

Deep Learning Approaches for Big Data Analysis

Prof. Dr. Naomie Salim
Faculty of Computing, Universiti Teknologi Malaysia
Email: naomie@utm.my

Good representations of data eliminate irrelevant variability of the input data, while preserving the information that is useful for the ultimate task. Among the various ways for learning representation is using deep learning methods. Deep feature hierarchies are formed by stacking unsupervised modules on top of each other, forming multiple non-linear transformations to produce better representations. In this talk, we will first show how deep learning is used for bioactivity prediction of chemical compounds. Molecules are represented as several convolutional neural networks to predict their bioactivity. In addition, a new concept of merging multiple convolutional neural networks and an automatic learning features representation for the chemical compounds was proposed using the values within neurons of the last layer of the CNN architecture. We will also show how the concepts of deep learning is adapted into a deep belief network (DBN) to enhance the molecular similarity searching. The DBN achieves feature abstraction by reconstruction weight for each feature and minimizing the reconstruction error over the whole feature set. The DBN is later enhanced using data fusion to obtain a lower detection error probability and a higher reliability by using data from multiple distributed descriptors. Secondly, we will show how we used deep learning for stock market prediction. Here, we developed a Deep Long Short Term Memory Network model that is able to forecast the crude palm oil price movement with combined factors such as other commodities prices, weather and news sentiments and price movement of crude palm oil. We also show how we combined stock markets price and financial news and deployed the Long Short Term Memory (LSTM), Recurrent Neural Network (RNN), and Word 2 Vector (Word2Vec) to project the stock prices for the following seven days. Finally, we will show how we exploited deep learning method for the opinion mining and later used it to extract the product's aspects from the user textual review for recommendation systems. Specifically, we employ a multichannel convolutional neural network (MCNN) for two different input layers, namely, word embedding layer and Part-of-speech (POS) tag embedding layer. We show effectiveness of the proposed model in terms of both aspect extraction and rating prediction performance.

Biography



Professor Naomie Salim's main research goal is to design of new algorithms to improve the effectiveness of searching and mining new knowledge from various kinds of datasets, including unstructured, semi-structured and structured databases. The current focus of her research is on chemical databases and text databases to support the process of computer-aided drug design, text summarisation, plagiarism detection, automatic information extraction, sentiment analysis and recommendation systems.

Professor Naomie Salim has been involved in 51 research projects out of which she heads 20 of the projects. The projects are in collaboration with colleagues within UTM or with external organisations and communities, to a total value of RM 6.18 million.

She has also authored over 170 journal articles out of which 138 journal articles are indexed under Scopus and 90 conference papers describing research into novel techniques for computerised information retrieval, with particular reference to textual, chemical and biological information. She has produced 9 book chapters, published under American Chemical Society and UTM Press. Her Google Scholar h-Index is 19, her i10index is 44 and she has 1594 citations to date. Her Scopus H-index is 13 with 634 Scopus-indexed citations. Professor Salim has one patent granted for probabilistic based molecular searching technique, 10 patent filing, 46 copyrights and 1 trademark registered. Professor Salim has had 23 PhD research students graduated under her supervision.

Among the research and innovation awards received by Professor Naomie Salim are the PECIPTA 2011 Gold Medal award for her UTMCLP cross-language semantic plagiarism detection system, the I-inova 2010 Gold Medal award for her Islamic Ontology-based Quran search engine, BioInnovation 2011 Bronze Award for UTMChem Workbench Molecular Database System, iPhex Gold Medal Award for innovation in teaching and learning, UTM 2011 Best Research Award, UTM 2014 Best Research Award and the INATEX Distinction Award (1998). She has also won the UTM Citra Karisma Indexed Journal Paper Award for 2009, 2011, 2012, 2013, and 2014. She has also received a fellowship from Japan Society for the Promotion of Science (JSPS) in 2016 to study application of deep neural networks and deep bayesian networks in computer aided drug design.

The output of the research has been incorporated into a number of software such as UTMChem Workbench and NADI Natural Products Database System to support drug design and drug optimisation process, UTMCLPD Cross Language Plagiarism Detection System to summarise documents and check for plagiarism and Orichack for cross-language idea similarity checking and plagiarism detection. The systems can be used by pharmaceutical scientists to search, retrieve, optimize and discover new drug compounds from chemical and natural product databases and help academic institutions preserve academic integrity by providing support to detect intelligent, idea plagiarism across different languages.