

Computational Intelligence and State-of-the-Art Data Analytics

William J. Yeager

Retired

Knowledge Systems Laboratory, Stanford University
Visiting Professor at iUniversity, Tokyo, Japan
byeager@fastmail.fm

Jean-Henry Morin

Institute of Information Service Science
University of Geneva, Switzerland
Jean-Henry.Morin@unige.ch

We are now in the era of, “Do everything on the Internet.” The amount of Internet data collected and stored in data warehouses has become explosive.

Among the long list of practitioners and technologists who must analyze and extract knowledge from this data we find members of large Internet companies; governments at all levels; Manufacturing; Advertising and Marketing; Health Care; Biomedical, and pharmaceutical companies; and their complementary research communities.

Each requires the use of a subset of the available state-of-the-art Data Analytic algorithms to extract their desired knowledge. For each such domain specific instance of these algorithms, the knowledge they extract from this data provides the “intelligence” that is indispensable for day-to-day operations, and research.

To this end, we looked for contributions from areas that included Artificial Intelligence, Machine Learning, knowledge graphs, and Expert Systems as well as natural language Processing for unstructured text, search, translation, questions and answers.

Given the above we selected the three following papers from the candidate papers that we received:

The first is entitled, “Book Success Prediction with Pretrained Sentence Embeddings and Readability scores.” The authors are Aminal Islam of the University of Louisiana at Lafayette; and Muhammad Khalifa of the University of Michigan. The authors’ model uses a convolutional neural network over pretrained sentence embeddings and leverages different readability scores through a simple concatenation operation as applied to Goodreads.com’ user datasets.

The second paper is entitled, “Rapid Selection of Machine Learning Models Using Greedy Cross Validation.” The author is Daniel S. Soper of California State University at Fullerton, California. This paper introduces a greedy method of performing k-fold cross validation and shows how the proposed greedy method can be used to rapidly identify optimal or near-optimal machine learning (ML) models.

The final paper is entitled, “An Algorithm for Generating Gap-Fill Multiple Choice Questions for an Expert System.” The authors are Pornpat Sirithumgul of Rajamangala University of Technology Phra Nakhon, Thailand; Pimpaka Prasertsilp of Sukhothai Thammathirat Open University, Thailand; and Lorne Olfman of Claremont Graduate University, Claremont, California. This research is aimed to propose an artificial intelligence algorithm comprising an ontology-based design, text mining, and natural language processing for automatically generating gap-fill multiple choice questions. These questions are to be ultimately used by an AI Expert System.

