

Lecture Notes on Data Engineering
and Communications Technologies 111

Mukesh Saraswat · Harish Sharma ·
K. Balachandran · Joong Hoon Kim ·
Jagdish Chand Bansal *Editors*

Congress on Intelligent Systems

Proceedings of CIS 2021, Volume 2

Lecture Notes on Data Engineering and Communications Technologies

Volume 111

Series Editor

Fatos Xhafa, Technical University of Catalonia, Barcelona, Spain

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at <https://link.springer.com/bookseries/15362>

Mukesh Saraswat · Harish Sharma ·
K. Balachandran · Joong Hoon Kim ·
Jagdish Chand Bansal
Editors

Congress on Intelligent Systems

Proceedings of CIS 2021, Volume 2

 Springer

Editors

Mukesh Saraswat
Department of Computer Science &
Engineering and Information Technology
Jaypee Institute of Information Technology
Noida, India

Harish Sharma
Department of Computer Science
and Engineering
Rajasthan Technical University
Kota, India

K. Balachandran
Department of Computer Science
and Engineering
CHRIST (Deemed to be University)
Bengaluru, India

Joong Hoon Kim
Korea University
Seoul, Korea (Republic of)

Jagdish Chand Bansal
South Asian University
New Delhi, India

ISSN 2367-4512

ISSN 2367-4520 (electronic)

Lecture Notes on Data Engineering and Communications Technologies

ISBN 978-981-16-9112-6

ISBN 978-981-16-9113-3 (eBook)

<https://doi.org/10.1007/978-981-16-9113-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This proceedings contains the papers presented at the 2nd Congress on Intelligent Systems (CIS 2021), organized by CHRIST (Deemed to be University), Bangalore, and Soft Computing Research Society, during September 4–5, 2021. Congress on Intelligent Systems (CIS 2021) invited ideas, developments, applications, experiences, and evaluations in intelligent systems from academicians, research scholars, and scientists. The conference deliberation included topics specified within its scope. The conference offered a platform for bringing forward extensive research and literature across the arena of intelligent systems. It provided an overview of the upcoming technologies. CIS 2021 provided a platform for leading experts to share their perceptions, provide supervision, and address participants' interrogations and concerns. CIS 2021 received 370 research submissions from 35 different countries, viz. Algeria, Bangladesh, Burkina Faso, China, Egypt, Ethiopia, Finland, India, Indonesia, Iran, Iraq, Kenya, Korea, The Democratic People's Republic of Madagascar, Malaysia, Mauritius, Mexico, Morocco, Nigeria, Peru, Romania, Russia, Serbia, Slovakia, South Africa, Spain, Switzerland, Ukraine, United Arab Emirates, UK, USA, Uzbekistan, and Vietnam. The papers included topics about advanced areas in technology, artificial intelligence, machine learning, and data science. After a rigorous peer review with the help of program committee members and more than 100 external reviewers, 135 papers were approved.

CIS 2021 is a flagship event of the Soft Computing Research Society, India. The conference was inaugurated by Fr Dr. Abraham VM, Honorable Vice-Chancellor, CHRIST (Deemed to be University), Bangalore, and Chief Patron CIS 2021. Other eminent dignitaries include Prof. Joong Hoon Kim, General Chair, CIS 2021; Fr. Joseph Varghese, Patron, CIS 2021; and Prof. K. Balachandran, General Chair, CIS 2021. The conference witnessed keynote addresses from eminent speakers, namely Prof. Xin-She Yang, Middlesex University, The Burroughs, Hendon, London, Prof. P. Nagabhushan, Indian Institute of Information Technology Allahabad, Prof. J. C. Bansal, South Asian University, New Delhi, India, Prof. Lipo Wang,

Nanyang Technological University, Singapore, and Prof. Nishchal K. Verma, Indian Institute of Technology Kanpur, India.

Noida, India

Kota, India

Bengaluru, India

Seoul, Korea (Republic of)

New Delhi, India

Mukesh Saraswat

Harish Sharma

K. Balachandran

Joong Hoon Kim

Jagdish Chand Bansal

Contents

A Comprehensive Survey on Machine Reading Comprehension: Models, Benchmarked Datasets, Evaluation Metrics, and Trends	1
Nisha Varghese and M. Punithavalli	
A Novel Feature Descriptor: Color Texture Description with Diagonal Local Binary Patterns Using New Distance Metric for Image Retrieval	17
Vijaylakshmi Sajwan and Rakesh Ranjan	
OntoINT: A Framework for Ontology Integration Based on Entity Linking from Heterogeneous Knowledge Sources	27
N. Manoj, Gerard Deepak, and A. Santhanavijayan	
KnowCommerce: A Semantic Web Compliant Knowledge-driven Paradigm for Product Recommendation in E-commerce	37
N. Krishnan and Gerard Deepak	
Removal of Occlusion in Face Images Using PIX2PIX Technique for Face Recognition	47
Sincy John and Ajit Danti	
Age, Gender and Emotion Estimation Using Deep Learning	59
Mala Saraswat, Praveen Gupta, Ravi Prakash Yadav, Rahul Yadav, and Sahil Sonkar	
Qualitative Classification of Wheat Grains Using Supervised Learning	71
P. Sarveswara Rao, K. Lohith, K. Satwik, and N. Neelima	
Pandemic Simulation and Contact Tracing: Identifying Super Spreaders	81
Aishwarya Sampath, Bhargavi Kumaran, Vidyacharan Prabhakaran, and Cinu C. Kiliroor	

Assessment of Attribution in Cyber Deterrence: A Fuzzy Entropy Approach	97
T. N. Nisha and Prasenjit Sen	
Predictive Maintenance of Bearing Machinery Using MATLAB	107
Karan Gulati, Shubham Tiwari, Keshav Basandrai, and Pooja Kamat	
Fuzzy Keyword Search Over Encrypted Data in Cloud Computing: An Extensive Analysis	137
Manya Smriti, Sameeksha Daruka, Khyati Gupta, and S. Siva Rama Krishnan	
Application of Data Mining and Temporal Data Mining Techniques: A Case Study of Medicine Classification	151
Shashi Bhushan	
A Deep Learning Approach for Plagiarism Detection System Using BERT	163
Anjali Bohra and N. C. Barwar	
A Multi-attribute Decision Approach in Triangular Fuzzy Environment Under TOPSIS Method for All-rounder Cricket Player Selection	175
H. D. Arora, Riju Chaudhary, and Anjali Naithani	
Machine Learning Techniques on Disease Detection and Prediction Using the Hepatic and Lipid Profile Panel Data	189
Ifra Altaf, Muheet Ahmed Butt, and Majid Zaman	
Performance Analysis of Machine Learning Algorithms for Website Anti-phishing	205
N. Mohan Krishna Varma, Y. C. A. Padmanabha Reddy, and C. Rajesh Kumar Reddy	
Analytical Analysis of Two-Warehouse Inventory Model Using Particle Swarm Optimization	215
Sunil Kumar and Rajendra Prasad Mahapatra	
Towards an Enhanced Framework to Facilitate Data Security in Cloud Computing	227
Sarah Mahlaule, John Andrew van der Poll, and Elisha Oketch Ochola	
Kinematics and Control of a 3-DOF Industrial Manipulator Robot	243
Claudia Reyes Rivas, María Brox Jiménez, Andrés Gersnoviez Milla, Héctor René Vega Carrillo, Víctor Martín Hernández Dávila, Francisco Eneldo López Monteagudo, and Manuel Agustín Ortiz López	
Enhanced Energy Efficiency in Wireless Sensor Networks	255
Neetu Mehta and Arvind Kumar	

Social Structure to Artificial Implementation: Honeybees 271
 Amit Singh

Lifetime Aware Secure Data Aggregation through Integrated Incentive-based Mechanism in IoT-based WSN Environment 287
 S. Nandini and M. Kempanna

Multi-temporal Analysis of LST-NDBI Relationship with Respect to Land Use-Land Cover Change for Jaipur City, India 299
 Arpana Chaudhary, Chetna Soni, Uma Sharma, Nisheeth Joshi, and Chilka Sharma

Analysis and Performance of JADE on Interoperability Issues Between Two Platform Languages 315
 Jaspreet Chawla and Anil Kr. Ahlawat

A TAM-based Study on the ICT Usage by the Academicians in Higher Educational Institutions of Delhi NCR 329
 Palak Gupta and Shilpi Yadav

An Investigation on Impact of Gender in Image-Based Kinship Verification 355
 Vijay Prakash Sharma and Sunil Kumar

Classification of COVID-19 Chest CT Images Using Optimized Deep Convolutional Generative Adversarial Network and Deep CNN 363
 K. Thangavel and K. Sasirekha

Intelligent Fractional Control System of a Gas-Diesel Engine 379
 Alexandr Avsievich, Vladimir Avsievich, and Anton Ivaschenko

Analysis on Advanced Encryption Standard with Different Image Steganography Algorithms: An Experimental Study 391
 Alicia Biju, Lavina Kunder, and J. Angel Arul Jothi

Diabetes Prediction Using Logistic Regression and K-Nearest Neighbor 407
 Ami Oza and Anuja Bokhare

Linear Regression for Car Sales Prediction in Indian Automobile Industry 419
 Rohan Kulkarni and Anuja Bokhare

Agent-Driven Traffic Light Sequencing System Using Deep Q-Learning 431
 Palwai Thirumal Reddy and R. Shanmughasundaram

System Partitioning with Virtualization for Federated and Distributed Machine Learning on Critical IoT Edge Systems 443
 Vysakh P. Pillai and Rajesh Kannan Megalingam

A Review on Preprocessing Techniques for Noise Reduction in PET-CT Images for Lung Cancer	455
Kaushik Pratim Das and J. Chandra	
Optimal DG Planning and Operation for Enhancing Cost Effectiveness of Reactive Power Purchase	477
Nirmala John, Varaprasad Janamala, and Joseph Rodrigues	
Prediction of User’s Behavior on the Social Media Using XGBRegressor	491
Saba Tahseen and Ajit Danti	
Model Order Reduction of Continuous Time Multi-input Multi-output System Using Sine Cosine Algorithm	503
Aditya Prasad Padhy, Varsha Singh, and Vinay Pratap Singh	
Smart E-waste Management in China: A Review	515
Yafeng Han, Tetiana Shevchenko, Dongxu Qu, and Guohou Li	
Artificial Intelligence Framework for Content-Based Image Retrieval: Performance Analysis	535
Padmashree Desai and Jagadeesh Pujari	
Priority-Based Replication Management for Hadoop Distributed File System	549
Dilip Rajput, Ankur Goyal, and Abhishek Tripathi	
Comparing the Pathfinding Algorithms A*, Dijkstra’s, Bellman-Ford, Floyd-Warshall, and Best First Search for the Paparazzi Problem	561
Robert Johner, Antonino Lanaia, Rolf Dornberger, and Thomas Hanne	
Optimizing an Inventory Routing Problem using a Modified Tabu Search	577
Marc Fink, Lawrence Morillo, Thomas Hanne, and Rolf Dornberger	
Classification of Histopathological Breast Cancer Images using Pretrained Models and Transfer Learning	587
Mirya Robin, Aswathy Ravikumar, and Jisha John	
Handwritten Digit Recognition Using Very Deep Convolutional Neural Network	599
M. Dhilsath Fathima, R. Hariharan, and M. Seeni Syed Raviyathu Ammal	
The Necessity to Adopt Big Data Technologies for Efficient Performance Evaluation in the Modern Era	613
Sangeeta Gupta and Rupesh Mishra	

Forecasting Stock Market Indexes Through Machine Learning Using Technical Analysis Indicators and DWT 625
 Siddharth Patel, B. D. Vijai Surya, Chinthakunta Manjunath, Balamurugan Marimuthu, and Bikramaditya Ghosh

Slotted Coplanar Waveguide-Fed Monopole Antenna for Biomedical Imaging Applications 639
 Regidi Suneetha and P. V. Sridevi

A Framework for Enhancing Classification in Brain–Computer Interface 651
 Sanoj Chakkithara Subramanian and D. Daniel

Measuring the Accuracy of Machine Learning Algorithms When Implemented on Astronomical Data 667
 Shruthi Srinivasaprasad

Artificial Intelligence in E-commerce: A Literature Review 677
 Richard Fedorko, Štefan Král, and Radovan Bačík

Modified Non-local Means Model for Speckle Noise Reduction in Ultrasound Images 691
 V. B. Shereena and G. Raju

Improved Color Normalization Method for Histopathological Images 709
 Surbhi Vijn, Mukesh Saraswat, and Sumit Kumar

Parametric Analysis on Disease Risk Prediction System Using Ensemble Classifier 719
 Huma Parveen, Syed Wajahat Abbas Rizvi, and Praveen Shukla

Sentimental Analysis of Code-Mixed Hindi Language 739
 Ratnavel Rajalakshmi, Preethi Reddy, Shreya Khare, and Vaishali Ganganwar

Enhanced Security Layer for Hardening Image Steganography 753
 Pregada Akshita and P. P. Amritha

Matrix Games with Linguistic Distribution Assessment Payoffs 767
 Parul Chauhan and Anjana Gupta

Interval-valued Fermatean Fuzzy TOPSIS Method and Its Application to Sustainable Development Program 783
 Utpal Mandal and Mijanur Rahaman Seikh

An Empirical Study of Signal Transformation Techniques on Epileptic Seizures Using EEG Data 797
 M. Umme Salma and Najmusseher

Rainfall Estimation and Prediction Using Artificial Intelligence: a Survey	807
Vikas Bajpai, Anukriti Bansal, Ramit Agarwal, Shashwat Kumar, Namit Bhandari, and Shivam Kejriwal	
Image Classification Using CNN to Diagnose Diabetic Retinopathy	821
S. Arul Jothi, T. Mani Sankar, Rohith Chandra Kandambeth, Tadepalli Siva Koushik, P. Arun Prasad, and P. Arunmozhi	
Real-Time Segregation of Encrypted Data Using Entropy	835
P. Gowtham Akshaya Kumaran and P. P. Amritha	
Performance Analysis of Different Deep Neural Architectures for Automated Metastases Detection of Lymph Node Sections in Hematoxylin and Eosin-Stained Whole-Slide images	845
Naman Dangi and Khushali Deulkar	
Linguistic Classification Using Instance-Based Learning	863
Rhythm Girdhar, Priya S. Nayak, and Shreekanth M. Prabhu	
CoFFiT-COVID-19 Fake News Detection Using Fine-Tuned Transfer Learning Approaches	879
B. Fazlourrahman, B. K. Aparna, and H. L. Shashirekha	
Improved Telugu Scene Text Recognition with Thin Plate Spline Transform	891
Nandam Srinivasa Rao and Atul Negi	
Analysing Voice Patterns to Determine Emotion	901
Amit Kumar Bairwa, Vijander Singh, Sandeep Joshi, and Vineeta Soni	
Limacøn Inspired Particle Swarm Optimization for Large-Scale Optimization Problem	917
Shruti Gupta, Rajani Kumari, and Sandeep Kumar	
Author Index	931

About the Editors

Dr. Mukesh Saraswat is an Associate Professor at Jaypee Institute of Information Technology, Noida, India. Dr. Saraswat has obtained his Ph.D. in Computer Science and Engineering from ABV-IIITM Gwalior, India. He has more than 18 years of teaching and research experience. He has guided two Ph.D. students, more than 50 M.Tech. and B.Tech. dissertations, and presently guiding four Ph.D. students. He has published more than 50 journal and conference papers in the area of image processing, pattern recognition, data mining, and soft computing. He was the part of successfully completed DRDE funded project on image analysis and SERB-DST (New Delhi) funded project on Histopathological Image Analysis. He is currently running a projects funded by and Collaborative Research Scheme (CRS), Under TEQIP III (RTU-ATU) on Smile. He has been an active member of many organizing committees of various conferences and workshops. He is also a guest editor of the *Array*, *Journal of Swarm Intelligence*, and *Journal of Intelligent Engineering Informatics*. He is an active member of IEEE, ACM, CSI Professional Bodies. His research areas include Image Processing, Pattern Recognition, Mining, and Soft Computing.

Dr. Harish Sharma is an Associate professor at Rajasthan Technical University, Kota in Department of Computer Science and Engineering. He has worked at Vardhaman Mahaveer Open University Kota, and Government Engineering College Jhalawar. He received his B.Tech. and M.Tech. degree in Computer Engineering from Government Engineering College, Kota and Rajasthan Technical University, Kota in 2003 and 2009 respectively. He obtained his Ph.D. from ABV—Indian Institute of Information Technology and Management, Gwalior, India. He is secretary and one of the founder member of Soft Computing Research Society of India. He is a life time member of Cryptology Research Society of India, ISI, Kolkata. He is an Associate Editor of *International Journal of Swarm Intelligence (IJSI)* published by Inderscience. He has also edited special issues of the many reputed journals like *Memetic Computing*, *Journal of Experimental and Theoretical Artificial Intelligence*, *Evolutionary Intelligence*, etc. His primary area of interest is nature inspired optimization techniques. He has contributed in more than 105 papers published in various international journals and conferences.

Dr. K. Balachandran is currently a Professor and Head CSE at CHRIST (Deemed to be University), Bengaluru, India. He has total 38 years' experience in Research, Academia and Industry. He Served as Senior Scientific Officer in the Research and Development unit of Department of Atomic Energy for 20 years. His research interest includes Data Mining, Artificial Neural Networks, Soft computing, Artificial Intelligence. He has published more than fifty articles in well-known SCI/SCOPUS indexed international journals and conferences, and attended several national and international conferences and workshops. He has authored/edited four books in the area of computer science.

Prof. Joong Hoon Kim a faculty of Korea University in the School of Civil, Environmental and Architectural Engineering, obtained his Ph.D. degree from the University of Texas at Austin in 1992 with the thesis title "Optimal replacement/rehabilitation model for water distribution systems". Professor Kim's major areas of interest include: optimal design and management of water distribution systems, application of optimization techniques to various engineering problems, and development and application of evolutionary algorithms. His publication includes "A New Heuristic Optimization Algorithm: Harmony Search", Simulation, February 2001, Vol. 76, pp. 60–68, which has been cited over 2300 times by other journals of diverse research areas. His keynote speeches include "Optimization Algorithms as Tools for Hydrological Science" in the Annual Meeting of Asia Oceania Geosciences Society held in Brisbane, Australia in June of 2013, "Recent Advances in Harmony Search Algorithm" in the 4th Global Congress on Intelligent Systems (GCIS 2013) held in Hong Kong, China in December of 2013, and "Improving the convergence of Harmony Search Algorithm and its variants" in the 4th International Conference on Soft Computing For Problem Solving (SOCPROS 2014) held in Silcha, India in December of 2014. He hosted the 6th Conference of The Asia Pacific Association of Hydrology and Water Resources (APHW 2013) and the 2nd International Conference on Harmony Search Algorithm (ICHSA 2015). And he is hosting another one, the 12th International Conference on Hydroinformatics (HIC 2016) in August of 2016.

Dr. Jagdish Chand Bansal is an Associate Professor at South Asian University New Delhi and Visiting Faculty at Maths and Computer Science, Liverpool Hope University UK. Dr. Bansal has obtained his Ph.D. in Mathematics from IIT Roorkee. Before joining SAU New Delhi he has worked as an Assistant Professor at ABV—Indian Institute of Information Technology and Management Gwalior and BITS Pilani. His Primary area of interest is Swarm Intelligence and Nature Inspired Optimization Techniques. Recently, he proposed a fission-fusion social structure based optimization algorithm, Spider Monkey Optimization (SMO), which is being applied to various problems from the engineering domain. He has published more than 70 research papers in various international journals/conferences. He has worked with the Liverpool Hope University UK for various research projects. He is the editor in chief of the journal *MethodsX* published by Elsevier. He is the series editor of the book series *Algorithms for Intelligent Systems* (AIS) and *Studies in Autonomic,*

Data-driven and Industrial Computing (SADIC) published by Springer. He is the editor in chief of *International Journal of Swarm Intelligence* (IJSI) published by Inderscience. He is also the Associate Editor of *IEEE Access* published by IEEE and *Array* published by Elsevier. He is the general secretary of the Soft Computing Research Society (SCRS). He has also received Gold Medal at UG and PG level.

A Comprehensive Survey on Machine Reading Comprehension: Models, Benchmarked Datasets, Evaluation Metrics, and Trends



Nisha Varghese and M. Punithavalli

Abstract Machine reading comprehension (MRC) is a core process in question answering systems. Question answering systems are capable to extract the answer from relevant resources automatically for the questions posed by humans, and machine reading comprehension brings attention to a textual understanding with answering questions. Recent advancements in deep learning and natural language processing pave the way to improve the accuracy in question answering systems with the major developments in neural machine reading comprehension, transfer learning, deep learning-based information retrieval, and knowledge-based information extraction. Herein, this research analysis included the comprehensive analysis of MRC tasks, benchmarked datasets, classic models, performance evaluation metrics, and modern trends and techniques on MRC.

Keywords Baseline models · Benchmarked datasets · Evaluation metrics · MRC tasks · MRC types

1 Introduction

Machine reading comprehension (MRC) is an indispensable task in the question answering system and the machines learn and answering from the text by using the series of modules in the MRC system. Read, learn, and understand the information and answer the most relevant answer for the question is one of the tedious tasks for machines. Each module in MRC plays a vital role in the question answering system to deliver the most accurate answer. In the MRC system, machines can understand the input documents and prepare the answers based on the given context, and it is also one of the promising tasks for direct information retrieval from natural language

N. Varghese (✉) · M. Punithavalli
Department of Computer Applications, Bharathiar University, Coimbatore 641046, India
e-mail: nisha.dca@buc.edu.in

M. Punithavalli
e-mail: punithavalli@buc.edu.in

documents in a more efficient manner rather than a sequence of associated Web pages. The search engine retrieves an enormous amount of Web pages, here comes the role of MRC. The user can input the query and the context into the MRC system and the output extracted as the most relevant answer for the query. To recapitulate the MRC for a given passage P and question Q , MRC tasks can retrieve the accurate answer A to question Q by the function F , $A = F(P, Q)$.

Since 2015, neural machine reading comprehension is a topic that caught fire in researchers and NLP Community. Before that, MRC systems were presented with substandard performance due to the limited and insufficient datasets and manually designed syntactic-rule-based techniques and machine learning techniques. Typical mathematical and statistical techniques and methods like pattern matching, similarity checking between query and context with automated natural language techniques like stemming, lemmatization, named entity recognition are inadequate to extract the exact answer in question answering systems.

Present MRC systems can gain success easily with the emerging trends in NLP and DNN techniques such as transfer learning, word vectorization methods, transformers, RNN, LSTM, GRU, attention mechanisms, and pointer networks in sequence to sequence modeling. To answer a question, the major challenges associated with question answering systems are seeking out and explore the relevant resources and then recognize and integrate the chunks of information from various repositories. MRC requires a comprehensive understanding of the given context to answer the question is one of the hurdles. This paper investigated through the most relevant and highly profiled papers from 2015 to 2020, and the purpose of this survey is to analyze and summarize the most promising models in MRC, various tasks or types of MRC systems and its representative datasets, modules and approaches, emerging trends on MRC and corresponding challenges.

2 Materials and Methods

This section contains machine reading comprehension tasks with examples, MRC datasets examples, various types of MRC questions, a comprehensive description of the latest MRC trends, MRC architecture with the elaborated explanation of each module and the models, techniques, and methods using in corresponding modules.

2.1 Machine Reading Comprehension Tasks

Cloze Tests. These tests are gap-filling tests or fill in the blanks like question answering; this type of question contains one or more blank space or gap in question. To answer the questions, MRC systems should be emphasized on the comprehensive understanding of context and the usage of vocabulary to fill in the blank with an accurate answer. Cloze test, for a given passage P from which a word or element

$A(A \in P)$ is absent, then the model fills the gap with the most appropriate answer A by the function $F, A = F(P - \{A\})$.

Multiple-Choice Questions (MCQ). MCQs are like factoid questions, which contain a list of answers with one or more correct answers. Proper understanding of the context is required to choose the right answer from the options. The formal definition of MCQ, for a given the passage P , the question Q and possible answers list $A = \{A_1, A_2, \dots, A_n\}$, the MCQ task is to choose the accurate result A_i from $A (A_i \in A)$ by, $F A_i = F(P, Q, A)$.

Span Extraction. Span extraction reading comprehension is much more natural than the cloze style and MCQ MRC tasks, certain words are inadequate to answer the questions, and there may be no proper candidate answer in some cases. The formal definition of span extraction MCQ [1], for a given passage P , with n tokens, $P = \{t_1, t_2, \dots, t_n\}$, and the question Q , the task needs to extract the continuous subsequence $A = \{t_i, t_{i+1}, \dots, t_{i+k}\}$ from passage P as the exact answer to the question Q by using the function $F, A = F(P, Q)$. Answer A should be a span that is inferred from passage P .

Free Form Answering. Free form answering or open-domain question answering (ODQA) delivers the correct information as the answer to the questions from various huge relevant data repositories, which produces more accurate and flexible answers and reduces the limitations of the previous tasks due to inadequate passage and inconsistent choices. Free answering has some possibilities like the exact answer contains in the context, the entities are from multiple contexts, all entities contained in the context and question, part of entities in output is unexplored in the context or question, and all entities in the answer are not contained in the context or question. Free answering, for a given passage P and the question Q , the exact answer A may not be the part of the input text, which means, the answer may or may not be is a subsequence of P . The free form task needs to deliver the correct answer A by $F, A = F(P, Q)$.

2.2 Types of Machine Reading Comprehension Questions

Cloze Style. It is typically with a placeholder in a statement and maybe a declarative sentence, imperative sentence, or any other pattern. Answer to question the system needs to seek out an entity or phrase that is appropriate to fill in the blanks.

Natural Form. It is a question statement and that completely follows the grammatical rules of the language. For example, "What is your name?"

Synthetic Style. These types of questions are the arrangements of words and that need not conform to any grammar. The question in Fig. 1 is an example of synthetic style.

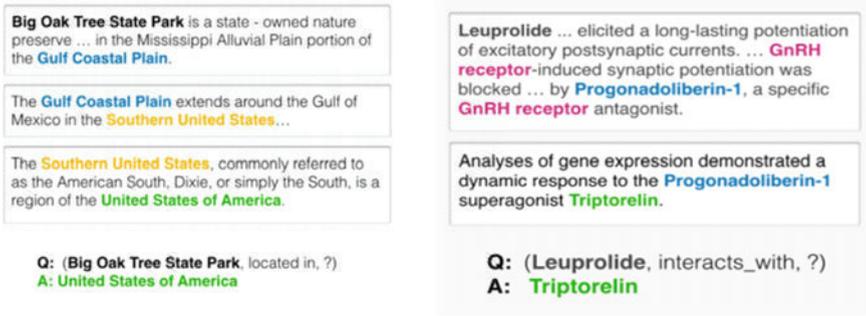


Fig. 1 Examples for synthetic style questions from QAngaroo dataset [2]

2.3 Latest Trends on Machine Reading Comprehension

In recent times, various machine reading comprehension datasets have been generated and that leads to the exponential increase in the research of the reading comprehension tasks and new trends. The type, size, model, characteristics, and application or approach to each model are varying from dataset to dataset. The corpus type is mainly divided into multi-modal and textual.

Multi-modal machine reading comprehension is an emerging trend in the QA system, which needs a comprehensive understanding of the textual data and the visual data. The textual corpus contains only plain texts, whereas, in the multi-modal MRC dataset, emphasis more on the given context of the text, architectures, images, audio, diagrams, and sometimes videos are also included to answer the multi-modal questions. Figure 2 is the example for multi-modal MRC (M^3RC), taken from TQA [3], and the image comprises the rich diagram parsing category. Answer is extracted based on the relevant information from the question and image with textual contents.

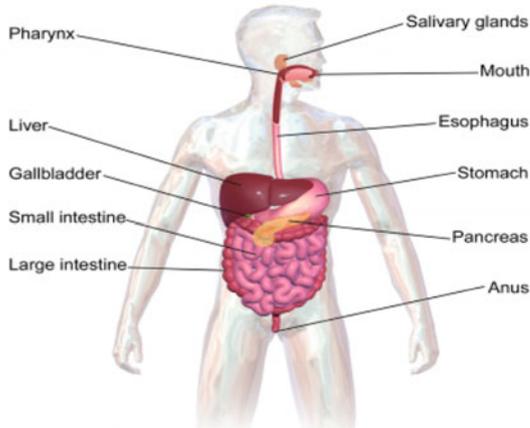
The latest challenges and trends are including MRC with unanswerable questions, reading comprehensions such as multi-hop, multi-modal, paraphrased paragraph, commonsense or world knowledge required MRC, complex logical reasoning, large-scale MRC dataset, the domain-specific MRC, open-domain (ODQA) dataset, and conversational MRC.

MRC with Unanswerable Questions. The complexity of the questions increases with a lack of evidence to answer the question. Then the system will answer the questions without any relevant content. It raises the problem of incorrect answer prediction. Consequently, these types of questions are challenging and difficult to retrieve answers from the context.

Multi-Hop Reading Comprehension (MHRC). The MHRC is effortful with the multi-passage and complex corpora. The model requires more emphasis on these kinds of documents with multi-hop searching and reasoning. The datasets for MHRC are SQuAD 2.0 [4], MS MARCO [5], natural questions [6] and NewsQA [7], and all these datasets are the example for MRC with unanswerable questions.

Q: This is the long narrow tube that carries food from the pharynx to the stomach.

- a. mouth
- b. salivary glands
- c. liver
- d. esophagus



The Components of the Digestive System

Fig. 2 Example for M^3RC from textbook question answering; the answer extraction requires the combined analysis of question and image with textual content [3]

Complex Reasoning MRC. Logic and complex reasoning are one of the most important challenges in the MRC. Facebook CBT bAbI [8], DROP [9], RACE [10], and CLOTH [11], LogiQA [12] are the datasets generated for logical reasoning.

Multi-modal Reading Comprehension (MMRC). Some complex or clear-cut concepts, architectures, logic diagrams, circuit diagrams can represent efficiently using images than text or a combination of both. The MMRC task includes heterogeneous data, which is a combination of images and language, which increases reading comprehension as an arduous task that may contain architectures, circuit diagrams, graphs, and much more with various colors and patterns. The representative dataset for the M^3RC is TQA [3], RecipeQA [13], COMICS [14], and MovieQA [15].

MRC with Commonsense or World Knowledge. Textual data is highly ambiguous and problematic due to word sense ambiguities like polysemous words and paraphrase sentences, semantic similarity conflicts, named entity recognition, and aspect term identification, and human language is more complex than a textual document. So, machine reading comprehension requires commonsense or world knowledge to answering these kinds of questions. Representative datasets are CommonSenseQA [16], ReCoRD [17], and OpenBookQA [18].

Conversational Machine Reading Comprehension (CMRC). The conversation is the natural way of communication; it includes actions or gestures, and the anaphoric, cataphoric, and exophoric references in compound statements are the other complex things to detect in the conversational MRC. To alleviate the problem, in the process of MRC on conversation needs, a proper identification of the previous conversations

and historical data on the conversation. Some of CMRC datasets are CoQA [19], QuAC [20], DREAM [21], and ShARC [22].

Domain-Specific Datasets. The context of a domain-specific dataset are taken from a specific domain, such as academics, cooking, health, or clinical reports. CliCR [23] is a medical domain cloze style MRC dataset, SciQ [24] is an MCQ–MRC dataset questions in science exam generated by crowdsourcing, and COVID-QA consists of 2019, COVID-19 questions which have been annotated into a broad category of transmission, prevention, and more information concerning COVID-19. Some other representative datasets are ReviewQA [25], SciTail [26], and PaperQA [27].

MRC with Paraphrased Paragraph. It alludes to the rephrasing passages using synonyms and altered words without changing the meaning of the context, which means the source context and paraphrased context are semantically similar. To extract the right answer to these questions in the task of paraphrased MRC, the system requires a comprehensive understanding of the true meaning of different versions of context. Some of the representative datasets for a paraphrased paragraph are DuoRC [28] and Who-did-What [29].

Open-Domain Question Answering (ODQA). The ODQA is a kind of task of identifying answers to natural questions from a large corpus of documents. Initially, the open-domain QA system starts with information retrieval to select a subset of documents from the corpus and collections of unstructured documents. Some of the representative datasets for ODQA are MCTest [30], CuratedTREC [31], Quasar [32], and SearchQA [33].

3 MRC Architecture, Modules, and Baseline Models

Neural machine reading comprehension (NMRC) is the core process in the question answering system. In NMRC process, context and question are the inputs and produce the answer as results. Figure 3 depicts the MRC system which includes four phases: word embedding, feature extraction, context–question interaction, and answer prediction.

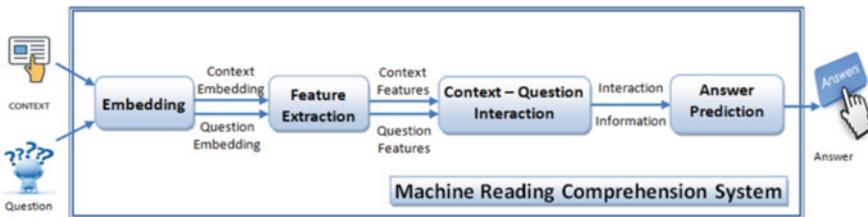


Fig. 3 Machine reading comprehension architecture

3.1 Word Embedding

Word embedding is used to represent the text tokens by fixed-size vectors, and then, the words which have related meaning have represented in a close vector space area. The conversion of the text tokens into vectors is known as vectorization. These word vectors can extract the information based on the meaning and structure of the input context and the relationship between words in the context. The meaning of a word depends on the context of words surrounded by it.

3.2 Feature Extraction

This phase takes the question and context embedding as input and extracts features individually. It also checks out the semantic and structural information of the sentences from the output of the word embedding. Some typical deep neural networks applied to mine the contextual features are recurrent neural networks (RNN), convolutional neural networks (CNN), and transformers.

3.3 Context–Question Interaction

Context–question interaction module extracts the semantic association between the context and the question; herein the MRC provides more focus to the context and question to transfer the most accurate information to the answer prediction module. Attention mechanisms and one-hop or multi-hop interaction are used to capture the relevant content from the input. The attention mechanism has concentrated more in an input sequence and decides the position of the context in sequence with higher precedence. One-hop interaction has calculated the context–query interaction only once, but in multi-hop interaction, the context–query interaction calculation may be more than once. Multi-hop interaction includes some methods to implement, first computes the semantics between the context and query, then introduces memory spaces to carry the prior memories process of the interaction, and finally, the hidden states of RNN store the previous information of context–question interaction.

3.4 Answer Prediction

Answer prediction is the final module, which delivers the answers to questions based on the actual context and also according to the specific tasks. To recapitulate, there are four types of answer prediction: word predictor, option selector, span extractor, and answer generator. Word predictor predicts the answer for close style tests and fills

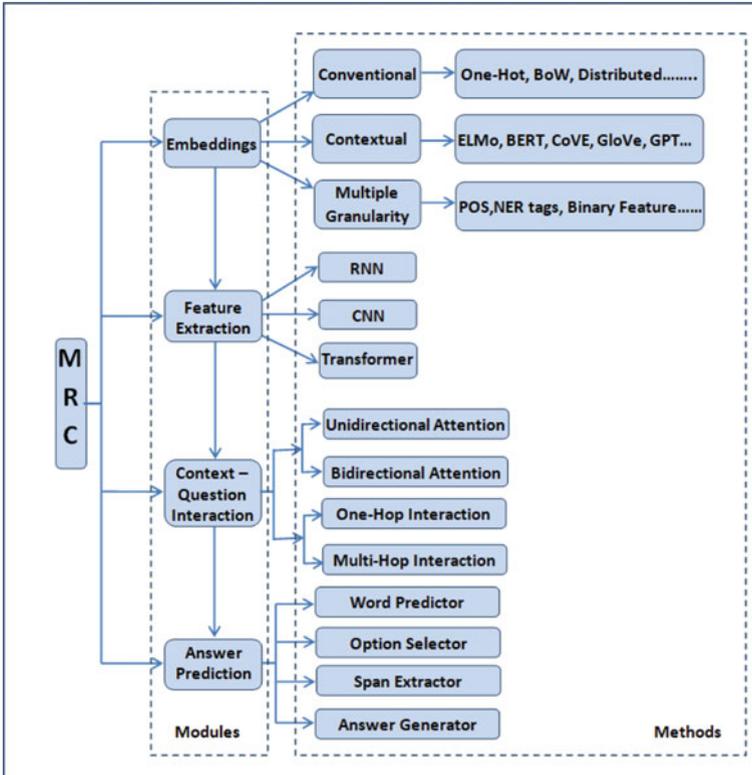


Fig. 4 Machine reading comprehension modules and methods of machine reading comprehension system

the corresponding words or entities to the place holder. Option selector selects the correct answer from candidate options for the multiple-choice tasks. Span extractor retrieves the answer for a question from a specific input context. In answer generator, the answers do not need to be exactly the subsequence of the input context, and answers may be generated from multiple pieces of evidence in other contexts. Figure 4 illustrates all modules and associated methods of MRC.

4 Performance Evaluation Metrics

Performance evaluation metrics measure the correctness of the MRC tasks using metrics. Evaluation metrics have been chosen for a specific dataset based on its characteristics. Some datasets may require more than one metric. Accuracy is the most common performance evaluation metrics for MRC models, and exact match, precision, recall, F1, ROUGE, BLEU, HEQ, and METEOR are the other metrics

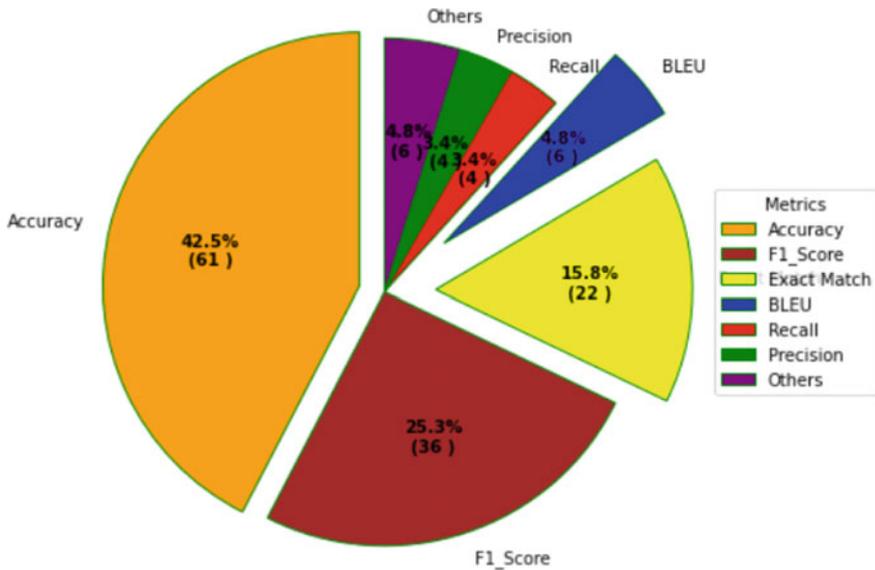


Fig. 5 Statistics of performance evaluation metrics

followed by accuracy. For multiple-choice or cloze style tasks are usually evaluated using accuracy, and the evaluation metrics such as exact match, precision, recall, and F1measure are usually used for span extraction–prediction tasks. ROUGE and BLEU are the popular measures for machine translation; recently, both metrics are using for the MRC tasks. HEQ is the latest evaluation metric in MRC and is most suitable for conversational MRC tasks. METEOR includes the features like synonym matching, and the other versions show higher accuracy in paraphrase matching, text normalization, proper interaction between content word and function word. The statistical analysis results of performance evaluation metrics included in Fig. 5. The pie chart shows that accuracy is the most common evaluation metric with 42.5%.

4.1 Accuracy

It is the percentage of the queries that correctly answers by an MRC system accurately, which means the ratio between the total number of questions (N) in MRC and the number of correctly answered questions (M). The accuracy can be calculated as in Eq. (1).

$$\text{Accuracy} = \frac{M}{N} \tag{1}$$

4.2 *Exact Match*

Exact match (EM) is the percentage of queries that match the correct answer with the system-generated answer; the answers can be an entity, phrase, or a sentence, but both answers should be equal. In the case of span extraction, both accuracy and the EM are the same. But in the case of factoid questions, EM does not work well, because there may or may not be a situation where the answer includes a part of the actual answer. The EM can be calculated using Eq. (2).

$$\text{Exact Match} = \frac{M}{N} \quad (2)$$

4.3 *Precision and Recall*

To define precision and recall, it is necessary to understand true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). False positives and false negatives affect the accuracy; these values occur when your actual value opposes the predicted value. Precision and recall are two types, token-level and question-level. Precision, recall, and accuracy can be calculated using Eqs. (3), (4) and (5).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (5)$$

4.4 *F1 Score (F Score or F Measure)*

It shows the balance between precision and recall, and it can be calculated using Eq. (6). Similar to precision and recall, the F1 score also in token-level F1 and question-level F1. Token-level F1 for a single question can be calculated as in Eq. (7), and the question-level F1 as in Eq. (8).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

$$F1_{TS} = \frac{2 \times \text{Precision}_{TS} \times \text{Recall}_{TS}}{\text{Precision}_{TS} + \text{Recall}_{TS}} \quad (7)$$

$$F1_Q = \frac{2 \times \text{Precision}_Q \times \text{Recall}_Q}{\text{Precision}_Q + \text{Recall}_Q} \quad (8)$$

4.5 Recall-Oriented Understudy for Gisting Evaluation (ROUGE)

ROUGE [34] is a performance evaluation metric for text summary systems, machine translation, MRC systems, and other NLP tasks. It includes five evaluation metrics.

ROUGE-N—ROUGE-1 (unigram), ROUGE-2 (bigrams), ROUGE-3 (trigrams) and so on, ROUGE-L—Longest Common Subsequence (LCS)-based statistics, ROUGE-W—Weighted LCS-based statistics, ROUGE-S—Skip-bigram-based co-occurrence statistics, ROUGE-SU—Skip-bigram plus unigram-based co-occurrence statistics. **ROUGE-N** is an n-gram recall between a candidate summary and a set of reference summaries and can be calculated as in Eq. (9).

$$\text{ROUGE-N} = \frac{\sum_{S \in \{RS\}} \sum_{\text{gram}_n \in S} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{S \in \{RS\}} \sum_{\text{gram}_n \in S} \text{Count}(\text{gram}_n)} \quad (9)$$

4.6 Bilingual Evaluation Understudy (BLEU)

Auto-machine translation represents the quality of a model using the BLEU [35] score, which shows the similarity between the context texts to the reference texts; the similarity between the context text and the reference text is higher than the values closer to one. BLEU can be calculated using Eq. (10). iBLEU is an interactive version of BLEU, which compares two different systems visually and interactively and also helps to visually estimate the BLEU scores obtained by the candidate translations. The interpretations of the BLEU score are like the following Table 1.

$$\text{BLEU} = \min\left(1, \exp\left(1 - \frac{\text{reference} - \text{length}}{\text{output} - \text{length}}\right)\right) \left(\prod_{i=1}^4 \text{Precision}_i\right)^{1/4} \quad (10)$$

Table 1 BLEU score and interpretation

BLEU score	Interpretation
< 10	Almost useless
10–19	Hard to get the gist
20–29	The gist is clear but has significant grammatical errors
30–40	Understandable to good translations
40–50	High-quality translations
50–60	Very high quality, adequate, and fluent translations
60–69	Quality often better than human

4.7 Metric for Evaluation of Translation with Explicit Ordering (METEOR)

METEOR [36] indicator uses recall and accuracy together to evaluate the model and based on the harmonic mean of unigram precision and recall, with recall weighted higher than precision. Unlike at the corpus level as in BLEU, the METEOR interlinks the better correlation with a sentence level or segment level human judgment. METEOR can be calculated using Eq. (11), where m is the total number of matched words, ch number of chunks, and α , β , γ are the parameters are tuned to maximize correlation with human judgments. The penalty and F_{mean} can be estimated through Eqs. (12) and (13), respectively.

$$\text{Meteor} = F_{\text{mean}} \times (1 - \text{Penalty}) \quad (11)$$

$$F_{\text{mean}} = \frac{\text{Precision} \times \text{Recall}}{\alpha \times \text{Precision} + (1 - \alpha) \times \text{Recall}} \quad (12)$$

$$\text{Penalty} = \gamma \times \left(\frac{ch}{m}\right)^\beta \quad (13)$$

4.8 Human Equivalence Score (HEQ)

The HEQ [20] is a new MRC evaluation metric, which estimates how the model is performing better than an ordinary person and that can be used in conversational MRC datasets, such kind of datasets contains questions with multiple valid answers, and then, the F1 score may be misleading. The HEQ score is computed by (14).

$$\text{HEQ} = \frac{M}{N} \quad (14)$$

5 Summary and Conclusion

This research article has presented a comprehensive survey on neural machine reading comprehension. First, it explained the need for an MRC system. MRC system contains four modules, word embedding, feature extraction, question–context interaction, and answer prediction. First, the system needs input as question and context; after that, both question and context will convert as word embedding. Any of the vectorization methods can be chosen for the embedding process, the output of the embedding transfer to the feature extraction module. In the second step, the most essential features for the answer prediction will be captured. For this process, most of the recent models use the transformer-based methodology, which emphasized transfer learning and fine-tuning. Then the extracted features transfer to the context–question interaction module, and finally, the system will predict the answer for the question. Based on the trends and pattern, the selection of a model to extract the accurate answer is an arduous task.

For the statistical analysis, 55 benchmarked datasets have been considered from the period of 2015–2020, and these datasets included textual and multi-model patterns. The tremendous increase in the large-scale datasets in MRC has contributed to the highly accurate answers in the question answering system, instead of producing the passages with chunks of spreading answers. Various types of MRC and the sources of generation of the dataset are also included in the statistical analysis. In addition to various performance evaluation metrics, the statistical analysis of metrics and the need and role of each metrics in different tasks are also incorporated.

With the development of new methods and models, especially the evolution of the transformer model in 2017, it has shown exceptional and dramatically accelerated performance not only in MRC but also for all natural language processing and deep learning tasks. With the evolution of the transformer, Google developed a pertained model BERT; in addition to that, numerous amounts of BERT versions have been generated, which makes higher accuracy in the leader board of every benchmarked dataset.

References

1. Liu S, Zhang X, Zhang S, Wang H, Zhang W (2019) Neural machine reading comprehension: methods and trends. *arXiv:1907.01118v5*. [cs.CL], 19(5), 1–45. <https://doi.org/10.3390/app9183698>
2. Welbl J, Stenetorp P, Riedel S (2018) Constructing datasets for multi-hop reading comprehension across documents. *arXiv preprint*, 18(6):287–302. <https://arxiv.org/abs/1710.06481>
3. Kembhavi A, Seo M, Schwenk D, Choi J, Farhadi A, Hajishirzi H (2017) Are you smarter than a sixth grader? Textbook question answering for multimodal machine comprehension. *Proc IEEE Conf Comput Vis Pattern Recognit* 17(3):4999–5007

4. Rajpurkar P, Jia R, Liang P (2018) Know what you don't know: unanswerable questions for SQuAD. In: Proceedings of the 56th annual meeting of the association for computational linguistics, Association for Computational Linguistics, 8(1), pp 784–789. <https://doi.org/10.18653/v1/P18-2124>
5. Nguyen T, Rosenberg M, Song X, Gao J, Tiwary S, Majumder R, Deng L (2016) MS MARCO: a human generated machine reading comprehension dataset. CoRR, [1611.09268], 18(3), 1–11. <http://arxiv.org/abs/1611.09268>
6. Kwiatkowski T, Palomaki J, Redfield O, Petrov S (2019) Natural questions: a benchmark for question answering research. Trans Assoc Computat Linguist 19(7):453–466. https://doi.org/10.1162/tacl_a_00276
7. Trischler A, Wang T, Yuan X, Harris J, Sordoni A, Bachman P, Suleman K (2017) NewsQA: a machine comprehension dataset. In: Proceedings of the 2nd workshop on representation learning for NLP; Association for computational linguistics, 17(1), 191–200. <https://doi.org/10.18653/v1/W17-2623>
8. Hill F, Bordes A, Chopra S, Weston J (2016) The Goldilocks principle: reading children's books with explicit memory representations. In: 4th International conference on learning representations, ICLR, 4(2), 1–13. <https://arxiv.org/abs/1511.02301>
9. Dua D, Wang Y, Dasigi P, Stanovsky G, Singh S, Gardner M (2019) DROP: a reading comprehension benchmark requiring discrete reasoning over paragraphs. Proc NAACL 19(1):2368–2378. <https://doi.org/10.18653/v1/N19-1246>
10. Lai G, Xie Q, Liu H, Yang Y, Hovy E (2017) RACE: Large-scale reading comprehension dataset from examinations. In: Proceedings of the 2017 conference on empirical methods in natural language processing; association for computational linguistics, vol 17(1), pp 785–794: Copenhagen, Denmark. <https://doi.org/10.18653/v1/D17-1082>
11. Xie Q, Lai G, Dai Z, Hovy E (2018) Large-scale Cloze test dataset created by teachers. In: Proceedings of the 2018 conference on empirical methods in natural language processing; association for computational linguistics, vol 18(1), pp 2344–2356. <https://doi.org/10.18653/v1/D18-1257>
12. Liu J, Cui L, Liu H, Huang D, Wang Y, Zhang Y (2020) LogiQA: a challenge dataset for machine reading comprehension with logical reasoning. [cs.CL], 7(1), 1–7. <https://arxiv.org/abs/2007.08124>
13. Yagcioglu S, Erdem A, Erdem E, Ikizler-Cinbis N (2018) RecipeQA: a challenge dataset for multimodal comprehension of cooking recipes. In: Proceedings of the 2018 conference on empirical methods in natural language processing, vol 18(1), pp 1358–1368. <https://doi.org/10.18653/v1/D18-1166>
14. Iyyer M, Manjunatha V, Guha A, Vyas Y, Boyd-Graber J, Daume H, Davis LS (2017) The amazing mysteries of the gutter: drawing inferences between panels in comic book narratives. In: Proceedings of the IEEE conference on computer vision and pattern recognition, vol 18(2), pp 7186–7195. <https://arxiv.org/abs/1611.05118>
15. Tapaswi M, Zhu Y, Stiefelhagen R, Torralba A, Urtasun R, Fidler S (2016) Movieqa: understanding stories in movies through question-answering. In: Proceedings of the IEEE conference on computer vision and pattern recognition, vol 6(2), pp 4631–4640. <https://arxiv.org/abs/1512.02902>
16. Talmor A, Herzig J, Lourie N, Berant J (2019) CommonsenseQA: a question answering challenge targeting commonsense knowledge. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, Volume 1 (Long and Short Papers), vol 19(1), pp 4149–4158. <https://doi.org/10.18653/v1/N19-1421>
17. Zhang S, Liu X, Liu J, Gao J, Duh K, Durme VB (2018) Record: bridging the gap between human and machine commonsense reading comprehension. arXiv preprint, 18(1):1–14. <https://arxiv.org/abs/1810.12885>
18. Mihaylov T, Clark P, Khot T, Sabharwal A (2018) Can a suit of armor conduct electricity? A new dataset for open book question answering. EMNLP 18(9):1–14. <https://www.aclweb.org/anthology/D18-1260.pdf>