Choose 120 topics once again
- > Remember, fitness scores, not probability distributions

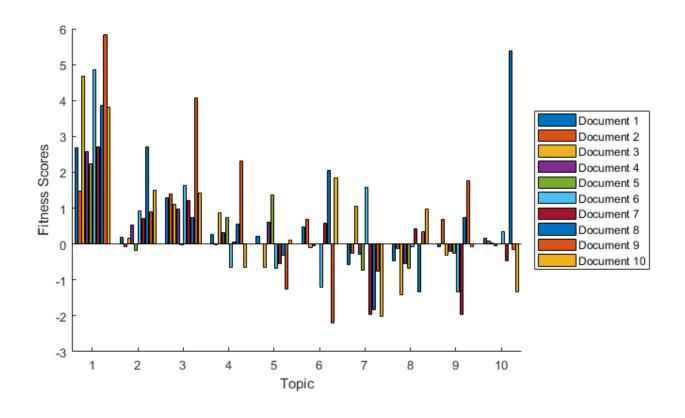


IEEEAll2018 First 10 Documents, First 10 Topics





IEEEAII2018 First 10 Documents, First 10 Topics





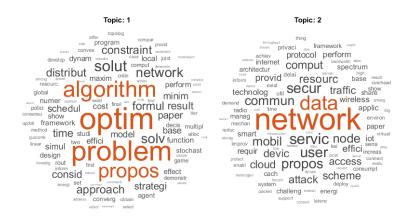
Overview

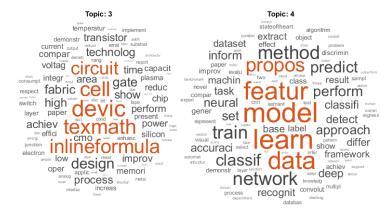
- > Pretty bad classification
 - Unusable number of topics
 - Topic assignments are messy

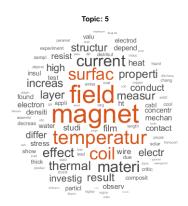
- > Re-run with fewer topics
 - 20 topics
 - Sticking with this assignment for the remainder of the presentation



IEEEAll2018 First 8 Topics (of 20)

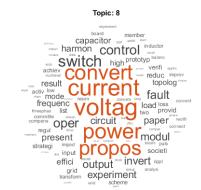






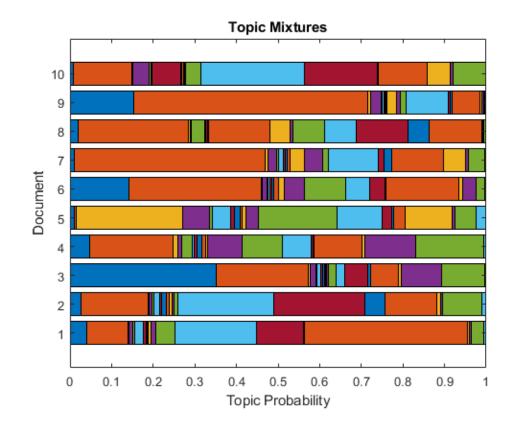






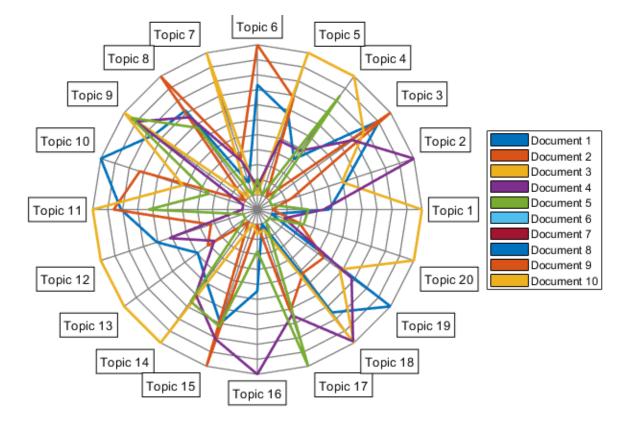


IEEEAll2018 First 10 Documents (20 topics)





IEEEAll2018 First 10 Documents (20 of 20 topics)





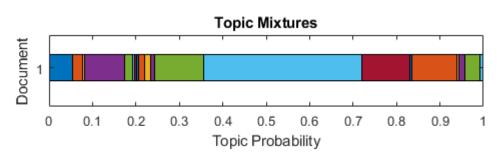
Outline

- > Introduction
- > 2018 Papers Analysis by LDA and LSA
- > Model Applicability
- > TF-IDF Analysis
- > Perplexity & Insights by Publication
- > Topic Accuracy
- > Conclusions and Ending Points



Model Applicability

- > Can we apply this model to any paper?
- > Dr. Wunsch's **Survey of clustering** algorithms



Abstract

Data analysis plays an indispensable role for understanding various phenomena. Cluster analysis, primitive exploration with little or no prior knowledge, consists of research developed across a wide variety of communities. The diversity, on one hand, equips us with many tools. On the other hand, the profusion of options causes confusion. We survey clustering algorithms for data sets appearing in statistics, computer science, and machine learning, and illustrate their applications in some benchmark data sets, the traveling salesman problem, and bioinformatics, a new field attracting intensive efforts. Several tightly related topics, proximity measure, and cluster validation, are also discussed.

Rui Xu and D. Wunsch, "Survey of clustering algorithms," in IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645-678, May 2005.

Topic 13



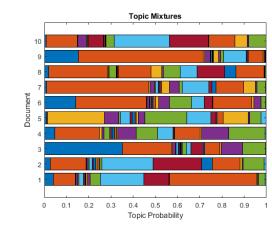


Question #1: How similar are journal papers?

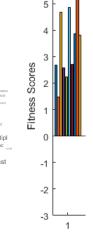
> Topic #1 - generic

> Overall Topic
Mixtures

> Topic #3 - TeX syntax









Topic: 3

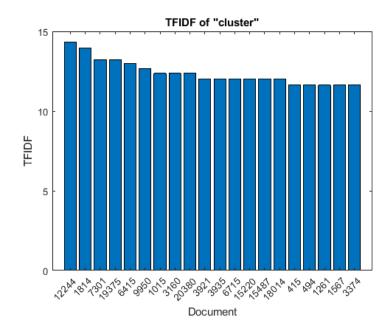


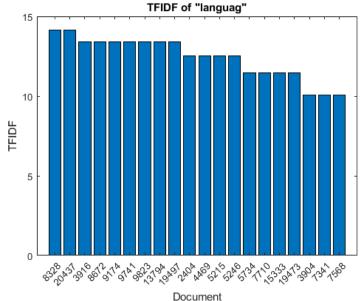
Outline

- > Introduction
- > 2018 Papers Analysis by LDA and LSA
- > Model Applicability
- > TF-IDF Analysis
- > Perplexity & Insights by Publication
- > Topic Accuracy
- > Conclusions and Ending Points



Term Frequency-Inverse Document Frequency Approach





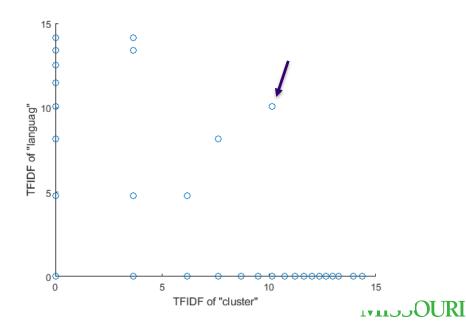


TF-IDF Results

- > Document 7568
 - On Pareto front
 - IEEE web search for "cluster language" returns this as well

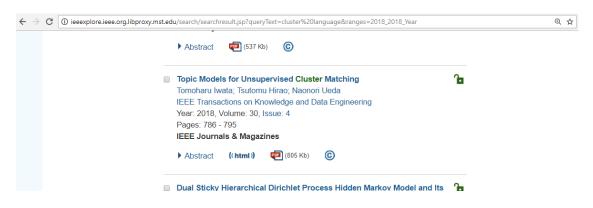
We propose topic models for unsupervised cluster matching, which is the task of finding matching between clusters in different domains without correspondence information. For example, the proposed model finds correspondence between document clusters in English and German without alignment information, such as dictionaries and parallel sentences/documents. The proposed model assumes that documents in all languages have a common latent topic structure, and there are potentially infinite number of topic proportion vectors in a latent topic space that is shared by all languages. Each document is generated using one of the topic proportion vectors and language-specific word distributions. By inferring a topic proportion vector used for each document, we can allocate documents in different languages into common clusters, where each cluster is associated with a topic proportion vector. Documents assigned into the same cluster are considered to be matched. We develop an efficient inference procedure for the proposed model based on collapsed Gibbs sampling. The effectiveness of the proposed model is demonstrated with real data sets including multilingual corpora of Wikipedia and product reviews.





TFIDF Conclusions

- > Not useful
- > Nothing more than a basic internet search query
- > As per the algorithm's design meant to be fast and simple
- > Will omit for rest of this project



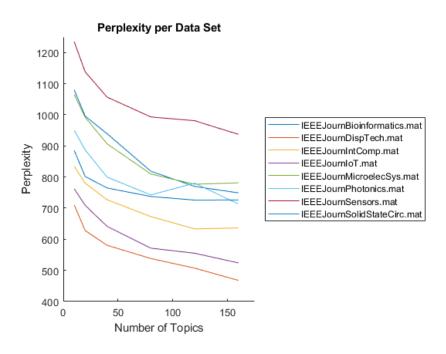


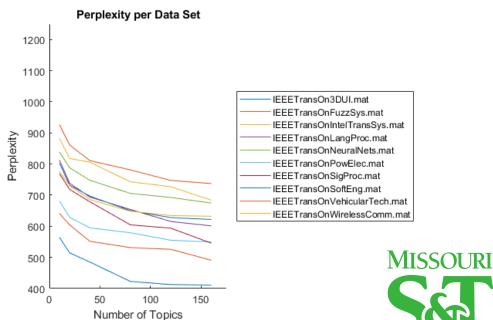
Outline

- > Introduction
- > 2018 Papers Analysis by LDA and LSA
- > Model Applicability
- > TF-IDF Analysis
- > Perplexity & Insights by Publication
- > Topic Accuracy
- > Conclusions and Ending Points

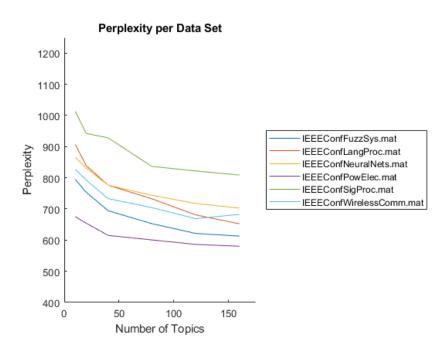


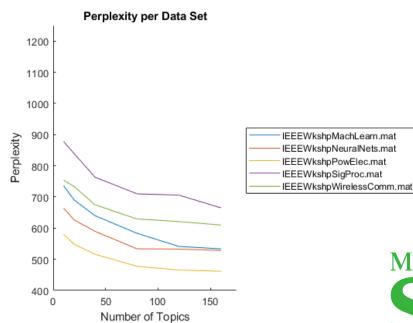
Perplexity for Journals and Transactions





Perplexity for Journals and Transactions

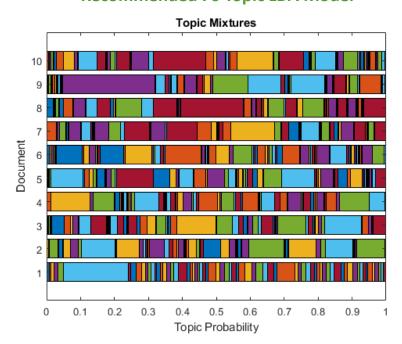




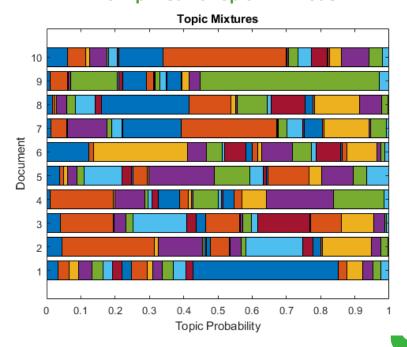
MISSOURI

IEEE Sensors Journal Insights

Recommended 70 Topic LDA Model

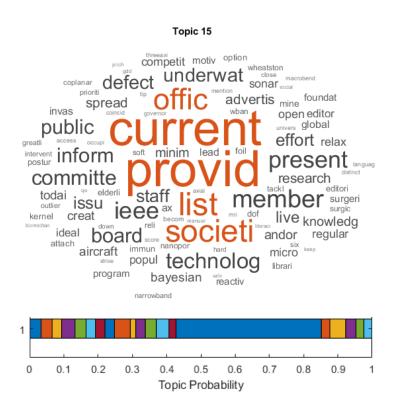


Simplified 20 Topic LDA Model



SOURI

Topic 15 (Primary Topic of Docs 1, 14, 58, 90...)



Abstract

Provides a listing of current society officers.

"IEEE Sensors Journal publication information," in IEEE Sensors Journal, vol. 9, no. 11, pp. C2-C2, Nov. 2009.

IEEE SENSORS JOURNAL

The IEEE SENSORS JOURNAL is published by the IEEE Sensors Council, which is sponsored by 26 IEEE societies (see the inside back cover). Subscriptions for IEEE members are \$33.00 per year. For information write to the IEEE Operations Center at the address below. Member copies of Transactions/Journals are for personal use only.

Editor-in-Chief EVGENY KATZ Milton Kerker Chair in Colloid Science Chemistry & Biomolecular Sci. Clarkson Univ. Potsdam, NY 13699 USA

M. NURUL ABEDIN

FRANCISCO J. ARREGUI

Pamplona 31006, Spain

PLAMEN ATANASSOV

Univ. New Mexico

Univ. Publica de Navarra

NASA Langley Research Center

Lasers & Electro-Optics Branch

Hampton, VA 23681-2199 USA

Ingenieria Electrica y Electroni

Univ. Calif. San Diego La Jolla, CA 92023-0357 Past Editor-in-Chief H. TROY NAGLE

UNC Chapel Hill-

NC State Univ.

Raleigh, NC 27695-7115

USA

ELLIOT R. BROWN

MASAYOSHI ESASHI

Univ. California, Santa Barbara

Senior Editor

GERT CAUWENBERGHS

Founding Chair Biomedical Engineering

Senior Editor Senior Editor RALPH ETIENNE-CUMMINGS IGNACIO R. MATIAS Univ. Maryland Dept. Elect. Comput. Eng. College Park, MD 20742 TISA

Public Univ. Navarra Dept. Elect. & Electron. Eng. 31006-Pamplona

Senior Editor MICHIEL I. VELLEKOOP Inst. Sensor and Actuator Systems Vienna Univ. Technol. Guschausstrasse 27/366 A-1040 Vienna Austria

Senior Editor JOHN R. VIG Concultant 33 Bucks Mill Rd. Colts Neck, NJ 07722 TISA

Founding Editor-in-Chief VLADIMIR LUMELSKY Univ. of Wisconsin Madison, WI 53706 USA

Editorial Board OKYAY KAYNAK Rogazici University

Elec. Comput. Eng. Santa Barbara, CA 93106 USA TIANHONG CUI Dept. Mechanical Engineering Nanofabrication Center KISEON KIM Univ. Minnesota, Twin Cities Campus Minneapolis, MN 55455 USA

34342 Istanbul, Turkey BAHRAM G. KERMANI Complete Genomics, Inc. Mountain View, CA 94043 USA Gwangju Inst. Sci. Tech. Gwangju, 500-712 Korea

JACQUES PISTRE F-333402 Talence Cedex, France PAUL P. L. REGTIEN Twente Univ. Wiskunde & Inform, Elektrotech, The Netherlands

V6T 1Z4 Vancouver, BC, Canada OOI KIANG TAN Nanyang Tech. Univ. Microelectronics Centre Singapore 639798 PATRICK RUTHER FRANCIS ENG HOCK TAY Microsystem Materials Natl. Univ. Singapore Laboratory (MML) Done Mach Eng

BORIS STOEBER

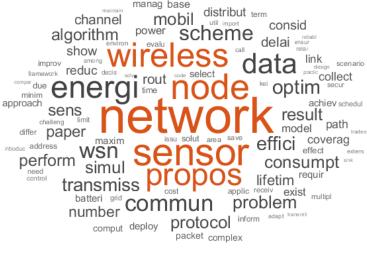
Mechanical Eng.

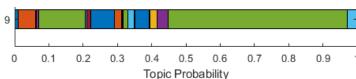
Univ. British Columbia

MISSOURI

Topic 19 (Primary Topic of Doc 9)







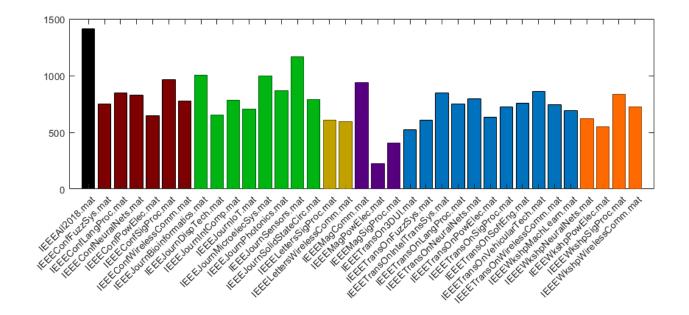
Abstract

As an extension of wireless sensor network in underwater environment, underwater acoustic sensor networks (UASNs) have caused widespread concern of academia. In UASNs, the efficiency and reliability of data transmission are very challenging due to the complex underwater environment in variety of ocean applications, such as monitoring abnormal submarine oil pipelines. Motivated by the importance of energy consumption in many deployments of UASNs, we therefore propose an energyefficient data transmission scheme in this paper, called energy-efficiency grid routing based on 3D cubes (EGRCs) in UASNs, considering the complex properties of underwater medium, such as 3D changing topology, high propagation delay, node mobility and density, as well as rotation mechanism of cluster-head nodes. First, the whole network model is regarded as a 3D cube from the grid point of view, and this 3D cube is divided into many small cubes, where a cube is seen as a cluster. In the 3D cube, all the sensor nodes are duty-cycled in the media access control layer. Second, in order to make energy efficient and extend network lifetime, the EGRC shapes an energy consumption model considering residual energy and location of sensor nodes to select the optimal cluster-heads. Moreover, the EGRC utilizes residual energy, locations, and end-to-end delay for searching for the next-hop node to maintain the reliability of data transmission. Simulation validations of the proposed algorithm are carried out to show the effectiveness of EGRC, which performs better than the representative algorithms in terms of energy efficiency, reliability, and end-to-end delay.

K. Wang, H. Gao, X. Xu, J. Jiang and D. Yue, "An Energy-Efficient Reliable Data Transmission Scheme for Complex Environmental Monitoring in Underwater Acoustic Sensor Networks," in IEEE Sensors Journal, vol. 16, no. 11, pp. 4051-4062, June 1, 2016.



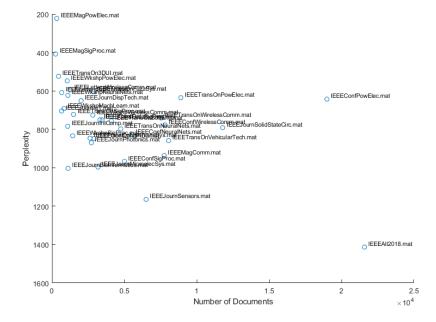
Perplexity by Journal





Data Set Size

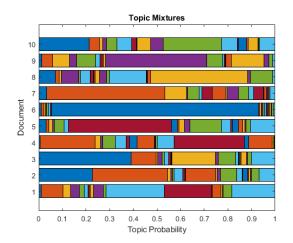
- > Smaller data set ∝ lower perplexity
 - Magazines vs. All2018
- > Plenty of exceptions though
 - IEEE Conferences on Power Electronics

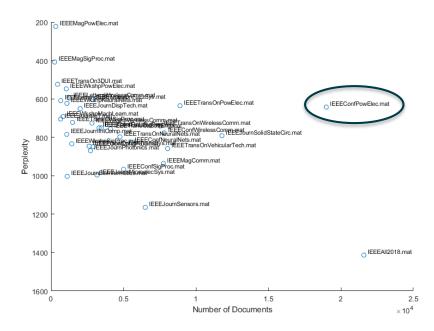




Data Set Size

- > Smaller data set ∝ lower perplexity
 - Magazines vs. All2018
- > Plenty of exceptions though
 - IEEE Conferences on Power Electronics

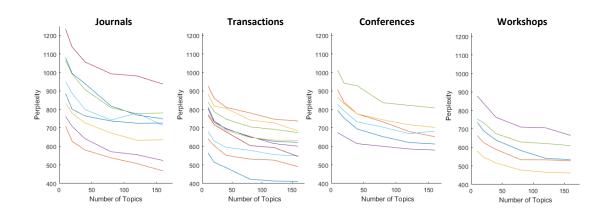






Question #2: Do Journals and Transactions vary more or less than Conference and Workshop Proceedings?

- > Not really
- > Journals are more varied
- > Subject matter is far more influential





Outline

- > Introduction
- > 2018 Papers Analysis by LDA and LSA
- > Model Applicability
- > TF-IDF Analysis
- > Perplexity & Insights by Publication
- > Topic Accuracy
- > Conclusions and Ending Points



Highly Specific Topic:

IEEE Transactions on 3D User Interfaces

chiecentr sovirtual multisensori california half toth review vision humancomput calculate use with the contact mechanic offset committee contact mechanic offset committee contact mechanic large provided in the sensor use product feedback uput the sensor use product feedback uput the sensor use product feedback uput the sensor offset may be uput the sensor of t optitrack thank children mass Symposium electron archeri anesthesia whist event audiovisu

produc is softwar actual character period of software period above the pure of the period of software per wand low indic result dof typic stimuli match accuraci significantli posit read postick at around rese assess waterpipter fundament

Topic: 6

Topic: 11

quests steecheart total highlight degrad out deal third procedur he determin domin quality regard full displai chang field confirm pattern lead dones applic displai caux platem pattern was papilic displai caux platem pattern was pattern with caux platem pattern was platem pattern with caux platem pattern was pattern with caux platem patte guidein full physic improv experi theshold marken exploit desktop same surround regard sea seethrough identifi perceiv signal malighed undestood most brended adult.

Topic: 16 against attent pecom current aspect shit music implement graphic sphere strength variou present order involv larger explor interact combin rensit new combin enhanc ransit mew combin enhanc rensit mew combin rensit re follow natur base allow focu specifi oenter possibl approach part mammet workspac demonstr express graph more workspac demonstr express converg telepras ward black program

tretch stretchald vertail with based individu multitool eart balder nois constraint mount nois constraint mount appli manipul average to deform decom group novel puzzi solv interpret decom group novel area advantag oot perper tott conduct rem
reg unv wai human work eesseel weght
related structur wind aid simpt dotal
equation algorithm
equation recovers withing

Topic: 7

The Walk Control of Fice I was a support of Fice I was relat import

Topic: 12 hierarchi neses gerner totan naradigm sati modup relationship paradigm statt attempt technolog valid potenti soptist attempt technolog side protein signature in the country of the cou welch even level game benefit boot stand tow develop train copper scheme hyperbee engin video percept destates advanc respons poster

Topic: 17 feethisp defract aboves some services feethis behavior particularly aboves and a service services and a service services and a service service service services and a service service services and a service service service service services and a service service service services and a service se there can be perceived to conduct the process of the perceived to the perc projected maintain world user integr voic grid popular give Cognit experi abl situat motic count count reach w increas regard appear decreas place problem session decreas producti psycholo steps facial wayfind vibrotactii diminsh steps consequi pass efectir

econstituce septice index into hoscori control aggregation of the control a phase front Cursor two increas class mice printer trajectori environ perspect region perspect defin. Perspect region region region represent defin. Perspect represent defin. Perspect region r

Topic: 8

pano gesturebas varieti embed lvo among close frame place acquisit concom precis move indic investig dunt smart rate perform task assist make mant rate perform task assist relations to the pair topic pair top dach potenti target demonstr down experiment high necessari

Topic: 13

manuscid gasher and the control property of the contro kload analysi goal bimanu krowledg makin bull accuraci lak bimanu krowledg makin bull accuraci lak bountstreen dawr samram oppor

Topic: 18

thought nonisomorph etal tradingso separ overal brescop lang short techingstores term effect conduct common and result three experi adjust have predict travel two sx task type woods time differ techniqu plec office complet - rotat test map midair pouts base mode translat Bur instrument smartphon makes isometr accomplish ridio

definition of the concepts of registry to the process process and the process proces evalu geometri subarre processi independi applici recordi malari subarre processi independi applici recordi malari subarre prototyp absiract erapa; absiract e

Topic: 9

nandamin conclus keyboard handheld dimens autgerform interfac preliminari distribution of the control of the c 2006 arrain organ Companies Companie

Topic: 14 pococogni est composition substituti pococogni est composition substituti pococogni est composition substituti pococogni est composition substituti pococogni est composition Desire MOTAL COLLEGE

Coppet Scannin Cisple is sens spend displetion of the competition o

Topic: 19 sechtacis enabl thu as four layout sechtacis enabl thu as four states button was law locomot set passiv resurr veloc rel show direct enter redirect control path wind space control path wind space control path wind space control path with space control path control path space control path - simultan immers mean window detail window discuss tool concept framework scenario

Topic: 5

nepord bushpusk deair interest expand until product challeng muturest oppose term of the proper interest commune of drag constructive present commune drag constructive present commune of drag constructive present commune of drag proper intuit of the present constructive present communication of the present communication of the proper communication of the property of the present communication of the presen sound IIIIag motion dui arri modi estadabh camera provid allow final mesh contest adappopri que de la depropri de la depropri final mesh contest appropri de la depropri final mesh contest appropri de la depropri de la depretación del del del depretación de la depretación de la depretación del del de

Topic: 10 massibes transpor distant represent throughput cloud placement correspond direct widget respect score extend Select of Sel sudhoove extend Select or such constitution of the substitution of immed techniqu convent around

Topic: 15 contenue realiz usabl contribut later contribute are recent warp after releas metaphor avail to new burst lost adapt as design inform difficult formation of the state o augment physic focus acute mani Visual certain travel emploi rester configur expect inspir attended printing print

Topic: 20 lowcost best avoid inerti wari remain mutition special frames often accur automat special finger eli multitouch viewpoint and multipud mul manual langibl obtain the case legal cortex steed except common behind expects.



Highly Specific Topic:

IEEE Transactions on 3D User Interfaces

chiecentr sovirtual multisensori california half toth review vision humancomput restance to the second of the societi note vainer sow march lieee haptic note vainer sow march lieee lilus current rapidii children mass Symposium electron archeri anesthesia whist event audiovisu

Topic: 6

produc is softwar actual character

period of software period above the pure of the period of software per

wand low indic result dof typic stimuli

Topic: 11

third procedur has determin domin quality

stereo chang field confirm pattern

dones applic displai caux platem pattern was papilic displai caux platem pattern was pattern with caux platem pattern was platem pattern with caux platem pattern was pattern with caux platem patte

exploit desktop same surround regard sea

seethrough identifi perceiv signal malighed undestood most brended adult.

guidein full physic improv experi theshold marken

Topic: 16

shit music implement graphic sphere

explor interact enhanc transit moving combin provided in the control of the contr

warplac already rather remot law technic against attent becom current aspect

strength variou present order involv larger

follow natur base allow focu specifi

oenter possibl approach part workspac demonstr express graph held program and back program

quests steecheart total highlight degrad out deal

match accuraci significantli posit read

joystick around men assess wakingingles

human work restrict weight around structur we high aid d simpl Topic: 7

deform group

derit light take

advanta



trech shelchald vertal such bread individu multitool exet for nois constraint mount

puzzl solv interpret

a conduct

ntroduc turn som and extens

rose motor appli manipul mount average to















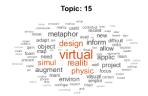












der consid facetric

stereoscop atract

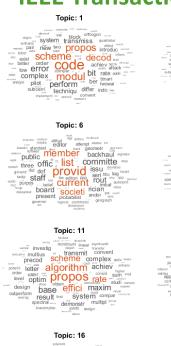
degre variat



loscost best avoid inerti vari remain mutivior special figurer often accur automat special finger eli multitouch viewpoint correct depth forc inspect futur requi matur orient gen good enabl framework tool concept pose room difficulti manual tangibl obtain free case displace proper recogn common behind impre-

Broader Topic:

IEEE Transactions on Wireless Communications























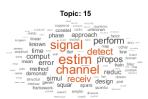


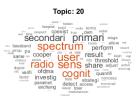














Topic Estimations

Modulation/coding Topic: 1 schemes

where the part of the part of

Meta papers (member lists, etc.)

power bear and adjug of water from the control of t

New algorithm proposition

ware investignment near septement of the proof of the pro

Throughput/packet

Cipable broducest for interest of the control of th

Power Optimization

throughput simular transmiss throughput transmiss tra

MIMO Topic: 7

Rate estimation

scheme

some result number regim most closed form most must be seen to be see

Improvements on Topic: 17 existing model

Topic: 3

week and the state of the state

Topic: 8

hybrid seitch rotiorhoon

cost threshold novel see verd

provid letter access

provid letter access

scheme multipl

scheme mu

Sensor/node networks

Topic: 13

process Senatory of the senatory of

Survey Papers Topic: 18

presence correct take direct Landques prove shown take direct Landques maintain adapt multipath and adapt multipath and adapt multipath and adapt multipath and adapt with a second prove shown adapt second prove shown and second prove show

Topic: 4

control of the contro

Security

multiarterna transmits state emiden versione haves transmits state emiden versione haves transmits state emiden versione pound of the p

Topic: 14

data of the point of the point

Channel interference

and the state of t

New model proposition

Topic: 5

develop literatur densiti extend description in the control in the cont

Topic: 10

west downlink depty tize number uplink west considered to the construction of the cons

Topic: 15

waste phase approach to the phase approach t

Topic: 20

Secondari primari saga sebagai sebagai



Question #3: Can papers be neatly clustered by subject matter?

- > Key aspects
 - Number of topics: smaller is better
 - > Optimum perplexity often way too high
 - Subject matter is important: broad subject accommodate models better
 - Human insight is always needed, ambiguous topics

> Conclusion? **Yes**, fairly well



Outline

- > Introduction
- > 2018 Papers Analysis by LDA and LSA
- > Model Applicability
- > TF-IDF Analysis
- > Perplexity & Insights by Publication
- > Topic Accuracy
- > Conclusions and Ending Points



Conclusions

- > Is this a valuable research tool for you?
- > Can reveal current research trends
- > What is this valuable for?
 - Classifying upcoming papers
 - Scoring your paper according to current trends



Code

- > Scripts and plots developed in MATLAB
- > Code available on my Github
 - https://github.com/stewythe1st/Research-Mining
- > IEEE Web API
 - https://developer.ieee.org/docs



Moving Forward

- > Figure out a good way to determine optimal number of topics
- > Try out with documents from other sources
- > Perhaps perform analysis on paper content, not just abstract



Citations

Research Used

- > K. S. Jones, "A Statistical Interpretation Of Term Specificity And Its Application In Retrieval", Journal of Documentation, vol. 28, issue 1, pp. 11-21, 1972
- > D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, vol. 3, pp. 993–1022, Jan. 2003.
- > S. C. Deerwester, S. T. Dumais, G. W. Furnace, R. A. Harshman, T. K. Landauer, K. E. Lochbaum, and L. A. Streeter, "Computer information retrieval using latent semantic structure," 13-Jun-1989.

Works Used in Examples

- > A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, "Internet of Things for Smart Cities," in IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22-32, Feb. 2014.
- > Rui Xu and D. Wunsch, "Survey of clustering algorithms," in IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645-678, May 2005.
- > T. Iwata, T. Hirao and N. Ueda, "Topic Models for Unsupervised Cluster Matching," in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 4, pp. 786-795, April 1 2018.
- > K. Wang, H. Gao, X. Xu, J. Jiang and D. Yue, "An Energy-Efficient Reliable Data Transmission Scheme for Complex Environmental Monitoring in Underwater Acoustic Sensor Networks," in IEEE Sensors Journal, vol. 16, no. 11, pp. 4051-4062, June 1, 2016 MISSOURI
- > "IEEE Sensors Journal publication information," in IEEE Sensors Journal, vol. 9, no. 11, pp. C2-C2, Nov. 2009.

Questions?

- > Questions/Goals
 - How similar are papers published in journals? More similar than you think!
 - Can papers be neatly clustered by subject matter? Yes!
 - Do Journals and Transactions vary more or less than Conference and Workshop Proceedings? Not really
- > Conclusions
 - Not much of a valuable research tool for most people
 - Does reveal current research trends
 - Is this valuable for...
 - > Classifying upcoming papers
 - > Scoring your paper according to current trends

