International Youth Nuclear Congress 2016, IYNC2016, 24-30 July 2016, Hangzhou, China

Natural Language Process: A New Kind of Nuclear Quality Assurance Management Tool

Yongqing GUAN[a,b,*], Yan SUN[a], Zeyu WANG[c], Qiyuan Zheng[b]

aSchool of Software,Shanghai Jiao Tong University, No.800 Dongchuan Road, Shanghai, 200030, P.R.CHINA
bState Nuclear Power Engineering Company, SNPTC, Building 2, No.888 Tianlin Road, Shanghai, 200233, P.R.CHINA
cSISE

Abstract

Quality Assurance Management is imperative in nuclear engineering. And quality assurance language is a rigorous and highly logical language in workplaces. We build machine-learning models of natural language process about quality assurance management activities to extract valuable information from texts for management intelligently. As technological means, the primary purpose of NLP tools here is that converting massive unstructured data (text) to structured data (data attribute relationship).The tasks include event classification, named entity recognition (NER) of engineering, event domain judgment, automatic summarization and event similarity computing. We focus on Labeled-LDA and SVMs algorithms to perform short text classification. And using them as primary content, we can perform more advanced nuclear quality assurance management in future.

1. Introduction

Quality Assurance is a systematic process of checking to see whether a product or service being developed is meeting specified requirements.[1] . It can provide enough confidence for administrators in all kinds of industries. In nuclear energy domain, the meaning of nuclear quality assurance is to ensure that "Nuclear Safety" target is effectively implemented. Nuclear quality assurance involves incident investigation , training, experience feedback and management of enterprise knowledge. Many countries are considering building the Generation III reactor, such as APWR (AP1000, EPR), ABWR etc. Nuclear Power Plants (NPPs) become more and more involved. We have to

*Corresponding author. Yongqing GUAN
Email address:* g.jkron@gmail.com

face more risks on information management in construction or operation. Today, in this information era, we have more tools to collect and process information. With the advent of big data technologies, cloud computing and artificial intelligence, people can use better tools to rapidly and deeply process high-dimension data, which was unbelievable in before. In the past, many world top 500 enterprises had developed a kind of expert system, which modelled the solution process on human expert's issue for subsequent usage. But the consequence was unsatisfied due to the limit of data storage and lack of computing capacity. Big data technology has changed this status. In various of industries, quality assurance management is regarded as a vital work, which relies on the high-quality input of the enterprise management data. Now, we define the establishment of the quality assurance information system as the Quality Assurance Engineering (QAE). In that way, the current quality assurance project will face more and more text type (unstructured) data, which are waiting for organising and processing.

Pattern Recognition is a vital capacity of a human. It is also a necessary task in artificial intelligence (AI). If we define the "environment" and object as "pattern", so that the pattern recognition in the artificial intelligence field could be defined as using statistics technique to study the automatic processing and interpretation of pattern through computers. Christopher M. Bishop in his famous article Pattern Recognition and Machine Learning[2] mentioned that the "pattern recognition" came from the industrial field and "machine learning" from computer science, but some activities from them could be considered two sides of one field. With the popularity of machine learning, researchers hoped that using computers to obtain higher confidence interpretation from large statistical samples, to extend the theories, methods, and technique of artificial intelligence. Nature language process is the combination of machine learning and computational linguistics in the artificial intelligence field. It is interdisciplinary science among cognitive science, computer science, linguistics, and could be well applied in management practice in the social organisation. It is the further development of old expert systems. With the help of nature language process (NLP) and progress of quality management, enterprises certainly will make great achievement at business intelligence decision and knowledge management.

The objective of this paper is to present NLP, a technological mean in nuclear quality assurance management. Chapter 2 introduces state of the art of Artificial Intelligence tool applied in quality management. Chapter 3 briefly reviews NLP in general. Chapter 4 describes the differences of nuclear quality assurance between the nuclear industries and normal machinery manufacturing industries, which texts are widely used in engineering, and exemplifies typical nuclear QA task. It also introduces the application of NLP in nuclear QA. Chapter 5 explains the steps of using NLP technology by constructing a nuclear NLP system.

Chapter 6 presents common NLP models and algorithms. Chapter 7 details the NLP experiment of the China Nuclear Power Plant Project and contains the result discussion. The conclusions are stated in Chapter 8.

2. State of the Art

In the year 2006, International Atomic Energy Agency (IAEA) published A System for the Feedback of Experience from Events in Nuclear Installations (Safety Guide NS-G-2.11) which proposes to utilise databases and analysis technologies on assessing the reliability of NPPs.

In 2009, some researchers (Seok-Won Hwang et al., 2010) gleaned real-time data associated with the ERP system [3] of NPPs. They proposed web-based algorithm model to compare with baseline data and analyses historical failure trend in the maintenance of Korean nuclear industry.

Other researchers (Pedro V. Marcal et al., 2009) acquires data from nuclear piping failure databases and use them to perform probability risk analysis. The main tool is the automatic natural language abstracting and processing (ANLAP)[4] tool (English). But the main application domain is technologies management of nuclear piping failure.