04

2020

Journal of Computer Science Research

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

APSCI | creative commons | CNKI 中国知网 www.cnki.net 中国知识基础设施工程 | Google scholar | Crossref | MyScienceWork

Volume 2 | Issue 4 | October 2020 | ISSN 2630-5151 (Online)

ISSN 2630-5151
04
9 772630 515209

Volume 2 Issue 4 • October 2020 • ISSN 2630-5151 (Online)

Journal of Computer Science Research

Editor-in-Chief

Dr. Lixin Tao

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Volume 2 | Issue 4 | October 2020 | Page 1-35
Journal of Computer Science Research

Contents

ARTICLE

REVIEW

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
http://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Machine Learning Approach to Mobile Forensics Framework for Cyber Crime Detection in Nigeria

Ibrahim Goni* **Murtala Mohammad**

Department of Computer science, Adamwa state University Mubi, Nigeria

ARTICLE INFO

Article history
Received: 13 July 2020
Accepted: 31 July 2020
Published Online: 30 September 2020

Keywords:
Cyber crime
Machine learning
NCC Mobile forensics

ABSTRACT

The mobile Cyber Crime detection is challenged by number of mobile devices (internet of things), large and complex data, the size, the velocity, the nature and the complexity of the data and devices has become so high that data mining techniques are no more efficient since they cannot handle Big Data and internet of things. The aim of this research work was to develop a mobile forensics framework for cybercrime detection using machine learning approach. It started when call was detected and this detection is made by machine learning algorithm furthermore intelligent mass media towers and satellite that was proposed in this work has the ability to classified calls whether is a threat or not and send signal directly to Nigerian communication commission (NCC) forensic lab for necessary action.

1. Background

World is in the state of unique period of history. The current technological advancement will be among the global transformations remembered by humanity ever before. We live in a connected world of digital devices which include mobile devices, workstations, control systems, transportation systems financial systems, base stations, satellites of different interconnected networks, Global positioning system (GPS) with their associated e-services in which internet provide a platform for the connection of this devices worldwide [1]. The dramatic changes accompanying the current global transformation have significantly altered the lives of average citizens across the developed and the developing countries in less than a millennium [53]. This technological advancement has long been the subject of science fiction and now it has become a reality. However, this advancement led to the increase in cybercrime, threat and attack the aim of this research work was to developed a robust mobile forensic framework for cybersecurity in Nigeria.

1.1 Cyber Forensics

Digital forensics or computer forensics is a branch of forensic science that described the technique of forensics investigation of crimes that take place in a computer network or computer system has been used as weapon for cyber-attack or conduct a criminal activities but with the regardless of any digital device that has been used to perpetrate the crime [2]. In addition to [3] described cyber forensics as a sub-branch of computer security that uses software and predefined techniques which is aim at ex-

**Corresponding Author:*
Ibrahim Goni,
Department of Computer science, Adamwa state University Mubi, Nigeria;
Email: algonis1414@gmail.com

 DOI: https://doi.org/10.30564/jcsr.v2i4.2147

tracting evidences from any form of digital device and can be presented to a court of law for criminal and/or civil proceedings provided that it satisfy this three conditions; comprehensiveness, authenticity and objectivity. Furthermore, they were able to reveal that digital forensic report should be able to show important facts about evidence; like who obtained the evidence? Where the evidence captured and was stored and what happened to the evidence. [4] Identified the goals of digital forensics as identification of the evidence, document the crime, collection and preservation of the evidences, packaging the evidence and transporting the evidences in an untemper manner and suggested that cyber forensics depend on the collection and analysis of incident in order to explore understand and show complex security breaches that have by-passed security mechanism. In [46] "Digital forensics can be said to be a scientific framework in system development to identify, locate, retrieve, and analyze evidence from computers, computer storage media, and other electronic devices and present the findings in a court case". Digital forensics it was graphically represented by [50] as

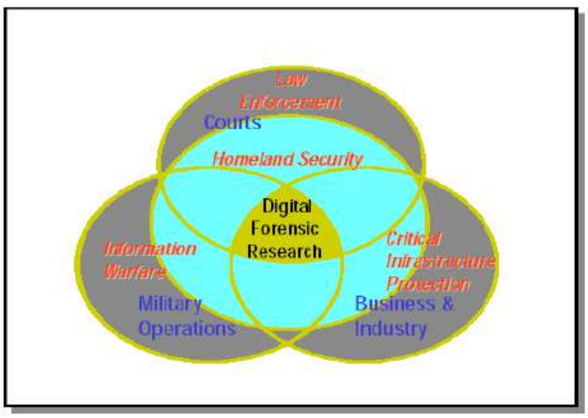

Figure 1. digital forensics science [50]

1.2 Cyber Space

Cyber space is recently considered a domain in science worth exploring, investigating and securing after lithosphere, hydrosphere, biosphere and atmosphere [3]. Cyber space has provide a dwelling environment and platform for technology today ranging from IoT, 5G, Fog, edge among others it confine to grow and expand and support all sorts of innovations in science and technology but the good and the bad. According to global cyber security index 2017 revealed almost half of the world population (3.5 billion users) are connected to the cyber space and they further estimated that there will be 12 billion device-to-device connections to the cyber space by 2020. It was also reported that, by the year 2020, 80% of adults on earth will have a smartphone [48], in addition to 49.7% of the total population are connected to Internet with the growth of 936% from 2000-2017 worldwide [49]. However threat and attack to this space is alarming day-by-day.

1.3 Cyber Crime

Cyber threats, attacks and breaches have become a normal incident in day-to-day life of internet users [6]. Furthermore, [10] ascertained that cyber terrorism is a conglomeration of cyber metrics and terrorism. They also believed that cyber terrorism is an illegal use of digital devices to perpetrate damage, intimidate or further terrorist's socio-economic political or religion agenda. Capgemini research institute reveal that in one case a hacker was able to access the GPS of 27000 vehicles which led to the shutdown the engine. "There is a pressing need for more research's and tool development to help digital investigator's obtain and analyses the increasing amount of digital evidence on smart phones, tablets, wearable devices, SatNav system, game console, automobile, IoT systems and cloud environment [5].

1.4 Cyber Security

Cyber security involves data security, network security, and computer security. It is also view by many researchers as an application of security preventions to provide a sense of confidentiality, integrity and availability, of data [28] but the major objectives of cyber security are prevention detection and reaction. Moreover, CIA revealed that the main goals of cyber security are confidentiality, integrity and availability. National cyber security center UK itemize ten steps to cybersecurity; network security, user education and awareness, malware prevention, removable media control, secure configuration, managing user privileges, incident management, monitoring and home and mobile working [35]. In addition to [33] revealed that AI and machine learning are the most important cyber tools for behavioral modeling, zero-day-attacks and advanced persistent threat.

1.5 Cyber Threats

According to US intelligent community in 2016 and 2017 there has been state sponsored cyber-attack against Ukraine and Saudi Arabia which resulted in targeting major infrastructure in both government and non-governmental organizations. They further indicated that known cyber security threats are classified under three headings; identity theft which includes; phishing, spoofing, masquerading, social engineering and password crackers. Unauthorized access includes; targeted data mining, backdoor and eavesdropping and tapping. Denial of service

  DOI: https://doi.org/10.30564/jcsr.v2i4.2147

(DoS, DDoS) includes; logic bomb and crypto-locker. Cyber security Ventures ascertained that in 2019 ransomware will damages as much as $11.5 billion [44]. "A ransomware attack targeting England's National Health Service affected 60 health trusts, 150 countries, and more than 200,000 computer systems" [45]. According to chief security officer of AT&T Bill O'Hern says "I see more than 100 billion potential vulnerability scans and probes across our global backbone every single day," [51].

1.6 Machine Learning Algorithm

Machine learning is a technique of using algorithm to parse data, learn from the data and make a decision, prediction, detection, classification, pattern recognition, responding and clustering based on the data collected. These algorithms are heavenly depend on the statistical and mathematical optimization. In broader sense machine learning algorithm are used in clustering, regression, (univariate & multivariate) anomaly detection, pattern recognition [34].

1.7 Supervised Learning

Supervised learning algorithms are machine learning algorithms that require datasets for training and testing the performance. This dataset has to be labeled and consist of features by which events or objects are defined as well as the expected outputs. The most common supervised learning algorithm are decision tree, logistic regression, support vector machine, relevance vector machine, random forest, K-NN, bagging neural networks, linear regression and naïve Bayes [33].

1.8 Unsupervised Learning

Unsupervised learning algorithm is a machine learning algorithm that required unlabeled datasets for training and testing the system performance the two major techniques used in unsupervised learning are principal component analysis (PCA) and clustering. The most common unsupervised learning algorithms are used especially in security is hierarchical, k-means, mixed model, DBSCAN, OPTIC, self-organizing mapping, Bolzan machine, auto encoder, adversarial network [34].

2. Related Work

Machine learning algorithms are recently applied to the following area of cyber security as in network security, data security, end-point security, identity access security, cloud security, IoT security, Fog security, but majority of the security systems depend on the detection, prediction and response. In [7] he explored the used of clustering algorithms such as K-means hierarchical clustering, k-means kernel, latent drichlet allocation and self-organizing mapping techniques for forensics analysis using text clustering in the large volume of data. [8] Presented a robust forensics analysis method using memetic algorithm. [9] Revealed how artificial intelligence techniques are applied to cyber-attacks security breaches. Machine learning algorithm was used to classified malware in android system in [16]. Machine learning and deep learning algorithm are combined and used for cyber security system in [15]. Machine learning algorithms are also applied to intrusion detection system in [14] and [52]. The researches of [13] systematic survey on the researches that combine machine learning algorithm and data mining to cyber security. In [12] presented how effectiveness of machine learning and deep learning in the feature of cyber security. Many surveys reviews and systematic reviews are conducted in the application of machine learning, deep learning and artificial intelligence techniques to cyber security, attack, intrusion detection system, network security as in [18,19,20-22]. Machine learning algorithm was also used to study cyber security in [29]. Security Framework was designed by [30] using fuzzy logic. [53] Highlight the role of intelligent system and artificial intelligences in addressing the challenges of cyber security but they didn't illustrate the framework on how to implement the system.

Furthermore, machine learning algorithms deep learning algorithms are applied in intrusion detection systems as in the research of [23] presented that machine learning based system can be used to detect intrusion for software defined networks. [24] Presented an extensive survey on anomaly based intrusion detection system. [25] Applied machine learning algorithm to intrusion detection in mobile cloud in a heterogeneous clients networks. In the work of [26] hybrid intrusion detection system for cloud computing. [27] They used machine learning algorithm to provide a roadmap for industrial network anomaly detection. Anomaly detection system for automobile network was presented by [31]. Deep neural network and fuzzy logic are used to identify abnormality in network traffic [32]. A systematic survey was made by [40] on the techniques that are used for malware detection, while [41] used APIs and machine learning algorithm to detect malware in android. In [42] they presented a general review on the malware detection in mobile devices based on parallel and distributed network. [43] They made a comparative analysis between the used of static, dynamic and hybrid technique to malware detection. Forensics analysis was also made on WhatsApp messenger to identify those that are using the application to perpetrate a crime or do illegal business as in the research of [36-39]. In [47] digital forensics framework was proposed and made a comparative analysis with other framework made with no AI techniques however, there framework has no instant detection and sending signals as compare to our proposed framework [17]. also explore extensively the roles of artificial intelligence, machine learning and deep learning algorithm to cyber space.

3. Method

The flow chart below illustrates the forensic procedure proposed in this research work, which serve as an intelligent algorithm to detect cyber threat with the help of government organization called Nigerian communication commission (NCC), in which they help the security agencies with the appropriate information for the next cause of action.

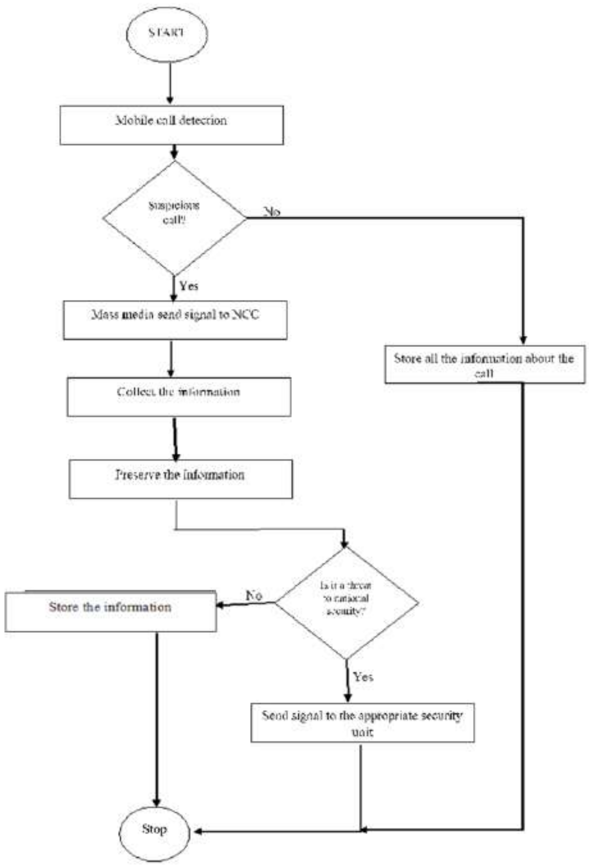

Figure 2. System flow chart

4. System Architecture

The system architecture or proposed framework as shown in the figure 2 is an integration of all communication gadgets including satellite, base stations, mass medias, mobile devices, mobile communication company, and Nigerian communication commission (NCC), intelligent devices that use various hardware and software components as well as communications networks to provide monitoring and controlling instantly. The major contribution of this research work is that the framework would provide instant information about all communication going on, if the communication is a threat to national security then appropriate security agencies would response immediately before the plan executed or it escalate.

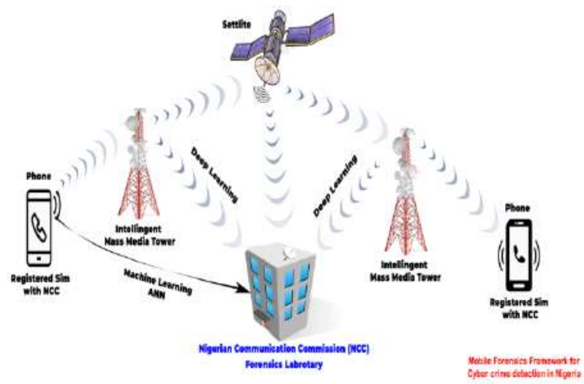

Figure 3. Mobile forensics framework for cyber-crime detection in Nigeria

5. Conclusion

In this research work a robust intelligent systems are connected and formulate mobile forensics framework for cybercrime detection in Nigeria.

References

[1] S. Shahzad. Protecting the integrity of digital evidence and basic human rights during the process of digital forensics. Ph.D. thesis Stockholm University, 2015.

[2] A. Abdalzim M., Amin B. A. M. a survey on mobile forensics for android smart phones IOSR Journal of computer engineering, 2015, 17(2): 15-19.

[3] M. Nickson K., Victor R.K., Venter H. Divergence deep learning cognitive computing techniques into cyber forensics Elservier Forensics Science international synergy 1, 2019, 61-67.

[4] A. Rukayat, Charles O. U., Florence A. O. computer forensics guidelines: a requirement for testing cybercrime in Nigeria now? 2017.

[5] E. Casey. Editorial-A sea change in digital forensics and incident response. Digital investigation evidence Elsevier Ltd 17, A1-A2, 2016.

[6] S, Ehsan, Giti J. Seminars in proactive artificial intelligence for cyber security consulting and research, Systematic cybernetics and informatics, 2019, 17(1): 297-305.

[7] A. Bandir, Forensics analysis using text clustering in the age of large volume data: a review. International

  DOI: https://doi.org/10.30564/jcsr.v2i4.2147

journal of advanced computer and application, 2019, 10(6): 72-76.

[8] Al-Jadir I., Wong K. W., Fing C.C., Xie H. Enhancing digital forensics analysis using memetic algorithm feature selection method for document clustering 2018 IEEE international conference on systems, Man and cybernetics, 2018: 3673-3678.

[9] Suid B., Preeti B. Application of artificial intelligence in cyber security. International journal of engineering research in computer science and engineering, 2018, 5(4): 214-219.

[10] O. David A., Goodness O., Etecte M.A. Unbated cyber terrorism and huma security in Nigeria. Asian social science, 2019, 15(11): 105-115.

[11] April. Threat start-SMS spam volume by month of each region SC magazine, 2014. Available online at: http//www.scmagazine.com/april-2014-threat-stats/slideshowz

[12] Apruzze G., Colajanni M. F., Ferreti L., Marchett M. on the effectiveness of machine learning for cyber security in 2018 IEEE international conference on cyber conflict, 2018, 371-390 .

[13] Buckza A. L., Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection IEEE communication survey and totorials, 2016, 18(2): 1153-1176

[14] Biswas S.K. intrusion detection using machine learning: A comparison study. International Journal of pure and applied mathematics, 2018 118(19): 101-114

[15] Y. Xin, Kong L., Liu Z., Chen Y., Zhu H., Gao M., Hou H., Wang C. Machine learning and deep learning methods for cyber security. IEEE Access, 2018, 6: 35365-35381.

[16] N. Miloseivic, Denghantanh A., Choo K.K.R. Machine learning aided android malware classification. Computer and electrical engineering, 2017, 61: 266-274.

[17] B. Geluvaraj, Stawik P.M., Kumar T. A. the future of cyber security: the major role of Artificial intelligence, Machine learning and deep learning in cyber space. International conference on computer network and communication technologies Springer Singapore, 2019, 739-747.

[18] H. Mohammed B., Vinaykumar R., Soman K. P. A short review on applications of deep learning for cyber security. 2018.

[19] M. Rege, Mbah R. B. K. Machine learning for cyber defense and attack. in the 7th International conference on data analysis, 2018, 73-78.

[20] D. Ding, Hang Q. L., Xing Y., Ge X., Zhang X. M. A survey on security control and attack detection for industrial cyber physical system. Neuro-computing. 2018, 275. 1674-1683.

[21] D. Berman S., Buczak A.L., Chavis J. S., Corbelt C.L. A survey of deep learning methods for cyber security information, 2018, 10(4).

[22] Y. Wang, Ye Z., Wan P., Zhao J. A survey of dynamic spectrum allocation based on reinforcement learning algorithms in cognitive radio network. Artificial intelligence review. 2019, 51(3): 413-506.

[23] A. Abubakar, Paranggono B. Machine learning based intrusion detection system for software defined networks. 7th International conference on Emerging security techniques IEEE. 2017, 138-143.

[24] S.Jose, Malathi D., Reddy B., Jayaseeli D. A survey on anomaly based host intrusion detection system. Journal of physics. Conference series, 2018, 1000(1).

[25] S. Dey, Ye Q., Sampalli S. A Machine learning based intrusion detection scheme for data fusion in mobile cloud involving heterogeneous clients network. Information fussion, 2019, 49: 205-215.

[26] P. Deshpande, Sharma S.C., Peddoju S.K., Junaid S. HIDS: a host based intrusion detection system for cloud computing environment. International journal of system assuarance engineering and management. 2018, 9(3): 567-576.

[27] M. Nobakht, Sivaraman V., Boreli R. A host-Based Intrusion detection and mitigation framework for smart IoT using open flow in 11th International conference on availability reliability and security IEEE. 2016, 147-156.

[28] A. Meshram, Christian H. Anomaly detection in industrial networks using machine learning: A road map. Machine learning for cyber physical system Springer Berlin Heldelberg. 2017, 65-72.

[29] R. Devakunchari, Souraba, Prakhar M. A study of cyber security using machine learning techniques. International journal of innovative technology and exploring engineering, 2019, 8(7): 183-186.

[30] E. Alison N. FLUF: fuzzy logic utility framework to support computer network defense decision making IEEE. 2016.

[31] A. Taylor, Leblanc S., Japkowicz N. Anomaly detection in auto-mobile control network data with long short term memory network in data science and advance analytics. IEEE international conference, 2016, 130-139.

[32] O. Amosov S., Ivan Y.S., Amosovo S.G. Recognition of abnormal traffic using deep neural networks and fuzzy logic. International Multi-conference on industrial engineering and modern technologies IEEE. 2019.

[33] M. Gyun L. Artificial Intelligence for development

series: Report on AI and IoT in Security Aspect. 2018.

[34] L. Matt. Rise of machine: machine learning & its cybersecurity applications, NCC group white paper, 2017.

[35] National cyber security center UK, www.ncsc.gov.uk

[36] A. Nuril, Supriyanto. Forensic Authentication of WhatsApp Messenger Using the Information Retrieval Approach. International Journal of Cyber Security and Digital Forensics (IJCSDF), 2019, 8(3): 206-212.

[37] A Marfianto, I Riadi. WhatsApp Messenger Forensic Analysis Based on Android Using Text Mining Method. International Journal of Cyber Security and Digital Forensics (IJCSDF), 2018, 7(3): 319-327.

[38] N Anwar, I. Riadi. Forensic Investigative Analysis of WhatsApp Messenger Smartphone Against WhatsApp Web-Based, Journal Information Technology Electromagnetic Computing and Information, 2017, 3(1): 1-10.

[39] S. Ikhsani, C. Hidayanto, Whatsapp and LINE Messenger Forensic Analysis with Strong and Valid Evidence in Indonesia. Tek. ITS, 2016, 5(2): 728-736.

[40] M. Ashawa, S. Morris. Analysis of Android Malware Detection Techniques: A Systematic Review. International Journal of Cyber Security and Digital Forensics (IJCSDF), 2019, 8(3): 177-187.

[41] W. Songyang, Wang, P., Zhang, Y. Effective detection of android malware based on the usage of data flow APIs and machine learning: Information and Software Technology, 2016, 75: 17-25.

[42] Anastasia, S., Gamayunov, D.: Review of the mobile malware detection approaches: Parallel, Distributed and Network-Based Processing (PDP). In: Proc. IEEE 23rd Euro micro International Conference, 2015: 600-603.

[43] D. Anusha, Troia, F. D., Visaggio, C. A., Austin, T. H., Stamp, M.: A comparison of static, dynamic, and hybrid analysis for malware detection. Journal of Computer Virology and Hacking Techniques, 2017, 13(1): 1-12.

[44] S. Morgan. Cyber security Business Report, 2017. Retrieved from CSO: https://www.csoonline.com/article/3237674/ransomware/ransomware-damage-costs-predicted-to-hit-115b-b y-2019.html

[45] R. Collier. NHS Ransomware attack spreads worldwide. CMAJ, 2017, 189(22), 786-787. https://doi.org/10.1503/cmaj.1095434

[46] H. Trisnasenjaya, I. Riadi Forensic Analysis of Android-based WhatsApp Messenger Against Fraud Crime Using The National Institute of Standard and Technology Framework. International Journal of Cyber Security and Digital Forensics (IJCSDF), 2019, 8(1): 89-97.

[47] H. Parag Rughani. Artificial Intelligence Based Digital Forensics Framework. International Journal of Advanced Research in Computer Science, 2017, 8(8): 10-14.

[48] Current State of Cybercrime, RSA Whitepaper, 2016.

[49] World Internet Users and 2017 Population Stats, accessed from: http://http://www.internetworldstats.com/stats.html

[50] R. Mark. Computer forensics: Basics. Lecture note Purdue University, 2004.

[51] LightReading. AT&T's Gilbert: AI Critical to 5G Infrastructure. September, 2018.

[52] I. Goni, Ahmed L. Propose Neuro-Fuzzy-Genetic Intrusion Detection System. International Journal of Computer Applications, 2015, 115(8): 1-5.

[53] Y. Harel, Irad Ben Gal, and Yuval Elovici. Cyber security and the role of intelligent systems in addressing its challenges. ACM Transaction on Intelligent System Technology, 2017, 8(4): 1-12.

  DOI: https://doi.org/10.30564/jcsr.v2i4.2147

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
http://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Researching the Research: Applying Machine Learning Techniques to Dissertation Classification

Suzanna Schmeelk[1*] Tonya L. Fields[2] Lisa R. Ellrodt[2] Ion C. Freeman[2] Ashley J. Haigler[2]

1. St. John's University, United States
2. Pace University, United States

ARTICLE INFO

Article history
Received: 24 July 2020
Accepted: 10 August 2020
Published Online: 30 September 2020

Keywords:
Machine learning
Natural language processing (NLP)
Abstract vs fulltext dissertation analysis
Industry-based
Dissertation research classification
GraphLab Create library
IBM Watson Discovery

ABSTRACT

This research examines industry-based dissertation research in a doctoral computing program through the lens of machine learning algorithms to determine if natural language processing-based categorization on abstracts alone is adequate for classification. This research categorizes dissertation by both their abstracts and by their full-text using the GraphLab Create library from Apple's Turi to identify if abstract analysis is an adequate measure of content categorization, which we found was not. We also compare the dissertation categorizations using IBM's Watson Discovery deep machine learning tool. Our research provides perspectives on the practicality of the manual classification of technical documents; and, it provides insights into the: (1) categories of academic work created by experienced fulltime working professionals in a Computing doctoral program, (2) viability and performance of automated categorization of the abstract analysis against the fulltext dissertation analysis, and (3) natual language processing versus human manual text classification abstraction.

1. Introduction

This research classifies industry-based doctoral research through the lens of machine learning algorithms to examine what Pace University's industry-geared doctoral students are researching. The Pace University Doctor of Professional Studies (DPS) in Computing began in the year 2000 [1,2]. The doctoral program is designed for experienced full-time working professionals (EFWPs) to study on campus with their faculty and advisers during the weekends. This research examines the first 114 dissertations that were successfully defended in the DPS program to understand industry trends and research needs through the dissertation topics. We employed the IBM Watson Discovery deep learning tool as well as Apple Turi's Graphlab Create in a Jupyter notebook running on an Amazon Web Services (AWS) Elastic Cloud (EC2) instance to classify the full-text of the DPS dissertations. This work extends the TF-IDF classification work of Ellrodt et al. [3,4], Freeman et al. [5], and Haigler et al. [6]; and, this research extends the EFWP research of Haigler et al. [7]. As the aforementioned publications were mutually exclusive studies of dissertation abstracts and fulltext dissertations, this research extends the work to compare the analysis methodologies give the performance differences from analyzing a page of text (dissertation abstract) versus analyzing over a hundred pages of text (dissertation fulltext).

**Corresponding Author:*
Suzanna Schmeelk,
St. John's University, United States;
Email: schmeels@stjohns.edu

1.1 Problem Statement

Semantically processing and deriving meaning from text are open research problems. Within the subset of natural language processing problems, there remains open questions about categorizing text. There has been very little meta-research on research text itself, specifically dissertations and theses. In this paper, we analyze the differences between manual and machine classifications of doctoral abstracts and full-text dissertations to understand what topics are being researched by senior-level and experienced fulltime working professionals.

1.2 Review of Literature

Employing machine learning to examine text has been evolving since the early 2000s. Textual-based machine learning has been successfully deployed in many computing fields such as computer security, networking, human computer interaction, medicine, and law.

Fautsch and Savoy [10] showed that adapting term frequency inverse document frequency (TF-IDF) is useful for domain specific information retrieval.

1.2.1 Building Recommender Systems via NLP

A thrust of literature which highly employs TF-IDF is text-based recommender systems. Duan, Gui, Wei, and Wu [11] proposed a personalized resume TF-IDF based recommendation algorithm to help job seekers find relevant jobs and enterprises find relevant talent. Yuan, and Zhang [12] employed TF-IDF classification within a recommendation system for seasonal events based on marketplace inventory.

1.2.2 Machine Learning for Domain Specific Information Retrieval

Yao, Mao Luo [13] proposed a new convolutional networks for text classification. They build a single text graph for a corpus based on word co-occurrence and document word relations They, then, learn a Text Graph Convolutional Network for the corpus. Their work showed promise of less training data in text classification.

Kumar, Alshehri, AlGhamdi, Sharma, and Deep [14] built and trained an artificial neural network (ANN) to detect skin cancer. Their work suggests that DE-ANN is best compared among other traditional classifiers in terms of detection accuracy of 97.4%.

Sinoara, Camacho-Collados, Rossi, Navigli, and Rezende [15] present a natural language processing approach based on embedded representations of words and word sense. Their approach results in semantically enhanced and low-dimensional representations.

Aggarwal, Rani, and Kumar [16] employ machine learning to authenticate license plates. Their method correctly captures the license plates with good performance metrics of 93.34% accuracy (e.g. detection rate and false positive rate).

1.2.3 Recent TF-IDF/K-Means Text Research

Text classification has become an effective means to discover trends in text. Yung [17] employed k-means with TF-IDF to explore all the Queens Memory program`s 400+ oral history interviews collected in Queens, New York. Frymire [18] employed k-means with TF-IDF to explore the Twitter feeds of the Social Movement #me-too.

1.2.4 Dissertation Text Classification

Ellrodt et al. [3,4], Freeman et al. [5] and Haigler et al. [6] examined text classification of these doctoral dissertations. Ellrodt et al. [3,4] examines the abstracts from these 114 dissertations through the lens of machine learning with natural language processing techniques. The goal was to learn about topic categories to understand what the student dissertation topics were and to cluster them to recognize different patterns. Freeman et al. [5] examined the same 114 dissertation abstracts through IBM Watson and additional machine learning algorithms. Haigler et al. [7] examined and reported on the clustering for the full text of 98 (of the 114) dissertations; however, they focused on a smaller cluster count than this research.

1.2.5 EFWP Dissertation Research

Haigler et al. [6,7] explored research topics selected for EFWPs to help understand the research categories and trends. Haigler et al. [6] reported on educational needs for EFWPs obtained from IRB-approved surveys of Pace University's DPS program participants, which was further discussed in Haigler et al. [7]

2. Methodology/Methods

This research performs meta-research on research through both manual and machine classifications. Specifically, we examine all the dissertations defended in Pace University's Doctor of Professional Studies (DPS) from the program inception in the early 2000s until 2018. We analyzed each defended dissertation both through the abstract (e.g. approximately one page) and a full-text analysis (e.g. approximately 150 pages) to gain insights if automated NLP processing on abstracts is an adequate categorical measure for dissertation content over fulltext analysis. As fulltext analysis of approximately 150 pages requires large quantities of memory and storage, we compared the results of

DOI: https://doi.org/10.30564/jcsr.v2i4.2230

the two distinct analysis methodologies. Additionally, we discuss both the natural language processing (NLP) performed on the dissertations as well methods we used to classify the texts.

2.1 Natural Language Processing (NLP)

One of the goals of artificial intelligence is to develop semantic context for human language; the machine learning field of natural language processing pursues this goal for text documents. The seminal textbook on NLP is written by Jurafsky and Martin [19]. At a high-level, the text book describes almost every use case of NLP.

In keeping with the techniques described by Jurafsky and Martin [19], we examined both clustering the dissertation and abstract text with and without cleaning. To clean the data, we used the Python Natural Language Toolkit (*NLTK*) learn library. The tasks included in the data cleaning were for text standardization. First, we made everything in the text lower case. Then, we removed markup symbols, section formatting markers, special characters, stop words, and punctuation from the text. Lastly, we ran classifications on both stemmed and non-stemmed words to see clustering differences. Stemming processes words by reducing inflected and derived words to their root or base language form. It removes different word variations so that the actual word usage is standardized throughout the text.

2.2 Dissertation Classification

One of the goals of artificial intelligence is to develop semantic context for human language; the machine learning field of natural language processing pursues this goal for text documents.

2.2.1 Manual Classification

To manually classify the dissertation abstracts, we divided them amongst 5 people. Each person was tasked with reading some assigned subset of abstracts and determining the correct category for each work. This first pass of the abstracts produced 176 categories and much debate, as discussed in Ellrodt et al. [3,4]. In order to reduce the categories an iterative approach leveraging domain knowledge would have been needed. The researchers found the human iterative approach excessively time consuming and had trouble exercising stable categories. This suggested that without adequate domain knowledge, it would be difficult to communicate these categories in any meaningful way; the whole process seemed unlikely to produce worthwhile results and was abandoned.

In most human topic assignments, the people involved have training in the specific distinctions between the types of documents they are likely to run across. However, developing ad hoc topics to distinguish similar products in a corpus requires a level of sophistication and ability to form consensus that our workers did not let emerge.

2.2.2 Machine Learning Classification

This research uses the approach of applying K-means clustering analysis to term-frequency inverse-document-frequency (TF-IDF) coding of DPS dissertations. We then compare the TF-IDF analysis to the topic analysis of their abstracts in IBM Watson Discovery.

2.2.3 TF-IDF

TF-IDF for "term frequency - inverse document frequency" is a characterization tool for text documents. Each abstract is regarded as a "bag of words", as if the meaning of each abstract were implicit in the words used in that abstract and the order of those words were unimportant. Each individual word is then deemphasized according to how often it occurs in the collected abstracts overall. The formula for TF-IDF can be seen in Figure 1.

TFIDF

For a term i in document j:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

tf_{ij} = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

Figure 1. TF-IDF Formula

2.2.4 K-Means Clustering

"K-means" is an unsupervised learning algorithm that solves clustering problem. It is one of the most basic clustering algorithms and works with numeric data only. The algorithm is composed of the following steps: (1) Pick a number (K) of cluster centers at random.; (2) Assign every item to its nearest cluster center.; (3) Move each cluster center to the mean of its assigned items.; (4) Repeat the prior two steps until convergence. After initialization, the k-means algorithm iterates between the following two steps: (1) Assign each data point to the closest centroid, as seen in Figure 2., and,

$$z_i \leftarrow \text{argmin}_j \|\mu_j - \mathbf{x}_i\|^2$$

Figure 2. Assign each data point to the closest centroid

 DOI: https://doi.org/10.30564/jcsr.v2i4.2230

(2) Revise centroids as the mean of the assigned data points, as seen in Figure 3.

$$\mu_j \leftarrow \frac{1}{n_j} \sum_{i:z_i=j} \mathbf{x}_i \Bigg|$$

Figure 3. Revise centroids as the mean of the assigned data points

The algorithm has convergence when the cluster assignments no longer change. There is no assurance that the cluster assignments are optimal using K-means. Clusters will be reasonable, however may not be robust to different start point selection. In k-means, the number of clusters must be selected beforehand. The algorithm is very sensitive to outliers. It can be proved that the running of the algorithm will always terminate. How can we tell if the k-means algorithm is converging? We can look at the cluster assignments and see if they stabilize over time. In fact, we'll be running the algorithm until the cluster assignments stop changing at all. To be extra safe, and to assess the clustering performance, we'll be looking at an additional criteria: the sum of all squared distances between data points and centroids, as defined in Figure 4.

$$J(\mathcal{Z}, \mu) = \sum_{j=1}^{k} \sum_{i:z_i=j} \|\mathbf{x}_i - \mu_j\|^2$$

Figure 4. Assessing convergence

2.2.5 IBM Watson Discovery

IBM Watson Discovery is a cloud platform which ingests and standardizes user data, providing services such as sentiment analysis, named entity extraction and concept tagging through an API [20]. In addition to providing chatbots and other workflow enhancements, Watson Discovery provides Smart Document Understanding, a clustering solution [21].

3. Results

We applied K-means analysis to term-frequency inverse-document-frequency (TF-IDF) coding of 114 of Pace University's DPS dissertations and then compare the output of that analysis to the topic analysis of their abstracts in IBM Watson Discovery.

3.1 IBM Watson Discovery Classification

We chose to examine the dissertations through machine learning using IBM Watson Discovery using both standard classification as well as enriched classification. This analysis extends the work of Ellrodt et al. [3,4] and Freeman et al. [5], where the dissertation abstracts were classified using a TF-IDF algorithm.

The IBM Watson Discovery system produced the top six enriched text concepts show in Table 1 as: Scientific method (14), Algorithm (11), Management (11), Computer (10), Mathematics (10), and Agile software development (9). The enriched text key-words were: Research (29), Dissertation (16), Model (11), Study (11), Approach (10), and Addition (9). None of this second list of individual words relate to any specific topic in the computing field, which is a comparative weakness of this approach to the TF-IDF analysis.

The system sentiment analysis examines sentence word choices with respect to sentiments. On the default configuration IBM labeled the 114 dissertation abstracts as follow: 85% (97) positive, 3% (3) neutral, and 12% (14) negative based on the word used. Interestingly, industry word choices like "false negative" triggered the negative analysis categorization.

Table 1. Watson Topic Analysis for Abstracts

Topic	Assigned Papers
Scientific method	14
Algorithm	11
Management	11
Computer	10
Mathematics	10
Agile software development	9
Software engineering	9
Education	8
Internet	8
Computer program	7
Computer science	7
Waterfall model	7

3.2 IBM Watson Full-Text Analysis

Using IBM Watson Discovery, we performed full-text PDF analysis, extending the work of Freeman et al. [5] and Haigler et al. [6]. Full text PDFs of the 114 dissertations were uploaded to Watson and evaluated via the basic Watson Discovery Natural Language Understanding (NLU) engine. The basic engine yields results such as sentiment analysis, related concepts and top entities.

The results of related concepts are listed using enriched text produced by Watson. The top six listed are: Software engineering (13), Agile software development (12), Computer (12), Software development (12), Biometrics (9),

DOI: https://doi.org/10.30564/jcsr.v2i4.2230

and Extreme Programming (9). This resulted in a different categorization of the dissertations than categorization of the abstracts alone.

IBM sentiment analysis labeled the 114 dissertation full text as follow: 82% positive, 2% neutral, and 16% negative.

3.3 Amazon EC2 Full-Text Analysis: K-Means with TF-IDF

The selection of k - the cluster count - is the primary hyperparameter for a k-means model. In order to select k, we let k vary from one to 25 - more than a fifth of the document count - and observed the heterogeneity. We assess the heterogeneity of a single cluster as the sum of all squared distances between data points in that cluster and its centroids.

Heterogeneity should decrease more quickly after the optimal point, at which point the model is overfit. The heterogeneity is plotted in Figure 1.

This plot does not have a clear point after which heterogeneity decreases, and so some further treatment is used to expose the optimal cluster count k. Therefore, we apply a log transformation and de-slope the output to look for discontinuity.

It is evident in Figure 5 that heterogeneity is discontinuous at a cluster count of eight and above. Therefore, we treat eight as the optimal value of k across all four experiments, which is unique from the work of Ellrodt et al. [3,4], Freeman et al. [5] and Haigler et al. [6,7].

Table 2 shows the keywords and cluster labels for k = 8.

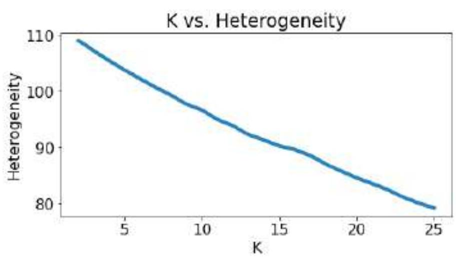

Figure 5. Heterogeneity v cluster count of non-stemmed full text dissertations

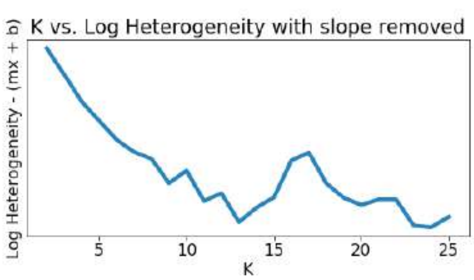

Figure 6. De-sloped log heterogeneity by cluster count of non-stemmed full-text dissertations

The results of running the TF-IDF/K-means clustering for eight clusters, produced the results seen in Table 2. Depicted in the table are the following: the cluster number, number of dissertations clustered into the category, top five key-words identified by the algorithms, our human interpretation of the category and the dissertation papers clustered in this category. Github [22] shows the abstract paper number title mapping given in the tables.

Table 2. K-Means with fulltext dissertations non-stemmed, k=8

#	*Instance Count*	*Top Five Keywords*	*Category*	*Paper Number*
1	17	cloud:0.168 compliance:0.068 security:0.067 hipaa:0.048 csp:0.047	Data security	16, 18, 19, 20, 22, 31, 49, 57, 66, 69, 72, 75, 77, 90, 99, 108
2	26	students:0.084 instructor:0.041 teaching:0.037 cics:0.037 abstraction:0.034	Education	2, 9, 10, 15, 24, 27, 32, 33, 35, 54, 55, 58, 63, 68, 71, 82, 83, 84, 89, 94, 102, 104, 105, 111, 112, 114
3	14	agile:0.172 team:0.116 retrospective:0.094 kms:0.068 pba:0.060	Agile software practices	3, 8, 26, 40, 42, 56, 62, 64, 67, 74, 76, 85, 93, 101
4	10	int:0.180 sa:0.128 annealing:0.071 fitness:0.070 patch:0.066	Optimization	1, 4, 6, 21, 36, 38, 39, 70, 103, 106
5	5	irs:0.219 impute:0.143 pottery:0.137 loyalty:0.123 recommender:0.084	Machine learning categorization	11, 51, 52, 88, 107
6	13	schematron:0.173 xml:0.164 owl:0.092 ontology:0.089 rdf:0.088	Ontology	7, 12, 28, 29, 37, 48, 53, 86, 87, 91, 92, 95, 98
7	19	keystroke:0.132 biometric:0.085 roc:0.057 classifier:0.052 svm:0.042	Biometrics	13, 23, 25, 30, 34, 41, 44, 46, 50, 59, 60, 65, 73, 78, 79, 80, 81, 97, 113
8	10	pda:0.097 channel:0.092 wireless:0.080 uumi:0.068 callback:0.061	Distributed software architecture	5, 14, 17, 43, 45, 47, 61, 96, 109, 110

The categories listed in Table 2 are meaningful human-assigned text labels to summarize the top five keywords extracted by the k-means on TF-IDF analysis. The categories we selected are discussed below in Table 3.

Table 3. Meaningful Human-Assigned Cluster Text Labels

• ***Data security*** is a current topic in the computing world. • ***Education*** is focus on the students, many of whom are trying to get credentials to advance an academic career. • ***Agile software practices*** are a focus of the program content. • ***Optimization*** is discussed in several classes and encouraged as a dissertation focus by faculty. • ***Machine learning categorization*** is another focus of the program class content. • ***Ontology*** is referred to in some courses and encouraged as a dissertation area. • ***Biometrics*** is a particular focus of research at Pace University. • ***Distributed software architecture*** is a mainstay of many students' work lives and supported by the program as a dissertation area.

Figure 7 is a visualization for the full non-stemmed dissertations cluster counts where k=8. In contrast, Figure 8 is a visualization for the non-stemmed abstracts cluster counts.

Cleaning data is a common methodology to improve the quality of natural language processing results. We employed the *NLTK* toolkit to clean the data (e.g. remove stop words, Porter stem, remove non-ASCII characters, etc.) Table 4 shows the keywords, and dissertations for eight clusters.

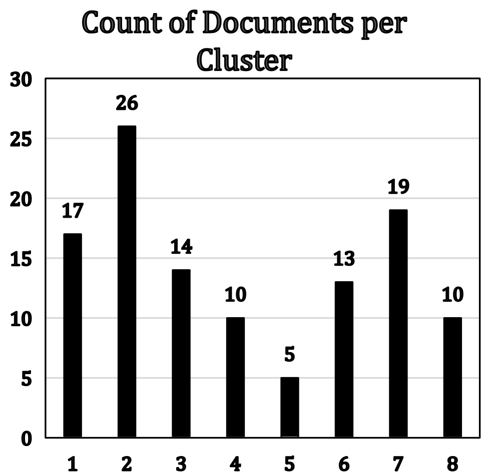

Figure 7. Counts of non-stemmed Full -Text Documents Per Cluster

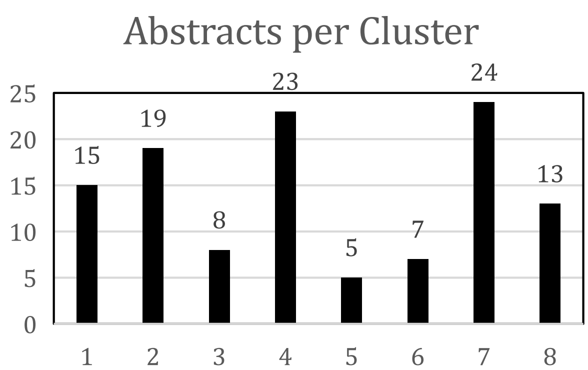

Figure 8. Counts of Abstracts per Cluster

Table 4. K-Means fulltext stemmed, k=8 where highlighted numbers changed from non-stemmed categories

#	*Instance Count*	*Top Five Keywords*	*Category*	*Paper Number*
1	14	cloud:0.225 csp:0.091 cic:0.069 packet:0.067 ip:0.063	Cloud Computing	**5, 16, 20, 36, 45, 47, 55, 63, 71, 72, 75, 90, 110, 114**
2	13	student:0.179 instructor:0.117 teach:0.090 game:0.078 cs:0.074	Education	10, 15, 27, **28**, **32**, 58, 68, **74**, 84, 89, 94, 102, 111
3	30	keystrok:0.094 biometr:0.069 imput:0.045 roc:0.039 distanc:0.031	Biometrics	**9, 11,** 13, **14, 21**, 23, **24**, 25, 30, 34, **44**, 46, **49**, 50, **51**, **54**, 59, 60, 65, 73, 78, 79, 80, 81, **83, 96**, 97, **101, 104**, 113
4	22	agil:0.109 team:0.088 retrospect:0.074 complianc:0.051 secur:0.046	Agile Software Development	8, **18, 19,** 26, **31**, 40, **52**, 56, 57, 62, 64, **66**, 67, **76, 77, 82**, 85, **87, 99, 100, 106, 108**
5	12	schematron:0.193 xml:0.190 schema:0.080 xs:0.069 recip:0.067	Data Validation	**3, 7, 12, 17, 29, 35, 48, 53, 61, 88, 92, 95**
6	12	int:0.163 sa:0.107 km:0.080 ler:0.073 ga:0.072	Optimization	1, 4, 6, 38, 39, **41, 42**, 70, **93**, 103, **105, 112**
7	3	pda:0.371 pervas:0.159 itamm:0.147 button:0.131 lotu:0.128	Human Computer Interaction	**22, 43, 109**
8	8	ontolog:0.158 rdf:0.122 owl:0.112 drug:0.104 pir:0.103	Medical Ontologies	**2, 33, 37, 69, 86, 91, 98, 107**

The heterogeneity for the stemmed full-text dissertations, to contrast with the above non-stemmed heterogeneity, is shown in Figure 9.

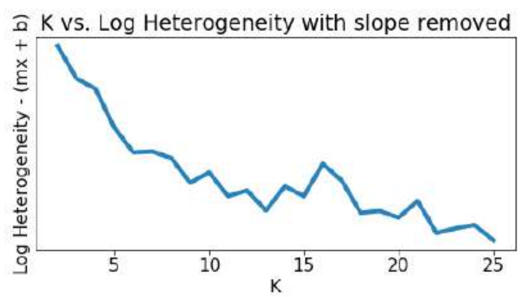

Figure 9. Heterogeneity v cluster count of stemmed full-text dissertations

The bolded dissertations listed in Table 4 show the category change from when a dissertation is categorized in a non-stemmed and stemmed-format. As we can see from the table, over half (i.e. 71 of 114) have changed categories after applying the stemming indicating that stemming is an essential standardization prior to categoriatzation.

Ellrodt et al. [1,2] and Freeman et al. [3] examined text classification of these same 114 doctoral dissertations; however, they report that the heterogeneity is discontinuous at a cluster

  DOI: https://doi.org/10.30564/jcsr.v2i4.2230

count of six and above. Therefore, the earlier abstract-only analysis treated six as the optimal value of k (i.e. categories).

Thus, to contrast to the clustering of eight-categories for the stemmed and non-stemmed 114 dissertations abstract, in Table 5, we clustered the non-stemmed dissertation abstracts into eight categories (i.e. k=8).

Table 5. K-Means with dissertation abstracts non-stemmed, k=8

#	*Instance Count*	*Top Five Keywords*	*Category*	*Paper Number*
1	15	keystroke:0.104 biometric:0.071 authentication:0.058 input:0.052 beta:0.052	Biometrics	4, 5, 16, 17, 34, 35, 61, 63, 70, 76, 80, 89, 91, 100, 101
2	19	agile:0.105 development:0.069 software:0.057 team:0.046 outsourcing:0.041	Agile Software Development	2, 14, 31, 39, 45, 46, 51, 52, 60, 62, 67, 69, 71, 74, 86, 96, 97, 102, 111
3	8	loyalty:0.079 user:0.073 sparse:0.065 capacity:0.059 unusual:0.057	Optimization	8, 15, 21, 22, 23, 41, 77, 104
4	23	security:0.051 components:0.037 cloud:0.037 application:0.033 requirements:0.033	Security	1, 11, 13, 19, 20, 24, 25, 32, 36, 40, 42, 47, 48, 50, 59, 73, 75, 78, 81, 87, 93, 113, 114
5	5	shape:0.159 pottery:0.143 images:0.102 classification:0.100 shapes:0.098	Machine Learning	28, 105, 107, 108, 110
6	7	students:0.124 erp:0.116 computer:0.077 schillinger:0.076 programming:0.070	Education	6, 7, 18, 29, 54, 90, 95
7	24	factors:0.059 cloud:0.044 genetic:0.040 algorithm:0.033 problems:0.033	Genetic Algorithms	1, 9, 12, 26, 27, 30, 37, 38, 43, 44, 49, 57, 58, 65, 66, 68, 79, 84, 85, 94, 99, 103, 106, 112
8	13	xml:0.164 documents:0.074 document:0.069 semantic:0.067 constraints:0.058	Ontologies	10, 33, 53, 55, 56, 64, 72, 82, 83, 88, 92, 98, 109

The heterogeneity for the non-stemmed abstract dissertations, to contrast with the above non-stemmed heterogeneity, is shown in Figure 10 and log heterogeneity in Figure 11.

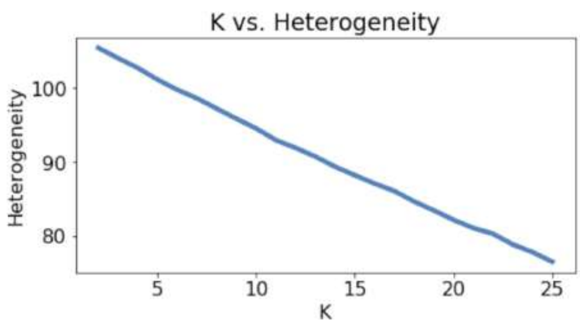

Figure 10. Heterogeneity v cluster count of non-stemmed abstract dissertations

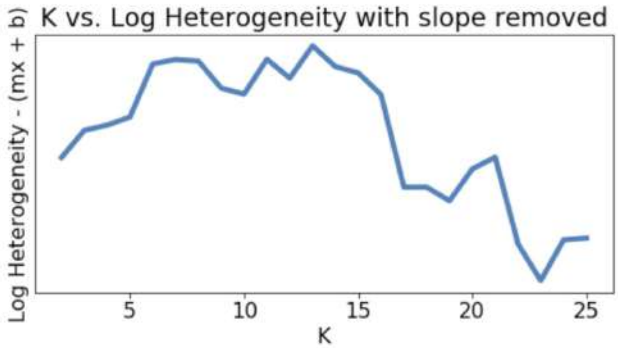

Figure 11. Developed log heterogeneity by cluster count of non-stemmed abstract dissertations

Thus, to contrast to the clustering of eight-categories for the stemmed and non-stemmed 114 dissertations dissertations, in Table 6, we clustered the stemmed dissertation abstracts into eight categories (i.e. k=8). The bolded dissertation abstract paper numbers changed categories from the stemmed fulltext categories.

Table 6. K-Means abstracts stemmed, k=8, where highlighted numbers changed from stemmed full categories

#	*Instance Count*	*Top Five Keywords*	*Category*	*Paper Number*
1	9	factor:0.094 shape:0.091 technolog:0.073 search:0.069 name:0.067	Machine Learning	**14, 25, 39, 44, 62, 63, 76, 78, 105**
2	18	secur:0.139 cloud:0.066 complianc:0.065 knowledg:0.057 risk:0.045	Cloud Computing Security	**8, 35, 52, 55, 58, 60, 65, 67, 71, 75, 77, 90, 100, 101, 108, 109, 110, 111**
3	21	agil:0.092 student:0.089 scienc:0.054 softwar:0.051 learn:0.048	Agile Software Development	**1**, **2**, **5**, 6, 7, 12, **13**, **27**, **29**, **30**, **43**, **49**, **59**, **66**, **69**, **85**, **87**, **92**, **94**, **99**, 102
4	12	algorithm:0.122 genet:0.086 problem:0.084 ann:0.083 optim:0.077	Genetic Algorithms	**10, 22, 24, 28, 42, 45, 47, 56, 74, 82, 84, 96**
5	7	estim:0.216 project:0.170 retrospect:0.107 elf:0.099 binari:0.092	Project Mangement	**21, 50, 64, 73, 81, 83, 93**
6	17	keystrok:0.111 biometr:0.088 featur:0.084 text:0.064 classif:0.056	Biometrics	**0**, **3**, **4**, **15**, 26, **32**, **33**, **41**, 61, 68, 72, 80, **86**, 88, **97**, 103, **104**
7	2	insur:0.281 cybersecur:0.256 gi:0.216 risk:0.201 financi:0.193	Security	**40, 57**
8	26	xml:0.088 busi:0.047 site:0.043 agent:0.042 constraint:0.042	Data Validation	**9, 11, 16, 17, 18, 19, 20, 23, 31, 34, 36, 37, 38, 46, 48, 51, 53, 54, 70, 79, 89, 91, 95, 98, 106, 107**

The heterogeneity for the stemmed full-text dissertations, to contrast with the above non-stemmed log heterogeneity, is shown in Figure 12.

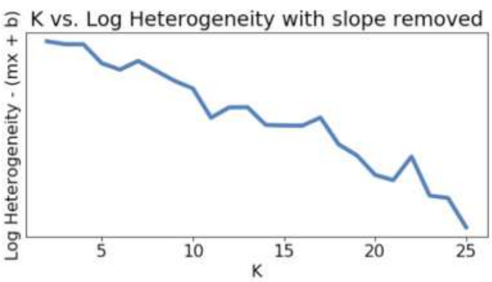

Figure 12. Developed log heterogeneity by cluster count of stemmed abstract dissertations

As we can see from Table 6 that all but eleven dissertations change categories indicating that abstract analysis alone is not a good indicator of content or dissertation category.

4. Discussion

Of the tools we examined, IBM Watson Discovery allowed for the quickest startup process. It easily ingested and categorized all 114 dissertations in less than 30 minutes from start to finish. In contrast, the k-means analysis using Turi's GraphLab Create TF-IDF coding was a more tedious process to set up the environment which required programming and debugging of the code, converting each dissertation into a text format, ingesting the dissertation, and finally running the algorithm on a large AWS EC2 server. Once running on the large server, the acual algorithm took approximately a minute to run.

Each of the machine learning algorithms produce different categories of dissertation research topics. The Watson Discovery system enriches the text using a natural language understanding module and the resulting categories shown in Table 1 are different from the categories produced by k-means as shown in Table 2, Table 4, Table 5, and Table 6. Overall, the dissertation abstracts appear to not be a good representation of the full dissertation. A future work study would involve an IRB-approved survey to ack the authors their own human interpretation for their own dissertation category.

In all cases we found that EFWP students tend to favor the emerging technologies they face in industry. The Pace University DPS program allows the student to select their research topic-then match an advisor versus the traditional Ph.D. program where students research tends to follow that of their academic advisor.

Future work involves further examination of the machine generated cluster category of each dissertation as compared to the authors actual intent. This work can further be extended to examine the year of the dissertation defense to determine if they are aligned with industry technology trends of the time.

5. Conclusion

There exists very little research on research itself from the perspective of text analysis. In this research we have performed cluster analysis on both the fulltext and abstracts for the first 114 dissertations defended in Pace University's DPS program to see what topics have been the doctoral focus of senior and experienced fulltime working professionals (EFWPs). We found that many students tend to focus their research on industry trends first; and, then, find an adviser. As such, the DPS dissertation research is typically different than the Ph.D. program model where students focus on their advisor's topics. We also showed that data preprocessing including stemming did slightly change the clusters identified by the machine learning algorithms. We also showed that fulltext analysis produces different categories than abstract analysis indicating that abstract analysis alone may not be sufficient for categorizing dissertation research. Future work such as examining longitudinal-trends, innovation, accountability, and automatic keyword generation can be further developed from our research. Lastly, we showed that machine learning on the abstract alone were not good indicators on dissertation content. As more and more text becomes digitally available, we must continue to develop methodologies to build semantic understandings from the available data.

Supplementary Data/Information

A full list of the dissertation full-text analysis mapping identifiers to their abstract-analysis identifiers can be found on GitHub [22].

References

[1] Susan M. Merritt, Allen Stix, Judith E. Sullivan, Fred Grossman, Charles C. Tappert, and David A. Sachs. Developing a professional doctorate in computing: a fifth-year assessment. In Working group reports from ITiCSE on Innovation and technology in computer science education (ITiCSE-WGR '04). ACM, New York, NY, USA, 2004: 42-46. DOI: http://dx.doi.org/10.1145/1044550.1041654

[2] Fred Grossman, Charles Tappert, Joe Bergin, and Susan M. Merritt. A research doctorate for computing professionals. Commun. ACM 54, 2011, 133-141. DOI: https://doi.org/10.1145/1924421.1924450

[3] L. R. Ellrodt, I. C. Freeman, A. J. Haigler, S. E. Schmeelk. Doctor of Professional Studies in Computing: A Categorization of Applied Industry Research. in 2018 IEEE Frontiers in Education Conference (FIE), 2018: 1-6.

  DOI: https://doi.org/10.30564/jcsr.v2i4.2230

[4] Lisa R. Ellrodt, Ion C. Freeman, Ashley J. Haigler, Lynne E. Larkin, Suzanna E. Schmeelk, Ronald G. Williams. Pace University DPS in Computing Studies: A Categorization of Applied Industry Research. The Michael L. Gargano 16th Annual Research Day. Pace University. May 2018 Retrieved from: http://csis.pace.edu/~ctappert/srd/index.htm

[5] I. Freeman, A. Haigler, S. Schmeelk, L. Ellrodt, T. Fields. What are they Researching? Examining Industry-Based Doctoral Dissertation Research through the Lens of Machine Learning. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018: 1338-1340.

[6] Ashley J. Haigler, Suzanna E. Schmeelk, Tonya L. Fields, Lisa R. Ellrodt, Ion C. Freeman. Employing Machine-Learning to Understand Research Trends of Full-Time Working Professionals. The Michael L. Gargano 17th Annual Research Day. Pace University, 2019. Retrieved from: http://csis.pace.edu/~ctappert/srd2019/schedule.htm

[7] Ashley J. Haigler, Suzanna E. Schmeelk, Tonya L. Fields, Lisa R. Ellrodt, Ion C. Freeman. Educational Needs in Computing of Experienced Full-Time Working Professionals. Future of Education. Florence, Italy, 2019.

[8] Dhillon, Paramvir, Amandeep Walia. A Study on Clustering Based Methods. International Journal of Advanced Research in Computer Science, vol. 8, no. 4, May 2017, 8(4): 1-5, 0967-5697.

[9] J. A. Hartigan, M. A. Wong. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society, 1979, Series C. 28(1): 100-108.

[10] Claire Fautsch, Jacques Savoy. Adapting the tf idf vector-space model to domain specific information retrieval. In Proceedings of the 2010 ACM Symposium on Applied Computing (SAC'10). Association for Computing Machinery, New York, NY, USA, 2010, 1708-1712. DOI: https://doi.org/10.1145/1774088.1774454

[11] Duan, Xiaolin Gui, Mingan Wei, You Wu. A Resume Recommendation Algorithm Based on K-means++ and Part-of-speech TF-IDF. In Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM 2019). Association for Computing Machinery, New York, NY, USA, 2019, Article 50: 1-5. DOI: https://doi.org/10.1145/3358331.3358381

[12] Ted Tao Yuan, Zezhong Zhang. Merchandise Recommendation for Retail Events with Word Embedding Weighted Tf-idf and Dynamic Query Expansion. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 2018, 1347-1348. DOI: https://doi.org/10.1145/3209978.3210202

[13] Yao, L., Mao, C., Luo, Y. Graph convolutional networks for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 7370-7377.

[14] Kumar M., Alshehri M., AlGhamdi R., Sharma P., Deep V., 2020. A DE-ANN Inspired Skin Cancer Detection Approach using Fuzzy C-Means Clustering. Mobile Network and Applications, 2020. DOI: https://doi.org/10.1007/s11036-020-01550-2

[15] Sinoara, R.A., Camacho-Collados, J., Rossi, R.G., Navigli, R., Rezende, S.O. Knowledge-enhanced document embeddings for text classification. Knowledge-Based Systems, 2019, 163: 955-971.

[16] Aggarwal A., Rani A., Kumar M. A Robust Method to Authenticate License Plates using Segmentation and ROI Based Approach. Smart and Sustainable Built Environment, 2019. DOI: https://doi.org/10.1108/SASBE-07-2019-0083

[17] Stephanie Yung. All the Queens Voices: An Oral History, Visualized. A data visualization of Queens Memory program`s 400+ oral history interviews collected in Queens, New York." Thesis, Parsons School of Design at The New School. New York, 2019. Retrieved from: http://parsons.nyc/thesis-2019/#13

[18] Ellie Frymire. An Exploration of the Social Movement #metoo. Thesis, Parsons School of Design at The New School. New York, 2018. Retrieved from: https://parsons.nyc/thesis-2018/#4

[19] Daniel Jurafsky and James H. Martin. Speech and Language Processing (2nd Edition). Prentice-Hall, Inc., USA, 2009.

[20] C. Forrest. IBM launches Watson Discovery Service for big data analytics at scale. TechRepublic. Retrieved from: https://www.techrepublic.com/article/ibm-launches-watson-discovery-service-for-big-data-analytics-at-scale/

[21] IBM. Corporation. Watson Discovery. IBM. Retrieved from: https://www.ibm.com/watson/services/discovery/

[22] Suzanna E. Schmeelk, Tonya L. Fields, Lisa R. Ellrodt, Ion C. Freeman, Ashley J. Haigler. GitHub: JSCR Machine Learning: Researching the Research, 2020. Retrieved from: https://github.com/schmeelk/jcsr_ml_researchingtheresearch

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
http://ojs.bilpublishing.com/index.php/jcsr

REVIEW

Design of Intelligent Home Security Alarm System under STC89C51 Single Chip Microcomputer

Ruizheng Wang* **Xuan Liu** **Xinhao Huang**

Artificial Intelligence College, North China University of Science and Technology, Tangshan, Hebei, 063000, China

ARTICLE INFO

Article history
Received: 15 July 2020
Accepted: 15 July 2020
Published Online: 30 September 2020

Keywords:
STC89C51 MCU
Intelligent
GSM
Pyroelectric infrared
Security alarm system

ABSTRACT

In order to improve the security of home residence, this paper studies and designs an intelligent home security alarm system, using STC89C51 single chip microcomputer as the main controller of the security system, and real-time monitoring by controlling the human pyroelectric infrared sensor and smoke sensor in the case of strangers invading the security range and showing signs of fire. Once the abnormal situation is found, the intelligent home security alarm system will start the acousto-optic alarm prompted by the LED lamp and pass through the information processing system of the GSM module. Send an abnormal text message to the user of the security system at the first time face, and finally realize the purpose of modern intelligent home security alarm.

1. Introduction

Based on STC89C51 single chip microcomputer, this paper makes a research and analysis on the design of intelligent home security alarm system. The purpose of this intelligent system is to provide users with an intelligent home security alarm tool which is convenient to use, easy to operate and very practical. This intelligent security alarm system can send fire accident message to users in the first time of accident, especially in the initial stage of home fire.

2. Brief Introduction of this Intelligent Home Security Alarm System

An intelligent home security alarm system designed in this paper is very practical, safe and sensitive. The system mainly includes the following modules: pyroelectric infrared sensor, smoke sensor, GSM,alarm system, which achieves the basic functions of preventing theft, preventing fire and remote alarm and preventing accidental text messages. [1]A thermoluminescence infrared sensor in the module can identify the infrared rays emitted by the human body very well. When a stranger enters the range of the security system, the acousto-optic alarm system will run automatically. Through the information processing of the GSM module, the alarm system can be quickly sent to the users of the intelligent home security alarm system at the first time Text messages. A smoke sensor module can sensitively detect the smoke in the intelligent security system. [2]When the fire appears and the smoke concentration

**Corresponding Author:*
Ruizheng Wang,
Artificial Intelligence College, North China University of Science and Technology, No. 21 Bohai Avenue, Caofeidian District, Tangshan, Hebei, 063000, China;
Email: 2840432403@qq.com

DOI: https://doi.org/10.30564/jcsr.v2i4.2159

reaches the set value of the security system, the smoke alarm module will automatically alarm, the GSM module will start immediately, and the fire emergency message will be sent to the system user immediately.

3. A Study on the Main Modules of 2System

3.1 System Design Main Controller Analysis

The main controller in this set of intelligent home security alarm system is STC89C51 single chip microcomputer, the advantage of STC89C51 single chip microcomputer is that it is a low-power, high-performance CMOS 8 bit micro controller, itself has 8 K of on-line programmable Flash memory, using high-density non-volatile memory technology research and manufacture, and industrial 80 C51 of all product instructions and pins can be fully compatible. [3] An on-chip Flash receiver memory for online programming also applies to other regular programmers. Very smart 8 for single chip bit CPU and the programmable Flash, of online systems enable STC89C51 to be provided with more flexible and efficient coping schemes by many other embedded control applications.

The main function of the P0.3 pin in the STC89C51 single chip microcomputer in this set of intelligent home security alarm system is to control the alarm of the acousto-optic system. The main function of the P2.4 pin is to control the pyroelectric infrared sensor in the system. The main function of the P2.5 pin is to control the smoke sensor. The function of the P3.0 and the P3.1 pin is to connect the GSM module to realize the mutual transmission of information.

3.2 Analysis of Pyroelectric Infrared Sensor in System Design

The pyroelectric infrared sensor used in this intelligent home security alarm system is RE200B type of sensor. The RE200B sensor itself uses pyroelectric material polarization to detect infrared radiation with different temperature changes.

[4] The pyroelectric infrared sensor module generally includes three pins, the signal terminal, the power supply positive pole and the power supply negative pole, respectively, the pin at the P2.4 of the STC89C51 single chip microcomputer, the power supply and the contact ground. pyroelectric infrared sensor can sensitively detect whether there are strangers invading the security range of the system, by setting infrared probe in the monitoring position, the infrared radiation emitted by the human body can be converted into electrical signals, and its electrical signals are sent to the OUT port by the cooperation of amplifier circuit and comparison circuit. As soon as the OUT port outputs a high-level signal, the signal passes through the transistor 9013 and is converted to a low-level transmitter to STC89C51 MCU. STC89C51 pin at the P2.4C of the single chip microcomputer is high level state in normal operation. if the low level is detected, the high voltage is output through the P0.4 port, the alarm is controlled to emit the alarm sound of the buzzer, and the alarm indicator is lit at the same time. the GSM module will send an alarm message that detects a stranger invading the security system.

3.3 Analysis of Smoke Sensor in System Design

The smoke sensor adopts MQ-2 type of smoke sensor. The advantages of the MQ-2 type sensor are its high sensitivity, fast reaction speed, good stability, long service life and simple and convenient driving circuit. It is suitable for the detection of natural gas, various kinds of smoke, gas and gasoline and so on.

[8] The main component of the MQ-2 type smoke sensor is the LM393 voltage comparator, which adopts the principle of comparing the voltage of 2 feet with 3 feet. If the voltage of 2 feet is higher than that of 3 feet ,1 foot will automatically output low level; if the voltage of 2 feet is lower than that of 3 feet ,1 foot will automatically output high level. The voltage of three feet is obtained by dividing the voltage through a 9103 potentiometer of 10 K, and the voltage of two feet will change with the change of smoke concentration in the range of MQ-2 detection. MQ-2 the smoke sensor is equal to a large resistance throughout the circuit, if the security range The smoke concentration inside increases, the blocking force of the MQ-2 will become smaller, the voltage output to 2 feet will naturally become larger, and the voltage of 2 feet will be larger than that of 3 feet. At this time, the LM393 voltage comparator will naturally send a low level to the STC89C51 MCU through 1 foot. [6,7]The normal state of the P2.5 pin of the single chip microcomputer is a high voltage flat state. Once the low level is detected, the high voltage will be output through the output port of the P0.4, indicating the alarm sound of the buzzer issued by the control buzzer, and the alarm indicator will be lit, and the GSM module will send the detection of a stranger invading security system. Alarm messages within the system.

3.4 Analysis GMS Module in System Design

[5]The SIM900A module is applied in the design GMS the intelligent home security alarm system. The advantage of the SIM900A module is that its interface is very rich, the function is complete and the performance is powerful, and the applicability in many fields, such as smart home sys-

tem, intelligent security, intelligent monitoring, on-board monitoring and intelligent alarm, is very high, which can completely meet the application requirements of the above system. circuit schematic diagram of the SIM900A module, as shown in figure 1.

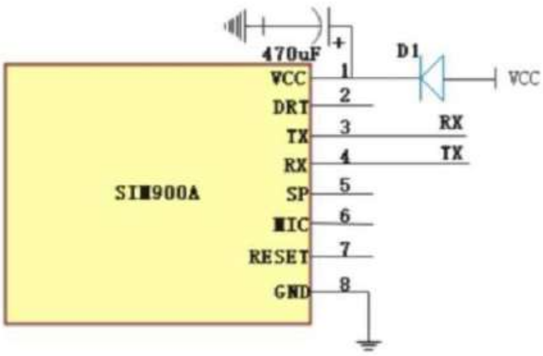

Figure 1. SIM900A Circuit diagram of the module

The pin 1 of the SIM900A module is connected to the LED lamp D1, and the main function is to indicate the operation of the GSM module. Pin 3 and pin 4 act on the transmission task of MCU pin P3.0 and P3.1, mainly to deal with alarm data, respectively.

4. Concluding Remarks

By deeply studying and analyzing the design of the intelligent home security alarm system under the action of STC89C51 single chip microcomputer, the function of the single chip microcomputer and the embedded application technology of the home security alarm system are successfully combined with the Internet of things technology. Exemption from family living around causing property, casualties, to provide people with a more intelligent, safe family living environment.

This design can be compatible with a large number STC89C51 control units under the design situation of single chip microcomputer, at the same time, it can combine these control units organically to form a whole set of intelligent control systems, make full use of STC89C51 single chip microcomputer, GSM the powerful function of the module to carry out systematic chip design for single chip microcomputer. STC89C51 single chip microcomputer has outstanding performance in its performance, excellent compatibility, low cost and strong practicability, and contributes to the stability of the whole security system. For the intelligent home security alarm system hardware, software design work carried out a lot of preparation work Through many experiments, we can overcome all kinds of difficulties and contradictions, and realize the design work of the intelligent home security alarm system.

References

[1] Jianfang Wen. Study on Home Intelligent Control System Based on AT89C51 Single Chip Microcomputer[J]. Electronic Production, 2019(09): 31-33.

[2] Qiaomei Ma. Design and Implementation of Fireproof and Anti-theft System for Dormitory Based on 51 Single Chip Microcomputer[J]. IT, 2018(03): 39-44.

[3] Chaoyang Zhang. The Design of Gas Alarm Based on AT89C51 Single Chip Microcomputer[J]. Journal of Tonghua Normal University, 2017, 38(12): 7-10.

[4] Tianxiang Guo. New Concept 51 MCU C Language Course[M]. Beijing: Electronic Industry Press.

[5] Xiangbin Ye. Sensor and test technology [M]. Beijing: Defense Industry Press, 2007.

[6] Peng Chen. LonWorks villa smart home system introduction[J]. Low Voltage Appliances, 2008(14): 59-62.

[7] Wei Xu, Dean Zhao. Detection and control of dissolved oxygen in aquaculture[J]. Agricultural Mechanization Research, 2007(1): 74-76.

[8] Guangdi Li, Yuexiu Zhu, Xiushan Wang. Foundation of Single Chip Microcomputer (Revised Edition)[M]. Beijing: Beijing University of Aeronautics and Astronautics Press, 2005.

DOI: https://doi.org/10.30564/jcsr.v2i4.2159

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
http://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Quality Models for Open, Flexible, and Online Learning

Ebba Ossiannilsson*

Swedish Association for Distance Education, International Council for Open and Distance Education, Lund, 22235, Sweden

ARTICLE INFO

Article history
Received: 7 September 2020
Accepted: 28 September 2020
Published Online: 30 November 2020

Keywords:
Covid-19
Crises
Flexible learning
Open online learning
OER
MOOC
Quality models
Stakeholders
Success factors

ABSTRACT

This article is based on research conducted for the European Commission Education & Training 2020 working group on digital and online learning (ET2020 WG-DOL) specifically regarding policy challenges, such as the following: 1) *Targeted policy guidance on innovative and open learning environments* under outcome; 2) *Proposal for a quality assurance model for open and innovative learning environments, its impact on specific assessment frameworks and its implication for EU recognition and transparency instruments*. The article aims to define quality in open, flexible, and online learning, particularly in open education, open educational resources (OER), and massive open online courses (MOOC). Hence, quality domains, characteristics, and criteria are outlined and discussed, as well as how they contribute to quality and personal learning so that learners can orchestrate and take responsibility for their own learning pathways. An additional goal is to identify the major stakeholders directly involved in open online education and to describe their visions, communalities, and conflicts regarding quality in open, flexible, and online learning. The article also focuses on quality in periods of crisis, such as during the pandemic in 2020. Finally, the article discusses the rationale and need for a model of quality in open, flexible, and online learning based on three major criteria for quality: excellence, impact, and implementation from the learner's perspective.

1. Introduction

This conceptual article is based on work done for the European Commission Education & Training 2020 working group on digital and online learning (ET2020 WG-DOL), specifically regarding policy challenges such as the following: (1) *Targeted policy guidance on innovative and open learning environments*, and (2) *Proposal for a quality assurance model for open and innovative learning environments, its impact on specific assessment frameworks and its implication for EU recognition and transparency instruments* [1]. The article aims to define quality in open, flexible, and online learning (OOL), particularly open education, open educational resources (OER), and massive open online courses (MOOC). Hence, quality domains, characteristics and criteria are defined and discussed, as well as how they contribute to quality and to personal learning so that learners can orchestrate and take responsibility for their own learning pathways. An additional goal is to identify the major stakeholders directly involved in open online education and to describe their visions, communalities,

**Corresponding Author:*
Ebba Ossiannilsson,
Swedish Association for Distance Education, International Council for Open and Distance Education, Lund, 22235, Sweden;
Email: ebba.ossiannilsson@gmail.com

 DOI: https://doi.org/10.30564/jcsr.v2i4.2357

and conflicts regarding quality in open, flexible, and online learning. Finally, this article discusses the rationale and need for a model of quality in open, flexible, and online learning based on three major criteria for quality: excellence, impact, and implementation from the learner's perspective.

Following this short introduction and description of the methodology, the subsequent section focuses on identifying the major stakeholders and their visions of quality. In the next section, open education is defined, and quality in open, flexible and online learning is described. Next, the concept of quality in open and flexible online learning, open educational resources (OER) and massive open online courses (MOOC) are elaborated. The final section discusses common and conflicting areas in quality dimensions in open and flexible online learning in OER and MOOC from the learners' perspectives. A model of quality dimensions in OOL, OER, and MOOC is presented. Also presented is a quality model of the three areas from the organizational perspective through the learner's lens, including the impacts on the organizational level, such as quality and excellence. The organizational and social perspectives on quality are not discussed in detail because they are beyond the scope of this article. The article concludes with recommendations for further research.

2. Methodology

This conceptual article is based on the findings of desktop research on the current discourse of quality models for open, flexible, distance, and online learning, as well as the future of education. In addition, quality-related approaches in times of crisis, such as the pandemic in the spring of 2020, are proposed. The content of this conceptual article is based on research by the author during the last two years, including work for the European Commission ET2020 WG-DOL. The article's focus is based on the author's research and perspectives during that period. The author has provided examples of the ongoing debate, which, however, do not always represent official perspectives. Moreover, the article does not provide a comprehensive review of developments in the field or events across the globe.

3. Results

In this section, the results of the author's research are presented. First, the stakeholders' interests in quality-related approaches and dimensions in open, flexible, and online education are explored, followed by a definition of open education. Descriptions of quality models in OOL, OER, and MOOC are then provided.

3.1 Stakeholders' Interests in Quality in Open Education, OER, and MOOC

Quality is related to the domains of economics, culture, and politics. Quality is also related to compliance, consumer protection, reputation, quality enhancement, and process improvements. Practical experience and academic research have shown that quality in open education is complex and is viewed from multifaceted perspectives that reflect the visions of those who consider the issue of quality [2]. It is obvious that "one size does not fit all." An aphorism that best describes this aspect is "*Quality is in the eyes of the beholder*."

Ossiannilsson, Williams, Camilleri, and Brown [2] identified three major groups of stakeholders who play roles in the definition and review of quality in open education, OER, and MOOC: learners, educational organizations, and society. The latter is concerned at regional, national, and international levels. Each stakeholder reviews quality from a unique perspective, which sometimes leads to dissonance among the various concepts of quality in open online learning, OER, and MOOC. This dissonance might lead to conflicts of interest with subsequent repercussions in the achievement of OFOL as well as in establishing a quality model that satisfies all stakeholders.

Stakeholders' interests in the quality of open education, OER and MOOC are related to their roles as leading partners, contributors, or participants. Hence, it is important to consider all three groups of stakeholder roles and to search for coinciding or conflicting views of quality among them. Traditionally, educational organizations may have a larger interest in purposes, such as branding, business models, security, marketing, competition, and goodwill. Society may be more interested in political aims such as democracy and policy issues such as equity, inclusion, relevance, sustainability, capacity building, and gender dimensions, as well as economic dimensions and consequences. Finally, learners may be more interested in their own motivations, *just-for-me learning*, their needs, autonomy, self-directed learning, lifelong learning, recognition, satisfaction, possibilities for employment, self-esteem, and self-realization [1]. Hence, it is expected that stakeholders define quality according to their needs. It is important to underline that sector-specific needs are focus in the research study by Ossiannilsson et al. Brown [2]. Figure 1 illustrates the main interests of the stakeholders in open online education.

DOI: https://doi.org/10.30564/jcsr.v2i4.2357

Figure 1. Stakeholders and their main interests in open online learning, OOL, OER, and MOOC

In addition to identifying the three major groups of stakeholders who play roles in the definition and review of quality in open education, OER, and MOOC, it is crucial to define the macro, meso, micro, and nano levels and how they are connected to and aligned with each other [2].

In the next section, a general definition of open education will be provided as a preface to the subsequent definition of quality. A discussion on opening up and open education will include OER and MOOC.

3.2 Definitions of Open Education

Education is viewed as a catalyst for the development of, societies and hence a key contributor to increasing universal access, democracy, and equality, which is stated in the United Nations (UN) Educational Scientific and Cultural Organization's (UNESCO) Sustainability Goal number four (SDG4). Education is also an essential condition in accelerating progress toward the achievement of global sustainability [3].

A general definition of open education encompasses resources, tools, and practices that employ a framework of open sharing to improve educational access and effectiveness worldwide. Open education combines knowledge sharing with 21st-century technology and competences to create a vast pool of openly shared educational resources, while harnessing today's collaborative spirit to enhance and facilitate educational approaches that are more responsive to learners' needs [1]. UNESCO has stated that open education contributes to the building of open and inclusive knowledge societies and to the achievement of the UN's Sustainable Development Goals [3-6]. The Commonwealth of Learning's approach to open education is to empower people through learning that leads to economic growth, social inclusion, and environmental conservation [7].

It is commonly accepted that open education includes resources, tools, and practices that operate within a framework of open sharing and aim to improve educational access and effectiveness worldwide. Open education combines knowledge sharing with 21st-century information and communication technology (ICT) to create a vast pool of openly shared educational resources, while harnessing today's collaborative spirit to enhance and facilitate educational approaches that are more responsive to learners' needs. The qualifier "open" in the term open education refers to the elimination of barriers (e.g., physical, mental, and organizational) that can preclude access, opportunities, and recognition of participation in institution-based learning. Open education has been defined by the Open Education Consortium [8] as:

"…*resources, tools and practices that employ a framework of open sharing to improve educational access and effectiveness worldwide*."

Through open education, learning and educational opportunities can be scalable by taking advantage of the power of the Internet, allowing rapid and essentially free dissemination, and enabling people around the world to access knowledge, connect, and collaborate. The traditional provision of education is limited by the capacity of educational institutions; consequently, this resource has been available to the few, not the many. Increased digitalization and digital transformation offer potential solutions to these limitations by giving a global audience access to free, open, and high-quality educational resources. By providing free and open access to education and knowledge, people can fulfill their desire to learn. Through open education, learners gain access to information, knowledge, and materials to help them succeed in their learning and study processes. Faculties and researchers can exchange materials, share data, develop networks, and draw on international resources. Employees can learn at their workplaces, supporting them in their work. People can connect with others with whom they otherwise would not meet to share information and ideas.

The wide interpretation of open includes not only access but also the ability to modify and use materials, information, and networks, which enables individuals to personalize their education and even use or reuse it in new ways. New audiences can be reached, and resources can be used in new contexts. In opening up education, learning can be available, accessible, modifiable, and free for anyone, everywhere, and at any time. Open education is now seen as a catalyst for the development and a key contributor to democracy, equality, and access for all. Education is also an essential condition in accelerating progress toward the achievement of global sustainability. In "opening up" education, openness embraces development and adoption. OER represent one of the main pillars of open education.

In their support framework for higher education institutions (HEI), Inamorato dos Santos, Punie, and Castaño

 DOI: https://doi.org/10.30564/jcsr.v2i4.2357

Muñoz [9] proposed a wide definition of the term "open education." Their framework includes a wide definition of the term, which accommodates different uses to promote transparency and a holistic approach to practice. The framework goes beyond OER, MOOC, and open access to embrace 10 dimensions of open education. The framework can be used as a tool by HEI staff to help them make strategic decisions regarding pedagogical approaches, collaboration between individuals and institutions, recognition of non-formal learning, and different ways of making content available. Because contemporary open education is enabled mainly by ICTs, there is great potential for innovation and access, which will contribute to the modernization of higher education in Europe and globally. The framework encompasses 10 dimensions, four transversals (i.e., leadership, strategy, quality and technology), and six core dimensions (i.e., content, pedagogy, recognition, collaboration, research, and access) (see Figure 2).

Figure 2. Support framework for higher education institutions (HEI) to open up education [9]

In opening up education, aspects of openness embrace the development and adoption of both OER and MOOC. OER and MOOC are defined and described in the following paragraphs.

An important step toward quality education and universal access to information was taken when the UNESCO General Conference adopted a recommendation on OER on 25 November 2019 [4]. This new recommendation supports the development and sharing of openly licensed learning and teaching materials, which will benefit students, teachers, and researchers worldwide. Moez Chakchouk, UNESCO Assistant Director-General for Communication and Information, stressed that the OER recommendation would contribute to building open and inclusive knowledge societies and achieving the objectives of the UN Sustainable Development Goals. The recommendation encompasses five areas [4,6]:

(1) Building the capacity of stakeholders to create, access, reuse, adapt, and redistribute OER

(2) Developing supportive policy

(3) Encouraging inclusive and equitable quality OER

(4) Nurturing the creation of sustainability models for OER

(5) International collaboration and networking

Furthermore, processes and indicators of monitoring and evaluation were adopted, and a new definition of OER was adopted [6]:

OER are learning, teaching and research materials in any format and medium, which are publicly available or under copyright, published under an open license allowing free access, reuse, reuse, adaptation and redistribution by others.

The UNESCO Recommendation supports the creation, use, and adaptation of inclusive and high-quality OER products, and it facilitates international cooperation in this field. Its objectives include the development of supportive policies and the creation of sustainability models of OER [6].

During debates at the General Conference, several representatives of Member States in different regions of the world expressed their support for the UNESCO Recommendation on OER and underlined their important role in providing access to quality education on digital platforms. The key contribution of the OER Recommendation to building open, inclusive and participatory knowledge societies was echoed throughout the discussion.

Following adoption of the Recommendation on Open Educational Resources (OER), UNESCO launched the OER Dynamic Coalition on 2 March 2020 [10]. The multi-stakeholders in the Coalition aim to expand and consolidate commitments to action and strategies, as well as reinforce international cooperation among all stakeholders in the five areas of the Recommendation.

OER can be considered a form of micro-learning because they are often delivered as stand-alone modules in formal and informal learning. According to UNESCO's SDG4, OER are crucial in sustaining lifelong learning for all. OER can be used as micro-learning in employment, personal interest, self-learning, "just-for-me," "just-in-time," and "just for fun" learning. Individuals who develop personal, social, and learning skills are better prepared to face the challenges of today's society and to seize the opportunities that change brings [11]. Furthermore, UNESCO stated that they believe that universal access to high-quality education is the key to peace building, sustainable social and economic development and inter-

  DOI: https://doi.org/10.30564/jcsr.v2i4.2357

cultural dialogue. OER provide a strategic opportunity to improve the quality of education and facilitate policy dialogue, knowledge sharing, and capacity building [4,6,10].

A second pillar of open education is represented by MOOC, which are related to the scalability of open and online education. Research has demonstrated that every letter in the acronym MOOC is negotiable; therefore, the term can have many definitions. According to EADTU, MOOC are courses designed for large numbers of participants, which can be accessed by anyone and anywhere as long as they have an Internet connection [12]. In several EADTU projects [12], such as the HOME project (Higher education Online: MOOC the European way), the ECO project (ECO: E-learning, Communication and Open-data: Massive Mobile, Ubiquitous and Open Learning), and OpenupEd, a definition of MOOC was shared by many European partners, who agreed that MOOC are courses designed for large numbers of participants. They can be accessed by anyone anywhere as long as they have an Internet connection, they are open to everyone without entry qualifications, and they offer a complete course experience online for free [13]. MOOC are also often seen as a political instrument and hence as a concept that should be broadly defined. EADTU (re-) defined MOOC in their quality model OpenupEd, which is described in detail below, so that it clarified the differences in the implementations of open and online courses. The first O in the abbreviation MOOC refers to open. Furthermore, because MOOC were built on OER, there is a strong link between them.

Having defined open flexible online learning (OOL), OER and MOOC, the most frequently used and common quality models are described in the following sections. These three types of open education-OOL, OER, and MOOC-use quality models that are specific to each type. However, in each type of open education, some features of quality display communalities with one or both of the other types.

3.3 Description of Quality Models in OOL, OER, and MOOC

The concept of quality is one of the most frequently discussed in education. Quality in online and e-learning usually refers to "*one in which the learner has a reasonable opportunity for success in reaching their goals*" [2]. Several authors and reports in the field have argued that quality in open online education and learning is defined as "*methods that successfully help learners develop knowledge and skills that they will require in a digital age*" [2].

In this section, quality models of open online learning, including OER and MOCCs, will be described. In general, the quality dimensions of OOL apply to OER and MOOC. However, both of the latter have specific quality dimensions, such as those that are absolute critical from the learner's point of view. In the following paragraphs, quality in open flexible online learning is discussed, followed by the most well-known and the most frequently used quality models of OER and MOOC.

3.3.1 Overview of Quality in Open Flexible Online Learning

A global research study was conducted by Ossiannilsson et al. on quality models in online and open education around the globe: "State of the art and recommendations" on behalf of the International Council for Online and Distance Education (ICDE) [2]. In this study, 40 of the most frequently used and best-known quality models of e-learning, online open learning, OER, and MOOC around the globe were reviewed and analyzed. The models' aims and rationales were analyzed, as well as their modality and application, including mature models of quality. In the report on the quality spectrum, norm-based (accreditation) versus process-based (enhancement) models were discussed, which included the entire spectra of accreditation, certifications, benchmarking, guidelines, and the issue of quality labels. A matrix of the analyzed models related to the spectra and the models was provided. Furthermore, stakeholders who had interests in the quality of open flexible online learning, OER, and MOOC were identified. Recommendations for stakeholders were included. The research showed that although many models were available on all continents, some quality indicators and quality dimensions were in common. Hence, sets of quality characteristics were presented. In all the reviewed quality models, the theoretical and practical approaches emphasized the importance of a holistic approach and the importance of placing the learner at the center. The quality indicators were considered standards, and best/next practice, especially in OOL (OER do not usually have a curriculum), were identified as being valid for OER and MOOC:

(1) Institutional support (vision, planning, and infrastructure)

(2) Course development

(3) Course structure or curriculum

(4) Course delivery

(5) Course design

(6) Teaching and learning (instruction)

(7) Student and faculty support

The model shown in Figure 3 and the study of more than 40 quality frameworks stemmed from research by Ossiannilsson [14] and were further developed in research by Ossiannilsson [15]. Products, services, and management are the three main domains that characterize this holistic model, which contributes to obtaining and developing a quality assurance mechanism, enabling individuals' success in their learning processes. A natural consequence is that both content and technology are crucial. Moreover, processes play a significant role because they are interactive and interdependent. The learner is placed at the center of this quality model. The ICDE research study confirmed the following characteristics of this model: the learner (student) is at the center, and services, products, and management are the three main quality domains. In Figure 3, the right side of the model represents the dimensions necessary to guarantee quality assurance in open and online learning from the learner's perspective.

In addition to these indicators, dimensions based on the learner's views, demands, and perspectives were identified as follows: transparency (learners can easily see and follow activities, demands, processes in a user-friendly way, so they can orchestrate their learning pathways); motivation (intrinsic and extrinsic); participation (involvement); productivity (being collaborators and/or *prosumers*), flexibility (time, path, place, learning style, content, and device); accessibility (anytime, anywhere, with any device, and the Web Accessibility Initiative [WAI]), interactivity (learner-to-learner, learner-to-academic, learner-to-material); and personalization (related to individual's needs, desires, study paths, and learning style). Flexibility was defined and considered crucial in the quality of OOL. These dimensions were emphasized as being crucial for student success in online learning, in which learners take control in orchestrating their learning processes. Moreover, these features of quality are iterative and interrelated.

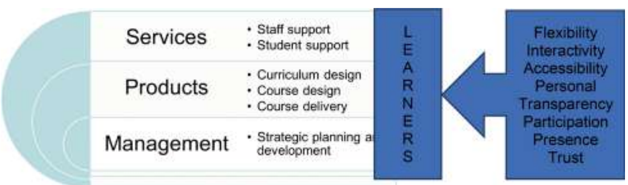

Figure 3. A model of quality in open and online learning

Ossiannilsson et al. [2] indicated that concepts of quality in OOL and education can be applied at different levels, such as the macro level (national/international), meso level (institutional), and micro level (individual practice). The review of the models in [12] addressed quality at the macro and meso levels. Less evidence of performance standards was found at the micro level, but it undoubtedly exists. For example, it is well-known that criteria for professional development and performance management are used by institutions engaged in the quality assurance of their open and flexible online learning programs. However, such quality models were developed in-house.

Because changes in the areas of open and flexible online learning, OER, and MOOC are occurring rapidly, it is necessary to have a flexible and agile approach to quality in open education. Although there are general dimensions of quality, it needs to be situational, cultural, and contextualized. In general, quality concerns the three Ps (i.e., people, products, and processes). The authors proposed that quality models require a holistic contextualized approach and a set of identified quality characteristics of quality assurance and quality enhancement, which are listed as follows:

(1) Multifaceted - e.g. Systems use a multiplicity of measures for quality and often consider strategy, policy, infrastructure, processes, outputs, and more to form a well-rounded view of holistic quality.

(2) Dynamic - e.g. Flexibility is built into systems to accommodate rapid changes in technology as well as social norms. For this reason, they rarely refer to specific technological measures and instead concentrate on services provided to users through that technology.

(3) Mainstreamed - e.g. While all the quality tools surveyed are aimed at high-level quality improvement, their benefits are intended to trickle down throughout the institution and be used in reflective practice by individual members of staff in their daily work.

(4) Representative - e.g. Quality systems seek to balance the perspectives and demands of various interested stakeholders, including students, staff, enterprises, governments, and society.

(5) Multifunctional - e.g. Most systems serve the triple function of instilling a quality culture within an institution, providing a roadmap for future improvement, and serving as a label of quality for outside perspectives.

Kahn and Ally [16] argued that quality in e-learning and online learning concerns the three Ps. Furthermore, similar to the findings in Ossiannilsson [15,21,22,24] and Ossiannilsson et al. [17], Frydenberg [17], Kahn and Ally [16], and Kahn [18] emphasized an eight-dimension framework of quality. The framework was designed to guide the process of developing content for e-learning purposes in both public and private institutions as they moved from traditional approaches to an electronic format. The eight dimensions of this e-learning framework are as follows: (1) institutional, (2) educational, (3) technological, (4) interface design, (5) evaluation, (6) management, (7) resource support, and (8) ethical (see Figure 4).

  DOI: https://doi.org/10.30564/jcsr.v2i4.2357

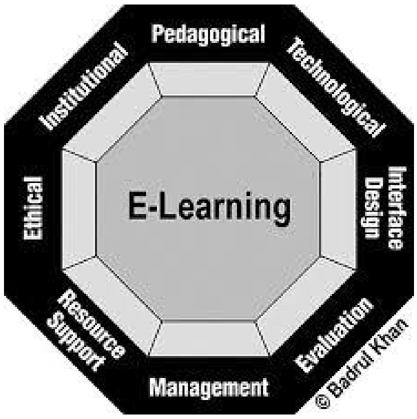

Figure 4. Kahn's quality framework [18]

Each of these eight dimensions includes a group of concerns or questions that need to be examined to assess and develop the e-skills of high-quality institutions. Kahn and Ally [16] and Kahn's [18] quality framework is shown in Figure 3. Similar to Frydenberg [17], Kear et al. [19], Inamorato dos Santos et al. [9], and Ossiannilsson [1,2,14,15], they argued that it is crucial to use a holistic ecosystem approach because the dimensions are interrelated.

In Contact North [20], the researchers argued that the quality agenda has to be rewritten and further developed by Ossiannilsson [21-24] to examine the effects of the dimensions over time on learning, careers, lifelong learning, community involvement, and benefits. Engagement is a key driver of quality, innovation, flexibility, the effective use of technology in learning, teaching, analytics, and assessment involving learners in practical applications of the content. In re-thinking the approach to quality, we should ask ourselves the following:

(1) How? How do students experience their learning?

(2) How? How do faculty experience their teaching?

(3) What? Focus on outcomes in depth.

(4) So what?

(5) Then what?

Accordingly, there are requirements to move to a much more experiential and outcome-based view of quality if it is to be the engine of transformation. Furthermore, agile approaches are needed to respond to rapid changes in the educational learning landscape. According to UNESCO Futures of Education, quality must be considered from the learner's perspective in line with the approach of learning to become [5,23].

3.3.2 Quality Models of Open Educational Resources (OER)

What does quality in OER mean? It is difficult to specify exactly what quality means in the context of OER, where traceability, accessibility, and availability are at least as important as the production values they represent. There is a difference in the emphasis on OER sharing, as third parties are actively encouraged to reuse, recycle, and remix resources. Overtime, as OER advocates, and I argue, this will lead to higher standards. However, the problem remains that the quality of learning resources is usually determined by the following:

(1) Accuracy

(2) Reputation of author/institution

(3) Standard of technical production

(4) Accessibility

(5) Fitness for purpose

The question of trust is an important factor in the issue of quality. OER that reference Wikipedia are obvious examples. While it is possible to abuse trust in relation to OER licenses, the community aspect and the inherent iterative model provide protection in the long run. Cultural issues are identified as important in relation to whether and how people share learning and teaching resources. Different institutions, sectors, and professional communities may all have their own established practices in terms of sharing teaching methods and learning materials. Academics may feel more connected to the culture of their discipline or professional practice community than to the institutional culture. It could be argued that within the higher education sector, there is no such thing as an institutional culture because there are many subcultures, which are often linked to different institutional roles, traditions, and approaches that can be more convincing than policy and policy documents. Moreover, some traditions and practices can lead to the slow uptake of new approaches and ideas [26].

The open movement challenges individuals and groups to change their existing practices. However, in large institutions with many subcultures, uneven development is very likely. An institution-wide approach to human resource development and promotion can help to overcome some of these cultural barriers and encourage the use of OER. Some institutions may choose to engage in such activities in order to move forward [21,27,30,31].

Stakeholders in different contexts have differing perspectives on what constitutes quality in OER. However, as emphasized above, quality is very much in the eye of the beholder. The most comprehensive, best-known, and most frequently used quality model related to OER is the TIPS model, which was developed by the Commonwealth Educational Media Center for Asia (CEMCA) [27].The acronym TIPS stands for teaching and learning process (T), information, material and content (I), presentation, product and format (P), and system, technical and technology (S).

According to CEMCA and Kawachi [27], three fields of quality can be clarified. The first two-quality as a product (content) and quality as a process-are well known. As products, OER can be released with the logo or the brand-name of the institution that produced them with the intention to preserve and/or improve its reputation. As a process, OER metadata tags can be completed by the end-users of the OER to offer feedback and comments for the benefit of future users. However, the educational experience is much more than simply producing free online content (regardless of whether it is high-quality content). For that purpose, the TIPS guidelines are oriented to nurturing the idea of quality as a culture, which constitutes the third field. Culture is shaped and developed by people. This idea resonates with Ossiannilsson and Kahn, who argued for the three Ps [15-17,21-24]. Developing a culture of quality may be the way forward rather than advocating resources as quality products or simply promoting quality practices and quality processes. Another common way to characterize OER, which is included in the TIPS framework, is a classification that embraces dimensions that describe the function of OER:

(1) Cognitive Domain: content knowledge, content skills, and reflective critical thinking skills to be learned

(2) Affective Domain: the motivations, attitude and decision to initiate performance, learner independence and autonomy

(3) Metacognitive Domain: understanding how the task is performed, and the ability to self-monitor, evaluate and plan own future learning

(4) Environment Domain: the localization, artistic presentation, language, multimedia, interactivity, and embedded links to other content

(5) Management Domain: discoverability, tagging, including for time management, transmissibility, business models

In the TIPS model by CEMCA and Kawachi [27], the three fields of product, process, and culture, as well as the five domains are integrated using 38 criteria under the four headings comprising the acronym TIPS, as shown in Figure 5. Additionally, the TIPS guidelines are intended to facilitate a culture of professional reflection, which was described above as the third field.

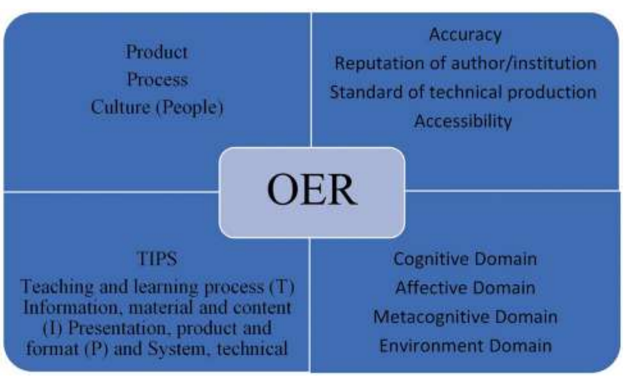

Figure 5. Free illustration of identified fields, domains, and features of the TIPS quality model related to OER developed by the Commonwealth Educational Media Center for Asia (CEMCA) [27]

Kahn's framework of e-learning has been applied to not only OER and MOOC but also micro-learning [18,28,29]. The eight dimensions of this framework are as follows: (1) institutional, (2) educational, (3) technological, (4) interface design, (5) evaluation, (6) management, (7) resource support, and (viii) ethics. It is obvious that the dimensions in this framework correspond highly to the features and domains in the TIPS framework [24].

3.3.3 Quality Model of MOOC

The EADTU's European quality benchmarking model OpenupEd offers a framework of common features in MOOC, which emphasizes learners' perspectives. The framework is not meant to be followed slavishly, as it is a benchmarking model, but it is intended to serve as a guideline for improving the quality of MOOC offerings. Because of its flexibility, some institutions and MOOC conform to the benchmarks more than others do. The framework is intended to be applied at the institutional level and in single MOOC. The background features of the OpenupEd framework of quality are intended to offer MOOC to everybody in a flexible manner that meets the needs of today's learners. The model comprises 35 benchmarks that relate to self-assessment or to external reviews required for the OpenupEd quality label. Any institution and course that meets the quality features of OpenupEd is awarded an international quality label.

The OpenupEd features guarantee that MOOC secure the fundamental values of open and free education in all societies. From the learner's perspective, OpenupEd facilitates appropriate incentives for learners to progress and succeed in learning by removing barriers such as costs and physical, mental, or requirements at entry into learning and along the learning path. As shown in Figure 6, OpenupEd is characterized by eight features:

(1) Openness to learners: free of charge, free admittance, open access, learn anywhere online, start anytime, self-paced, and diversity in languages and cultures, a spectrum of approaches and contexts, accounting for variety, and profiling.

(2) Digital openness: open sources (software), open access (scientific output), open content, and open educational resources.

(3) Learner-centered: all unnecessary barriers to learning are removed, while aiming to provide students with a reasonable chance of success in education. The focus is more on innovation in open pedagogical thinking and less on technology and platforms. Students construct their own learning in a rich environment, and they share and communicate it with others in learner-centered activities.

(4) Independent learning: providing high-quality ma-

  DOI: https://doi.org/10.30564/jcsr.v2i4.2357

terials and a rich learning environment to enable an independent learner to progress through self-study.

(5) Media supported learning: course materials should make use of online affordances (interactivity, communication, and collaboration) as well as rich media (video and audio) to engage students.

(6) Recognition: OpenupEd partners offer a full/complete course experience, including recognition options.

(7) Quality focus: quality assured, accredited, and European Qualifications Framework level.

(8) Spectrum of diversity: a course should be inclusive and accessible to a wide diversity of citizens. In short, it should appeal to everyone, be diverse in language and culture, provide a spectrum of approaches and contexts, and account for variety and profiling. Diversity is considered one of the key advantages of new and emerging learning technologies.

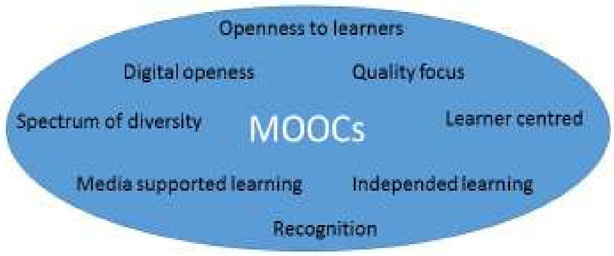

Figure 6. Identified features of the OpenupEd quality model related to MOOC by EADTU [12].

3.3.4 Quality dimensions of OOL, OER, and MOOC from the learner's perspective

Figure 7 presents a summary of the areas of open, online learning, OER, and MOOC and the main quality domains, dimensions, and features. In all three perspectives, the learner is at the center.

In Figure 8, OOL, OER, and MOOC are defined, and the most frequently used quality models in those areas are included. However, the quality dimensions of the models are described in general terms, and they do not necessarily focus on the learner as the core or consider the learner's perspective, even if they claim that personalization is a core quality dimension. Although personalization is important, the concept can be interpreted in at least two ways: literally and aligned with organizations' offers and services.

The learner's perspective on quality in OOL, OER, and MOOC is taken literally. Furthermore, several quality dimensions have been identified through experience and research, which are related to motivation, success, and self-determined learning [2]. In this context, success means motivation, passion, learning to learn, outcomes in the form of completion, employability, and "just-for-me" learning. Previous studies showed that if learners can take control and orchestrate their own learning and if the offers and services embed quality dimensions such as flexibility, transparency, personalization, there is a positive effect on completion rates, which unfortunately used to be a highly ranked quality indicator in traditional education and a marker of success in institution [15,21].

As shown in Figure 8, additional quality dimensions for learners to succeed in the learning process can be identified in OOL, MOOC, and OER. Some dimensions are commonly considered quality dimensions that are crucial not only for success from the learner's perspective but also for completion rates. Unfortunately, many quality dimensions are not mentioned, considered, or explicit in the models presented above.

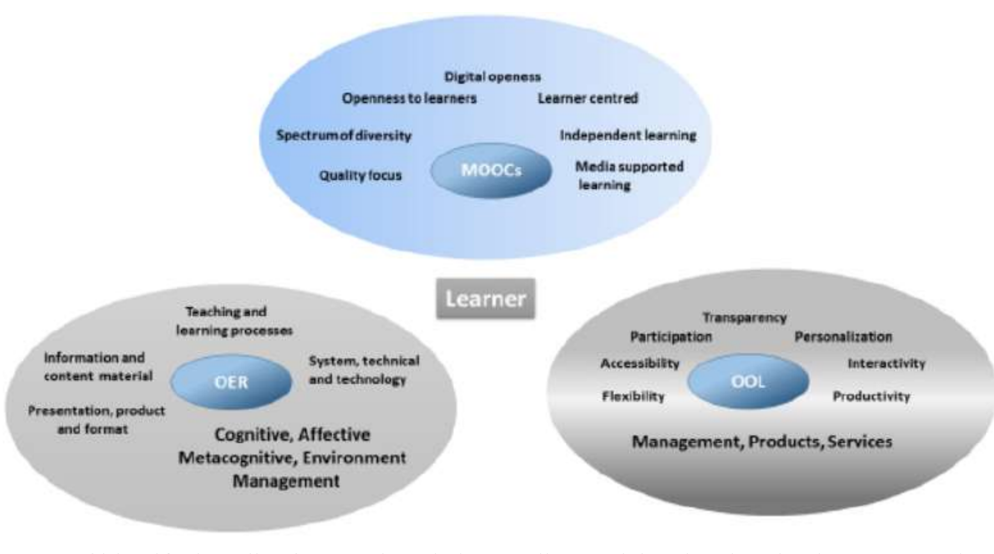

Figure 7. Summary of identified quality features in existing quality models related to the three perspectives OOL, OER, and MOOC

Figure 8. Quality dimensions from the learner's perspective on the three domains of OOL, OER, and MOOC. The common areas are identified and clustered

The common quality dimensions in OOL, OER, and MOOC from the learner's perspective are identified in Figure 6. Table 1 provides a detailed explanation of the concepts and what they relate to.

Table 1. Common Concepts in OOL, OER, and MOOC

Accessibility	Related to anytime, anywhere, with any device to Web Accessibility Initiative (WAI)
Content	Related to needs and desires; have to learn, want to learn, must learn
Fitness for purpose	Meets learners' desires and needs, just-in-time and purpose
Just-for-me	Meet learners' desires and needs, just-in-time and purpose at the right level for individual's knowledge, skills and attitudes, here and now
Just-in- time	Timing exactly when individuals have needs, motivation, passion and requirements related to work and leisure time
Motivation	Related to both intrinsic and extrinsic motivation
Passion	A powerful feeling. An extreme interest in or wish for doing something, such as a hobby, activity, and/or to learn something
Personalization	To personalize related to individual's needs, desires, study paths and style. Level of flexibility in all means
Security	Personal security, as well as technical, moral, and legal security
Quality	Quality of content, process, culture, and at all means and levels; quality assured, accreditation, certificated, peer reviewee
Recognition	Valued for time and effort spending time and energy and sometimes cost in learning settings and contexts
Trustworthiness	Trust and reliability, security, current, professional
User-friendly interface	Easy to navigate, easy overview, interactive, and intuitive

4. A Quality Model Embracing the Threefold Perspectives on OOL, OER, and MOOC

A fragmented picture of quality assurance in digital and online learning arose from the analysis presented in this conceptual research article and from the peer learning activities performed within the mandate of the WG-DOL. This article attempted to present a quality model that embraces all three main areas described above. The learner at the center, that is, at the heart of the model, symbolizes single individuals and their motivation, purpose, enjoyment, and passion for learning in open OOLs with MOOC, OER, and other open sources, which allows them to have options to become autonomous, self-directed, self-paced lifelong learners.

As shown in Figure 9, through identification and clustering, three domains of quality emerged: (1) the excellence of the open education offered; (2) the effects of the learning process; (3) the implementation from the learners' point of view in relation to the material and methods used to learn. Each domain has a set of characteristics based on and clustered from the quality models described above. Depending on the context, they can be situated and elaborated in detail. However, it should be noted that although most of the obvious quality features are included, other features, dimensions, and benchmarks might be emphasized, depending on the specific educational purpose of open education. However, any model must be flexible, which was argued in a quality research report by Ossiannilsson et al. [2], because open learning environments are changing rapidly.

Figure 9 shows a framework of quality domains in OOL, OER, and MOOC from the learner's perspective. This framework is the result of research on the most commonly used quality models in OOL, OER, and MOOC, as well as clustering quality domains, indicators, and dimensions. The quality of the product and the process from the learner's perspective has implications and consequences for *excellence, impact (Outcome)*, and *implementation*. The model can be elaborated and described as follows:

  DOI: https://doi.org/10.30564/jcsr.v2i4.2357

(1) Excellence is related to the comparison of the quality of a concept or an object with its peers and its maximum potential. Excellence relates not only to quality at all levels but also to efficacy, accuracy, and research. OOL, OER, and MOOC are excellent when quality not only exists at all levels but also is efficient. Moreover, offerings must allow and meet learners' expectations of equity, access, participation, collaboration engagement, passion, and motivation. User-friendliness, accuracy, affordance, fitness for purpose, and just-for-me learning are also related to excellence.

(2) Impact is related to both availability and accuracy, as well as to the measure of the extent to which an object or concept is effective. It is also related to the consequences or implications of the object or concept, the context in which it is applied, and the use to which it is put by the user. Impact also related to learning outcomes, equity, access, participation, collaboration, availability, and accuracy. There is an impact if learners experience access, availability, usability, and accuracy in their learning process. The levels of just-in-time and just-for-me learning are valid as high-impact factors. Similarly, recognition, self-esteem, and opening up new possibilities for life and employability are highly valid factors.

(3) Implementation is related to efficacy, fitness for purpose, services, and recognition. Implementation is also related to just-for-me learning.

Figure 9. Framework of quality domains in OOL, OER, and MOOC from the learner's perspective

Embracing and embedding learners' perspectives on quality in OOL, OER, and MOOC are important for organizations and institutions if they take personalization seriously by creating learner-centered educational settings, learning arenas, research, offers, and services. Embedding learners' perspectives on quality means that they can increase their own levels of quality and expand in relation to excellence, impact, and implementation. For organizations, excellence and outstanding performance can be related to enhancing quality at all levels. For stakeholders, quality can be related to efficacy, accuracy, and research. For organizations, impact is related to both availability and accuracy. Impact is also related to visible and sustainable outcomes, innovation, and change for others and society, which is often related to high ranking. Considering the learner's perspective on quality implementation from the beginning could serve to embed quality dimensions that positively affect learners' access, engagement, efficacy, fitness for purpose, equity, and inclusiveness. By implementing this approach, organizations can increase excellence for all learners, markets, and societies worldwide.

Figure 10 shows the framework of quality domains in OOL, OER, and MOOC from the organization's perspective. The framework is based on the quality models in OOL, OER, and MOOC and through the clustering of quality domains, indicators, and dimensions. The model can be elaborated for organizations to suit their aim and business model:

(1) Excellence is related to comparison of the quality of a concept or an object or to its peers and to its maximum quality potential. Excellence is related to quality at all levels and all stakeholders, as well as to efficacy, accuracy, and research.

(2) Impact is related to both availability and accuracy, as well as to the extent to which an object or concept proves effective. Impact is related to consequences or implications for the nature of the object or concept itself as well as the context in which it is applied and the use to which it is put by the user. Impact is also related to learning outcomes, equity, access, participation, collaboration, availability, and accuracy.

(3) Implementation is related to efficacy and the fitness for purpose of the object and concept being assessed. Implementation is also related services, cost, strategy, recognition, pedagogy, technology, and leadership.

Figure 10. Framework of quality domains in OOL, OER, and MOOC from the organization's perspective

5. Discussion and conclusion

This conceptual research article addresses peer learning activities performed within the work of WG2 DOL, output 2 on the quality assurance model for open and innovative learning environments, its impact on specific assessment frameworks, and its implications for EU recognition and transparency instruments. Through the identification of the most well-known and used quality models in the areas of OOL, OER, and MOOC, their concordance, and the identification of stakeholders and their individual and common interests, a tentative quality model was identified based on the learner's perspective and hence the organization's perspective. Organizations must review quality through the lenses of the learners.

In summary, quality in OOL, OER, and MOOC can be described and discussed at differing levels and from several perspectives. Several models of quality in OOL, OER, and MOOC have been applied. However, as discussed in this conceptual article, common core quality dimensions are crucial from the learner's point of view, which must be considered in discussing and defining quality. In opening up education, there are greater possibilities for and demands on increased personalization and *just-for-me learning*. Additionally, different stakeholders have different rationales, purposes, and interests. Moreover, they do not always have the same values regarding quality. Three main domains were derived from this overview of quality models, stakeholders' perspectives, and the learner-centered approach, which are crucial to consider for both learners and organizations: *excellence, impact, and implementation*.

Further research could focus on changes in learning landscapes, such as those caused by the pandemic in the spring of 2020 [22,32,33,34,35]. Additional topics recommended for further research are the future of education [5,22] in preparation for the changing learning and educational agenda, and the upcoming paradigm of *the new normal*, in which learner is firmly at the center and agile leadership and management approaches are required.

References

[1] Ossiannilsson, E. Quality models for open online learning (OOL), open educational resources (OER) and massive open online courses (MOOC). Brussels: European Commission ET Working Group on Open and Distance Learning (ODL), 2016. (Unpublished raw data).

[2] Ossiannilsson, E., Williams, K., Camilleri, A., Brown, M. Quality models in online and open education around the globe: State of the art and recommendations. Oslo: International Council for Open and Distance Education. 2015.
http://www.icde.org/en/icde_news/Global+overview+of+quality+in+online+and+open+education.b7C_wRvG43.ips

[3] UNESCO. SDG4, 2015.

[4] UNESCO. OER recommendation, 2019.

[5] UNESCO. Futures of education. 2019a.
https://en.unesco.org/news/new-unesco-recommendation-will-promote-access-educational-resources-all

[6] UNESCO. Recommendation on open educational resources (OER). Ref: CL/4319, 2020.
http://portal.unesco.org/en/ev.php-URL_ID=49556&URL_DO=DO_TOPIC&URL_SECTION=201.html

[7] Commonwealth of Learning.

[8] Open Educational Consortium. Definition of open education, 2015.

[9] Inamorato dos Santos, A., Punie, Y., Castaño Muñoz, J. Opening up education: A support framework for higher education institutions. Seville: Joint Research Center, 2016.
https://publications.jrc.ec.europa.eu/repository/bitstream/JRC101436/jrc101436.pdf

[10] UNESCO. Dynamic coalition, 2020.

[11] Sala, A., Punie, Y., Garkov, V., Cabrera-Giraldez, M. LifeComp: The European framework for personal, social and learning to learn key competence. Luxembourg, 2020.
DOI:10.2760/302967 (online). DOI:10.2760/922681 (print).

[12] EADTU. Definition of massive open online courses (V1.1). 2015.
http://www.openuped.eu/images/docs/Definition_Massive_Open_Online_Courses.pdf

[13] Jansen, D., Rosewell, J., Kear, K. Quality frameworks for MOOC. In M. Jemni, K. Khribi, and M. Koutheair (Eds.), Open education: From OER to MOOC. Springer, 2017: 261-28.

[14] Ossiannilsson, E. Benchmarking e-learning in higher education: Lessons learned from international projects. (Doctoral dissertation, Oulu University, Finland), 2012.
http://jultika.oulu.fi/Record/isbn978-952-62-0041-5

[15] Ossiannilsson, E. Opening up education: Quality matters and personalization. In M. T. Restivo, A. Cardosa, & A. M. Lopez (Eds.), Online experimentation: Emerging technologies and IoT. International Frequency Sensor Association (IFSA), 2015: 247-260.

[16] Kahn. B. H., Ally, M. International handbook of e-learning, theoretical perspectives and research. Routledge, 2015, 1.

[17] Frydenberg, J. Quality Standards in eLearning: A matrix of analysis. The International Review of Re-

  DOI: https://doi.org/10.30564/jcsr.v2i4.2357

search in Open and Distributed Learning, 2002, 3(2). https://doi.org/10.19173/irrodl.v3i2.109
[18] Kahn. B. H. E-learning framework and models, 2020. http://asianvu.com/bk/framework/
[19] Kear, K., Rosewell, J., Williams, K., Ossiannilsson, E., Covadonga Rodrigo, C., (UNED), Sánchez-Elvira Paniagua, A, Santamaría Lancho, M., André Vyt, A., Harvey Mellar, H. Final Editing: Karen Kear. K., & Rosewell, J., coordinated by George Ubachs, G. (EADTU), and Lizzie Konings, L. (EADTU). Quality assessment for e-learning: A benchmarking approach (3rd ed.). Heerlen: EADTU, 2016.
[20] Contact North. Rethinking quality dimension, 2018.
[21] Ossiannilsson, E. OER and OEP for access, equity, equality, quality, inclusiveness, and empowering lifelong learning. The International Journal for Open Educational Resources, 2019, 3.
[22] Ossiannilsson, E. After Covid19, some challenges for universities. In A. Tlili, D. Burgos, and A. Tabacco (Eds.), Education in a crisis context: COVID-19 as an opportunity for global learning. Springer, LNET Series, 2020a.
[23] Ossiannilsson, E. OER: open to ideas, open to the world. Springer Briefs in Open and Distance Education. New York: Springer. (Manuscript in preparation), 2020.
[24] Ossiannilsson, E. Toward a quality framework for OER: Application of Kahn's quality framework. In B. H. Khan, S. Affouneh, S. Salha, and Z. N. Khlaif (Eds.), Challenges and opportunities for the global implementation of E-learning frameworks. Manuscript submitted for publication, 2020c.
[25] JISC. (n.d.) OER quality.
[26] Kawachi, P. Quality assurance guidelines for open educational resources: TIPS framework. CEMCA, 2024.
http://oasis.col.org/bitstream/handle/11599/562/TIPS-Framework_Version%202%5B1%5D%20Copy.pdf?sequence=1&isAllowed=y
[27] Ossiannilsson, E., Glapa-Grossklag, J., Zhang, X. Opening pathways for access, inclusion, flexibility, and quality. Paper presented at the ICDE WCOL 2019, Dublin, IE, 2019.
[28] Corbell, J. R., Kahn, B. H., Corbell, M. E. (Eds.). Microlearning in the digital age: The design and delivery of learning snippets. Routledge, 2020.
[29] Fifilan, K. Could micro-learning make employees more engaged? HR learning and development, 2019. https://www.raconteur.net/hr/micro-learning-engagement?utm_source=email&utm_medium=Newsletter&utm_campaign=Weekly&utm_term=July-23&utm_content=1
[30] Ossiannilsson, E., Altinay, F., Altinay, Z. Analysis of MOOC practice from learner experiences and quality cultures. Educational Media International, 2015, 4: 272-283.
[31] Ossiannilsson, Altinay, Z., Altinay, F. Towards fostering quality in open online education through OER and MOOC practices. In M. Jemni, K. Khribi, and M. Koutheair (Eds.), Open education: From OER to MOOC. Springer, 2017: 189-204.
[32] Ossiannilsson, E., Zhang, X., Wetzler, J., Gusmão, C. M. G., Aydin, C.H., Jhangiani R. S., Glapa-Grossklag, J., Makoe, M., Harichandan, D. From open educational resources to open educational practices: Moving from the margins to the mainstream [Special issue]. From open educational resources (OER) to open educational practices (OEP): Examining re-appropriation of resources in open and distance education. DMS-DMK Distance et mediations des saviors - Distance and mediations of knowledge, 2020. (Manuscript submitted for publication)
[33] Bozkurt, A., Jung, I., Xiao, J., Vladimirschi, V., Schuwer, R., Egorov, G., Lambert, S. R., Al-Freih, M., Pete, J., Olcott, Jr., D. Rodes, V., Aranciaga, I., Bali, M., Alvarez, Jr., A. V., Roberts, J., Pazurek, A., Raffaghelli, J. E., Panagiotou, N., de Coëtlogon, P., Shahadu, S., Brown, M., Asino, T. I. Tumwesige, J., Ramírez Reyes, T., Barrios Ipenza, E., Ossiannilsson, E., Belhamel, K., Irvine, V., Sharma, R. C., Adam, T., Janssen, B., Sklyarova, T., Olcott, N. Ambrosino, A., Lazou, C., Mocquet, B., Mano, M., Paskevicius, M. A global outlook to the interruption of education due to COVID-19 pandemic: Navigating in a time of uncertainty and crisis. Asian Journal of Distance Education, 2020, 15(1): 1-128.
[34] Sala, A., Punie, Y., Garkov, V., Cabrera-Giraldez, M. LifeComp: The European framework for personal, social and learning to learn key competence. Luxembourg, 2020.
DOI:10.2760/302967 (online). DOI:10.2760/922681
[35] Uhomoibhi, J., Rojão Lourenço Azevedo, A., Monteiro Lopes Azevedo, J. M., Ossiannilsson, E. Learning analytics in theory and practice: Guest editorial. The International Journal of Information and Learning Technology, 2019, 36(4).
DOI: https://doi.org/10.1108/IJILT-08-2

BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
http://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Human-Centered A.I. and Security Primitives

Alex Mathew*

Department of Cybersecurity, Bethany College, USA

ARTICLE INFO

Article history
Received: 9 November 2020
Accepted: 20 November 2020
Published Online: 30 November 2020

Keywords:
Artificial Intelligence
Deep learning
Digital signatures
Machine learning
Private information retrieval

ABSTRACT

The paper reviews how human-centered artificial intelligence and security primitive have influenced life in the modern world and how it's useful in the future. Human-centered A.I. has enhanced our capabilities by the way of intelligence, human informed technology. It has created a technology that has made machines and computer intelligently carry their function. The security primitive has enhanced the safety of the data and increased accessibility of data from anywhere regardless of the password is known. This has improved personalized customer activities and filled the gap between the human-machine. This has been successful due to the usage of heuristics which solve belowems by experimental, support vector machine which evaluates and group the data, natural language processing systems which change speech to language. The results of this will lead to image recognition, games, speech recognition, translation, and answering questions. In conclusion, human-centered A.I. and security primitives is an advanced mode of technology that uses statistical mathematical models that provides tools to perform certain work. The results keep on advancing and spreading with years and it will be common in our lives.

1. Introduction

Artificial intelligence (AI) is the study and design of algorithms that perform tasks or behaviors that a person could reasonably deem to require intelligence if a human were to do it [1]. Many mature technologies nowadays adopted by the general public have historically switched their design approaches from machine-centered to human-centered [2]. Increased digitization in various spheres of life has led to a shift towards the implementation of digital technologies for data management and administration [3]. The business sector offering A.I. services benefits the most from the importance of Human-Centered A.I. Programmers are directed to ensure a continuous flow of income from the user, meaning that humans have to be dependent upon the software in a certain sense.

Artificial Intelligence is of three different types:

(1) Narrow or Weak A.I.

(2) General or Strong A.I.

(3) Artificial Superintelligence

Current progressing improvements draw scientists closer to achieving general AI with several theories speculating its future where intelligent killer robots will take over the world by either wiping out the human race or enslaving all humanity. The optimistic theory relates a blending co-relation between humans and robots, where humans use artificial intelligence as a tool to enhance their life experience. These machines' speed and accuracy are already gaining root in human lives, performing tasks that would

*Corresponding Author:
Alex Mathew,
Department of Cybersecurity, Bethany College, USA;
Email: Dr.alex.soh@gmail.com*

DOI: https://doi.org/10.30564/jcsr.v2i4.2534

have otherwise been impossible for humans to complete independently.

However, humans' creativity and emotions are incredibly unique, that it becomes challenging for any programmer to replicate the machines' traits. Security Primitive is also an essential part of the A. I since it determines whether the selected portions of the image match with the password and also provides random access to the selected image if the password of the user is known. It refers to an integrated protocol within the software where specific procedural steps have to be taken before to accomplish an intended purpose. For example, a person may commit to a chosen value while keeping it hidden from others, with the ability to reveal it later. Even so, no one should trust the artificial system by default. Verification procedures that allow testing the AI "black box" behavior and outputs using different solutions already available should comply with legislation [4].

2. Proposed Methodology

This research paper's methodology section portrays the steps that can be followed to identify our current position with artificial intelligence. Understanding the primitives is a secure way of grasping the root of artificial intelligence.

3. Commonly Used Security Primitives

3.1 Digital Signatures

A digital signature is a mathematically designed scheme meant to verify the authenticity of digital documents. Once the digital prerequisites have been satisfied, the recipient will have no reason to doubt whether the known sender created the message and if its integrity was not altered during transit. It is implemented by having all scanners in, say, an office setting, sign each scan, and have the signature verified by the viewing applications to avoid fraud as any alteration is detected [5].

3.2 Private Information Retrieval

A file from a given server can be retrieved without leaving a trace of the database from where it was initially. It is particularly useful where the user cannot access the other information from the same database. The only possible protocol that can grant the user information-theoretic privacy is when the entire copy of the database is sent to them. Clients want to know that the servers perform correctly, to understand the reasoning behind their actions, and to know how to use them appropriately to guarantee safety of their information [6].

3.3 Mix Networks

Mix networks refer to complicated protocols created by hard-to-trace routing in a chain of proxy servers called mixes. The secret of a Mix Network is that it takes messages from multiple senders then sends them back at a random order to the next destination. The whole process becomes hard to trace since the link between the source and the goal is destroyed.

3.4 Algorithm

The three different steps in this research are:

(1) Supervised Learning

Programs set in the model with unrelated data are compared with the actual target outputs; therefore, minimizing errors. There are six kinds of supervised learning classifiers: random forest, gradient boosting machine, conditional random forest, naive Bayes, neural network, and support vector machine [7].

(2) Unsupervised Learning

Unsupervised learning is essentially incomprehensible without inductive biases both on the considered learning approaches and the data sets [8].

(3) Reinforcement Learning

Artificial neural networks are under the heading of regression and clustering algorithms, which either can fall under supervised or unsupervised methods. Reinforcement Learning is used in virtual games, in building collaborative multi-agent systems [9]. Virtual reality increases the performance of training due to immersion and realistic spatial objects [10].

Flow charts:

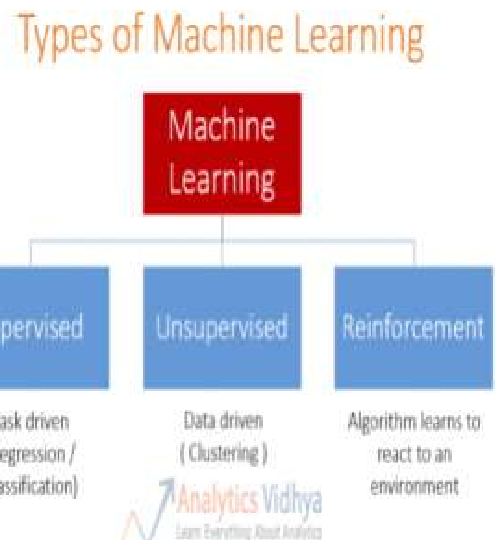

Figure 1. Types of Machine Learning algorithms

Source: https://www.analyticsvidhya.com/blog/2015/06/machine-learning-basics/

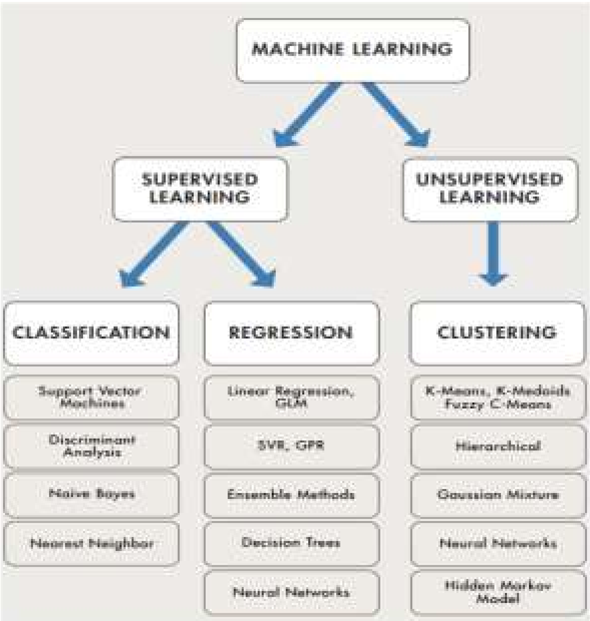

Figure 2. A summary of traditional machine learning methods

Source: http://www.datasciencecentral.com/profiles/blogs/machine-learning-summarized-in-one-picture

4. Result Analysis and Discussion

A more in-depth evaluation of the subject of artificial intelligence reveals ideas, thought, ambitions, and desires of man to progress to a higher level of human achievement. The world's rapid advancement towards an artificial future comes with worries and fear. The accuracy level of man is not infallible. Therefore, some loopholes may still prevail causing massive destruction at the epitome of the advancement. Below, the future of Artificial Intelligence and the possibilities of merits or demerits associated with it is discussed.

4.1 Artificial Narrow Intelligence (ANI)

The 21st century celebrates its achievement in this level of artificial progress. If a 19th-century corpse could resurrect today, it would be stunned by the rapid progress in technology vivid in this generation. Phones operated by fingerprints and unlocked by facial recognition are common advancements that a large population enjoys today. All credit goes to the obedient programs that are set to accomplish the given tasks. Narrow A.I. fitly pictures the constraints that the machines have been put to operate on. Human intelligence is here not replicated, but it merely simulates human behavior under a narrow set of parameters. Systems that excel at specific tasks, yet cannot apply their resources outside fairly narrow domains are said to have this Artificial Narrow Intelligence [11]. Examples of Narrow A.I. include:

(1) Self-driving cars
(2) Siri by Apple and other virtual assistants
(3) Email spam filters
(4) Disease prediction tools
(5) Face recognition software

Despite all its advantages, the selection of AI in numerous ventures is seen as a danger to low-and-middle talented specialists, as it will radically chop down dependence on the human labor force. The increasing level of unemployment created by the replacement of human labor by these machines is alarming and companies could gain a disproportionate advantage over conventional companies that still depend on normal, shift-based systems leading to unfair competition [12].

4.2 Artificial General Intelligence

The goal of AGI is the capacity to solve multiple problems, not just one [13]. Advanced AI, can mirror human behavior, and at times make conversations with humans forming relationships. This is capable of breaking the fabric holding the society as we know it.

4.3 Artificial Super Intelligence

Artificial Super Intelligence is the capability of human-made systems that can surpass humans [14]. The machine acts naturally mindful and goes past human insight and capacities. The ASI has a better memory and faster ability to process data. The reasoning and decision making of ASI will also be higher than that of humans. The thought of having such kind of machines sound appealing, especially when we imagine that we shall be sitting somewhere relaxed as we command the "non-exhausted humans" with ease. It is this level of technological achievement that scientists are busy striving to attain. The thought of the ASI taking over the dominion of man is also a speculated potential danger. Positively, malware identification techniques are used to improve cognitively-inspired inference [15].

5. Conclusion

Will the Artificial Super Intelligent level of technological advancement ever be achieved? In case it is completed, what of the accelerated hacking, A.I. terrorism, and more negative issues will be associated with them? Such questions and more may crisscross the mind of a thinker and possibly stigmatize the mind of a pessimist. If the current machines can still find their way to the hands of terrorists, the militia groups may even use the advanced devices for

  DOI: https://doi.org/10.30564/jcsr.v2i4.2534

evil. A thought like this poses a danger to the forthcoming generations. It appears more profitable to appreciate the level of advancement we have attained and think of how humanity can benefit from it other than focusing on the higher. In contrast, society suffers from them.

References

[1] M. O. Riedl. Human-centered artificial intelligence and machine learning. Hum. Behav. Emerg. Technol., 2019, 1(1): 33-36.

[2] E. Coronado, G. Venture, N. Yamanobe. Applying Kansei/Affective Engineering Methodologies in the Design of Social and Service Robots: A Systematic Review. Int. J. Soc. Robot., 2020.

[3] G. Leiras. European Journal of Public Health. 2020, 3(5): 2020.

[4] A. Zorins, P. Grabusts. Safety of artificial superintelligence. Vide. Tehnol. Resur. - Environ. Technol. Resour., 2019, 2: 180-183.

[5] P. Hunter. The advent of AI and deep learning in diagnostics and imaging. EMBO Rep., 2019, 20(7): 1-4.

[6] B. W. Israelsen, N. R. Ahmed. "Dave...I can assure you ...that it's going to be all right ..." A Definition, Case for, and Survey of Algorithmic Assurances in Human-Autonomy Trust Relationships. ACM Comput. Surv., 2019, 51(6): 1-37.

[7] K. Mizukami et al. Genetic characterization of pancreatic cancer patients and prediction of carrier status of germline pathogenic variants in cancer-predisposing genes. EBioMedicine, 2020, 60: 103033.

[8] F. Locatello et al. Challenging common assumptions in the unsupervised learning of disentangled representations. RML@ICLR 2019 Work. - Reprod. Mach. Learn., 2019.

[9] L. Han et al. Grid-wise control for multi-agent reinforcement learning in video game AI. 36th Int. Conf. Mach. Learn. ICML 2019, 2019: 4558-4571.

[10] I. Petukhov, L. Steshina, A. Glazyrin, D. Velev. Design Model of a Training Simulator in Virtual Reality. FASSI 2019 Fifth Int. Conf. Fundam. Adv. Softw. Syst. Integr., 2019: 1-7.

[11] H. Shevlin, K. Vold, M. Crosby, M. Halina. The limits of machine intelligence. EMBO Rep., 2019, 20(10): 1-5.

[12] S. M. Abdullah. Artificial Intelligence (Ai) and Its Associated Ethical Issues. Islam Civilisational Renew., 2019, 10(1): 124-126.

[13] R. Thawonmas, J. Togelius, G. N. Yannakakis. Artificial General Intelligence in Games : Where Play Meets Design and User Experience. NII Shonan Meet., 2019, 130.

[14] M. L. How. Future-ready strategic oversight of multiple artificial superintelligence-enabled adaptive learning systems via human-centric explainable ai-empowered predictive optimizations of educational outcomes. Big Data Cogn. Comput., 2019, 3(3): 1-43.

[15] R. Thomson, C. Lebiere, S. Bennati, P. Shakarian, E. Nunes. Malware identification using cognitively-inspired inference. 24th Conf. Behav. Represent. Model. Simulation, BRiMS 2015, co-located with Int. Soc. Comput. Behav. Model. Predict. Conf. SBP 2015, 2015: 18-25.

BILINGUAL
PUBLISHING CO.
Pioneer of Global Academics Since 1984

Tel:+65 65881289
E-mail:contact@bilpublishing.com
Website:www.bilpublishing.com