

sakshi_2022_recent_trends_in_mathematical_expressions_recognition_an_lda_based_analysis

Year

2022

Author(s)

Sakshi and Vinay Kukreja

Title

Recent trends in mathematical expressions recognition: An LDA-based analysis

Venue

Expert Systems With Applications

Topic labeling

Manual

Focus

Primary

Type of contribution

Established approach

Underlying technique

Manual labeling

Topic labeling parameters

/

Label generation

The authors have reviewed high-loading articles of all topic solutions in the current study. Further, the labeling of all the topic solutions has been performed individually to formulate the conclusive topic label.

The task involved examining extracted terms and abstracts of documents related to a particular factor or topic, thus analyzing and interpreting the underlying research area or trend.

The two authors have consensually termed the research zones, areas, and trends based on the keywords fetched for every topic solution.

According to the frequent practice in classical factor analysis, the authors have related each topic label to its key terms or high-loading terms and documents to assist in labeling every topic solution.

Further, for every topic solution, a table listing has been maintained that lists the prominent high-loading terms (key terms) and documents that load satisfactorily well (highly associated) on a particular topic or factor.

Thus, the labeling has been accomplished manually.

Table 2 presents ten high loading terms as key terms and the high loading documents for each topic or factor in two, five, and ten topic solutions, along with a contribution value or factor-loading value.

The contribution value or factor loading value for each topic specifies the extent of the relation of the related key term with a specific topic solution.

The values shown in the "Contribution" column of Table 2 mean the probability values based on the estimated topic distributions.

For example, "79.37" in the first row indicates that the occurrence probability of the first topic in the topic distribution of paper with title 285 is 0.793.

The high loading terms and documents are those terms and studies, respectively, that load or relate sufficiently well to the identified factor or topic label.

Table 2

High loading Research Articles for Topic Solutions.

Topic ID	Key Terms	Topic Label	Count	High Loading Documents	Contribution (%)
2.1	Two Topic Solution method, analysis, result, problem, base, segmentation, stroke, structure, user, propose, recognize, approach, structural, input, character, process, document, present, technique, spatial	Recognition Techniques and Analysis Methods	96	Title 285 (Shan et al., 2021)Title 286 (Rhee et al., 2008)Title 97 (Le & Nakagawa, 2017)Title 126 (Dong & Liu, 2017)Title 108 (D.-H. Wang et al., 2020)	79.37 79.31 79.25 77.09 76.21
				Title 35 (Yingying et al., 2009)Title 291 (J. Fitzgerald et al., 2006)Title 19 (Taranta et al., 2016)Title 133 (Taranta and LaViola, 2015)Title 34 (Gong et al., 2015)	79.65 79.38 79.28 79.26 79.21
5.1	Five Topic Solution structure, approach, analysis, tree, result, base, relationship, structural, propose, parse, spatial, graph, process, relation, problem, technique, present, recognize, interpretation, information	Parsing Techniques	45	Title 271 (Naderan, 2017)Title 123 (Ray Genoe et al., 2006)Title 215 (Celik & Yanikoglu, 2011a)Title 226 (H.-J. Winkler et al., 1995)Title 229 (J. Fitzgerald et al., 2006)	79.46 78.45 78.31 77.28 76.28
				Title 166 (Chan & Yeung, 2000)Title 46 (Xiaorong & Chaoying, 2004) Title 35 (Yingying et al., 2009)Title 186 (J. Huang et al., 2020)Title 285 (Shan et al., 2021)	79.28 77.25 76.81 76.25 74.78
5.2	method, image, character, base, document, problem, segmentation, propose, present, provide, analysis, study, error, technique, performance, research, print, math, box, form	Character-Based Recognition methods	74	Title 277 (Raymond Genoe, 2010)Title 285 (Shan et al., 2021)Title 286 (Rhee et al., 2008)Title 283 (H. Wang & Shan, 2020)Title 95 (L. Hu & Zanibbi, 2013)	79.59 79.21 78.59 77.31 73.31
5.3	feature, model, propose, classification, base, rate, result, show, set, classifier, level, recognize, stroke, high, formula, achieve, accuracy, dataset, present, distance	Segmentation and classification procedures	92	Title 143 (Álvaro, Sánchez, & Benedí, 2014b)Title 255 (Mahdavi et al., 2019)Title 88 (J. Wang et al., 2019)	79.69 78.47 70.05
5.4	online, task, dataset, neural, attention, network, competition, datum, propose, feature, model, result, image, performance, accuracy, base, approach, stroke, offline, crohme	CROHME and Neural Network Model	61	Title 237 (L. Chen, 1992)Title 243 (Naderan & Zaychenko, 2013)Title 244 (S. J. Rani & Kumari, 2016)	79.73 79.63 76.51
5.5	user, stroke, interface, recognize, time, input, result, order, stage, single, method, line, online, enter, network, support, computer, write, combine, processing	Structural Analysis mechanism	53		
10.1	Ten Topic Solution problem, character, base, input, provide, math, develop, computer, mathematic, application, make, accuracy, equation, recognize, present, exist, research, dimensional, design, solve	Dimensional Model Construction and Offline recognition	24	Title 218 (Wells, 1976)Title 293 (Smirnova & Watt, 2010)Title 149 (Lee et al., 2018)Title 43 (Vuong et al., 2010)Title 185 (J Zhang & Hong, 2008)	74.39 73.22 66.28 64.18 63.28
				Title 188 (Phong et al., 2017)Title 140 (Álvaro & Sánchez, 2010)Title 221 (Pillay, 2014)Title 135 (Phong et al., 2020)	79.52 77.75 69.52 68.49
10.2	method, document, image, propose, segmentation, print, problem, base, technique, work, extract, stage, accuracy, character, result, present, information, experiment, scientific, analysis	Parse Tree-Based recognition model	34	Title 215 (Celik & Yanikoglu, 2011b)Title 209 (Álvaro, Sánchez, & Benedí, 2014a)Title 197 (Shi et al., 2011)Title 172 (Le et al., 2016)Title 83 (L. Hu & Zanibbi, 2011)	79.61 78.61 77.14 72.13 70.46
10.3	stroke, graph, order, parse, result, context, rate, model, grammar, time, base, score, parsing, recognize, algorithm, crohme, online, process, top, free	Contextual Mapping and Graph-based recognition	46	Title 217 (Büyükbayrak et al., 2007)Title 243 (Naderan & Zaychenko, 2013)Title 244 (S. J. Rani & Kumari, 2016)Title 101 (Ernesto Tapia & Rojas, 2005)Title 232 (Mahmoud et al., 2011)	99.79 99.79 99.79 99.57 99.53
10.4	user, interface, recognize, result, box, input, network, write, draw, matrix, enter, support, combine, output, stroke, neural, enable, mathematic, online, fuzzy	Input Methods	27	Title 233 (H.-J. Winkler et al., 1995)Title 48 (Kaplan, 2016)Title 29 (Phan et al., 2015)Title 164 (Phan et al., 2018)Title 179 (Xiangwei & Abaydulla, 2010)	79.59 77.58 61.44 59.32 43.13
10.5	method, approach, process, analysis, generate, present, reduce, propose, structural, segmentation, efficient, base, language, time, technique, problem, experiment, alternative, structure, strategy	Performance parameters and analysis	21	Title 96 (A. M. Awal et al., 2009)Title 247 (G. Chen & Tang, 2013)Title 38 (A.-M. Awal et al., 2010a)Title 121 (Medjkoune et al., 2012)Title 201 (Rhee & Kim, 2009)	79.79 77.79 71.34 67.13 63.31
10.6	propose, level, segmentation, stroke, spatial, online, set, information, classifier, structure, relationship, train, interpretation, offline, accuracy, label, input, result, global, evaluate	Segmentation and Spatial Constructs	33	Title 277 (Raymond Genoe, 2010)Title 283 (H. Wang & Shan, 2020)Title 285 (Shan et al., 2021)Title 286 (Rhee et al., 2008)Title 27 (Shan et al., 2021)	79.77 78.71 76.35 73.26 72.49
10.7	model, attention, network, propose, neural, end, base, latex, achieve, art, decoder, state, crohme, deep, structure, convolutional, dataset, accuracy, sequence, dimensional	Attention and Deep Networks	14	Title 118 (A.-M. Awal et al., 2010c)Title 123 (Ray Genoe et al., 2006)Title 131 (Li & Tian, 2010)Title 34 (Gong et al., 2015)	79.44 78.44 77.62 76.32
10.8	structure, analysis, tree, error, propose, approach, base, structural, spatial, relationship, problem, result, performance, method, evaluation, pen, construct, process, give, relation	Spatial relations and symbol identification	41	Title 143 (Álvaro, Sánchez, & Benedí, 2014b)Title 127 (Yousefi et al., 2010)Title 204 (Álvaro et al., 2013)Title 265 (Okamoto & Higashi, 1995)Title 268 (Littin, 1995)	79.82 79.69 79.12 78.24 76.39
10.9	feature, base, classification, dataset, model, result, image, method, present, set, approach, distance, match, math, datum, online, recognize, database, evaluation, show	Features based model development	43	Title 223 (H. J. H.-J. H. J. Winkler & Lang, 1997)Title 224 (Y. Hu et al., 2014)Title 295 (H. J. H.-J. H. J. Winkler & Lang, 1997)Title 60 (Guo & Liu, 2018)	79.72 79.34 78.37 78.33
10.10	graph, feature, dataset, result, task, online, competition, datum, performance, accuracy, application, present, provide, test, evaluate, problem, research, base, experiment, offline	Online and Offline Recognition	42		

Motivation

Depicting the (two, five or ten) core research areas in the collected work.

Topic modeling

LDA

Topic modeling parameters

α : $1/T$ (where T is the nr of topics)

β : 0.01

Nr of topics (T): 2, 5, 10

Nr. of topics

17 (2, 5, 10 topics solutions)

Label

Multi-word manually assigned labels identifying core research areas in MER

Label selection

The task of labeling topic solutions has been carried out jointly by the two researchers, and the names of topic labels culminated after several rounds of brainstorming sessions and discussions.

Label quality evaluation

\

Assessors

Two researchers

Domain

Domain (paper): Mathematical expression recognition (MER)

Domain (corpus): Mathematical expression recognition (MER)

Problem statement

Identify and associate (semantic mapping) the leading research zones, core research areas, and research trends steering in the mathematical expression recognition (MER) domain.

Identifying prominent recognition models based on extracted research areas. To develop the development chart from extracted research trends for directing the future works in this direction.

Corpus

Origin: Various digital libraries

Nr. of documents: 325

Details:

- Corpus of 325 research papers published from 1967 to 2021
- Origin: ScienceDirect, IEEExplore, Wiley, and ACM

Document

Title and abstract of a single article published in one of the queried libraries

Pre-processing

Example of pre-processing steps:

Corpus Pre-processing.

Pre-processing Steps	Results
Sample Abstract	We address the problem of handwritten symbol classification in the presence of distortion modeled by an affine transformation. We consider share rotation scaling and transformation since these type of transformation that occurs in practice and focuses on shear in this framework.
After Tokenization	'We', 'address', 'the', 'problem', 'of', 'handwritten', 'symbol', 'classification', 'in', 'the', 'presence', 'of', 'distortion', 'modeled', 'by', 'affine', 'transformation', '.', 'We', 'consider', 'share', 'rotation', 'scaling', 'and', 'transformation', 'since', 'these', 'type', 'of', 'transformation', 'that', 'occur', 'in', 'practice', 'and', 'focus', 'on', 'shear', 'in', 'this', 'framework', '.'
After Stop Word Removal	'address', 'problem', 'handwritten', 'symbol', 'classification', 'presence', 'distortion', 'modeled', 'affine', 'transformation', 'consider', 'share', 'rotation', 'scaling', 'transformation', 'type', 'transformation', 'practice', 'focus', 'shear', 'framework'
Stemming	'address', 'problem', 'handwrit', 'symbol', 'classif', 'presen', 'distor', 'model', 'affine', 'transform', 'consider', 'shear', 'rotat', 'scal', 'transform', 'type', 'transform', 'practice', 'focus', 'shear', 'framework'
Lemmatization	'address', 'problem', 'handwriting', 'symbol', 'classify', 'present', 'distort', 'model', 'affine', 'transform', 'consider', 'shear', 'rotat', 'scal', 'transform', 'type', 'transform', 'practice', 'focus', 'shear', 'framework'

@article{sakshi_2022_recent_trends_in_mathematical_expressions_recognition_an_l
da_based_analysis,

abstract = {Context

Although recognition works on mathematical expressions have been explored for four decades, the current literature and trends are varied and frequently influenced by distinct emerging methods and technology. This situation instigates the necessity of an organized review to provide heedful insight into research trends and patterns currently prevailing in the domain of mathematical expression recognition (MER).

Objective

To identify and associate (semantic mapping) the leading research zones, core research areas, and research trends steering in the MER domain. Identifying

prominent recognition models based on extracted research areas. To develop the development chart from extracted research trends for directing the future works in this direction.

Method

A manual and automatic search has been performed across the reputed digital libraries for corpus formation. The formulated corpus is used for topic modeling, and Latent Dirichlet Allocation is deployed for information modeling for achieving defined objectives.

Result

The corpus of 325 research papers published from 1967 to 2021 has been processed using LDA. The five major research areas and ten research trends are identified. Leading research area is ``Segmentation and Classification Procedures'', and the trend with the highest related publications is ``Contextual and Graph-based recognition''. ``Attention and Deep Networks'' has emerged as the newborn trend, and the identified newborn, young, and matured trends impetrate more exploration from the MER research community.},

```
author = {Sakshi and Vinay Kukreja},
date-added = {2023-03-10 12:13:35 +0100},
date-modified = {2023-03-10 12:13:35 +0100},
doi = {https://doi.org/10.1016/j.eswa.2022.119028},
issn = {0957-4174},
journal = {Expert Systems with Applications},
keywords = {Mathematical expressions, Research trends, Pattern recognition,
Latent dirichlet allocation, Topic modelling},
pages = {119028},
title = {Recent trends in mathematical expressions recognition: An LDA-based
analysis},
url = {https://www.sciencedirect.com/science/article/pii/S0957417422020462},
volume = {213},
year = {2023}}
```