Mathematical Algorithms in Natural Language Processing

Divya Paikaray, Assistant Professor

Department of Engineering & IT, Arka Jain University, Jamshedpur, Jharkhand, India

Email Id-divya.p@arkajainuniversity.ac.in

ABSTRACT: ***So far, many systems have been suggested to integrate machines in the solving of mathematical problems in a variety of disciplines, including algebra, geometry, physics, and mechanics. This letter provides an up-to-date technical overview of these systems and methods offered by a number of research organizations. The differences between the systems are highlighted, and a unifying design is found that has been used in the majority of these methods. The main achievements of each approach are emphasized. The main advantages and drawbacks of the methods are also addressed. This article addresses the difficulty of understanding mathematics expressed in plain language. This paper examines the role of ad mathematics in the algorithmic workings of natural language. So far, many methods for solving mathematical problems in a variety of fields, including as algebra, geometry, physics, and mechanics, have been described. This communication includes state-of-the-art technical evaluations on various systems and methods from different research organizations.***

KEYWORDS: ***Algorithm, Automated Reasoning, Knowledge Engineering, Mathematical Problems, Natural Language Processing.***

1. INTRODUCTION

Natural Language Processing (NLP) is a field of study and application that studies how computers can utilise text or speech for natural language purposes to comprehend and modify. NLP researchers are working to gain understanding of how people interpret and use language in such a way as to build appropriate tools and strategies to enable computer systems to comprehend and manipulate natural languages in order to do desired tasks. NLP's underpinnings include computer sciences, information science, languages, mathematics, electronics and engineering, artificial and robotic intelligence and psychology[1].

NLP applications cover a range of disciplines, including machine translation, the processing and summary of natural language texts, user interfaces, multilingual and multi-language retrieval of information (MRRIs), speech recognition, artificial understanding and expert systems. Several academics have highlighted the necessity for proper research in order to make it easier to collect multi-language or cross-language information, including multi-lingual text processing and multi-lingual user interface systems[2].

Artificial intelligence (AI) systems, which could interpret, evaluate, resolve or answer questions, situations, facts or issues in the natural language environment from the start of the 1960s, or perhaps before, had been a unique field in computers research. The ultimate aim is to study the capacity of a computer to serve as a substitute for human intellect and reasoning with the support of many subject expertise. However, there was minimal progress in automating the resolution of mathematical problems in a natural language setting, in comparison to the overall research efforts in a higher direction, to imitate the human problem-solving technique[3].

With the word 'mathematical issue,' we imply mostly problems involving some type of scientific theory/regulation or theory or logic, and mathematics is essential for solving them—be it a mechanics problem, physics issue, geometry, or algebra. It is not easy to develop intelligent systems that grasp and resolve mathematical issues in the natural language. It focuses on a distinct challenge from one managed during the development of a menu or control-driven programme for solving computational mathematical problems. These standard software handle user input using a pre-designed structured input method, whereas the smart systems should take into account the natural language description of a problem which various users can identify in diverse ways[4].

To grasp the problem of all lexical complications, the natural language system should employ AI, interpret it in terms of recognised ideas, and then implement the suitable solution logic. On the other hand, software packages frequently follow a mechanical processing pattern, such that no language understanding module or expertise of the domain can address any problem. The development of these intelligent systems is also tough by mapping an optimum range of rules from a broad range of existing rules or technologies to address all sorts of difficulties related to every scientific field. This study gives an extensive review of several issue solvers from early systems created in natural language[5].

Out of the scope of this study are systems that deal with the natural language other than English. An example of issues may be found in the techniques used by the different solver systems. The evaluation of relative qualities and the overall application of techniques has been attempted. The primary aim of this study is to identify certain common fundamental architectures that are in use in systems and to track the helpful paradigms developed in the field of research. A new level of complexity can lead to the generalisation from the current small experiments systems to domain-specific comprehensive issue solvers, and some current methodologies or inventions need to be changed and completely alternative ways developed. The full study contained in this paper can give a brief representation of past research and can serve to guide research towards the development of mathematical problem solvers in natural languages[6].

1.1 Understanding NLP:

The key issue of natural language understanding lies at the heart of any NLP work. There are three main issues in the process of constructing computer programmes, which comprehend natural language: firstly, thinking processes, secondly, linguistic inputs, and thirdly, global information. The NLP system can therefore start at the word level by identifying the morphological structure and nature of the word (for example, part-speech or sense), followed by the sentence level in order to determine the order, the grammar and the significance of the words in the whole sentence, then the context and the overall environment or domain. A given word or sentence may have a specific meaning or connotation in a given context or domain, and may be related to many other words and/or sentences in the given context[7].

Liddy and Feldman suggest that in order to understand natural languages, it is important to be able to distinguish among the following seven interdependent levels that people use to extract meaning from text or spoken languages[8], [9]:

- Level that deals with phonetic or phonological pronunciation.
- Level morphological that addresses smallest portions of meaningful words and suffixes and prefixes.
- Lexical level, which covers words and speech-analysis, pieces having lexical significance.
- Level of syntax addressing grammar and sentence structure.
- Semantic level dealing with word and phrase significance.
- Level of discourse that addresses structures of various text hands with document structures.
- Level of pragmatism that addresses knowledge from outside the world, i.e. from external material.
- All or any of these layers of analysis can include a natural language processing system.

No sentence is an essential NLP method used to collect information. One of the main areas of research is the combination of classical keywords and syntactic techniques with semantical techniques to text processing to enhance information retrieval quality. To analyse the possibilities for Tolle and Chen to isolate noun phrases from medical journal databases, Tolle and Chen (2000) analysed four noun phrase production methods. The category of each case is determined by a text grammar represented as a semantic network. No sentence is an essential NLP method used to collect information. One of the main areas of research is the combination of classical keywords and syntactic techniques with semantical techniques to text processing to enhance information retrieval quality. To analyse the possibilities for Tolle and Chen to isolate noun phrases from medical journal databases, Tolle and Chen (2000) analysed four noun phrase production methods. The category of each case is determined by a text grammar represented as a semantic network[10].

1.2 Abstracting:

The interest recently shown in automated abstraction and text synthesising is reflected in the huge research output presented by ACL, the American Association for Artificial Intelligence (AAAI) and the ACM SIGIR at a variety of international, national and regional conferences and workshops. For automated abstracting and resuming text, several strategies are utilised.

For the extraction and classification of news article summaries, Goldstein, Kantrowitz and Mittal and Carbónell employ traditional IR approaches and language indicators. Recent reports have been made on a variety of studies on text summarisation. Roux and Ledoray have reported on a project, Aristotle, which attempts to create an automated medical data system that produces a semantic canonical text representation.

Moens and Uyttendaele are describing the project SALOMON (Summary and Legal Text Analyses for the Management of Online Needs). In the complete text, such as the name of the court issuing the judgement, the date of the decision, crimes charge, legal provisions disclosed by the court, as well as the legal principles appliqued in this case, the system extracts significant information.

1.3 Information Extraction:

Morin says that while many IE systems can successfully draw words from texts, it is still difficult to disclose interconnections between words. PROMETHEE is a method that collects lexico-syntactic patterns from technical companies in reference to a particular conceptual relationship. Glasgow, Mandell, Binney, Ghemri, and Fisher described the Metlife Intelligent Text Analyzer (MITA), a system that derives information from life insurance applications. They use back propagation (BPI learning technique for training the event detector and use NLP technology to help identify names as the characteristics of documents. These substances are maintained as a knowledge base in an ontology and are utilised to extract information from e-mail communications.

Cowie and Lehnert evaluated IE research and found that the NLP research community was not equipped for addressing the challenges of semantinational feature scoring, co-referencing, and discourse analysis. From its origins into the artificial intelligence field in the 1960s and 1970s, also have evaluated the IE research. They talked about important IE projects carried out in many industries such as research, employment, fault diagnostics, finance, law, medical, military intelligence, police, the specifications and technology/product tracking. software system's needs.

The research Chowdhury examined the usage of model mining approaches in many contexts: the removal of proper names from completeness documents, the extraction of data from news releases, the abstraction of scientific articles. It also explored the usage of templates for information retrieval by several web search engines and suggested that each web author complete a template to characterise hash document, in order to regularise the production of document surrogates.

Smeaton believes that IE is the topic of a large number of research and development projects and has provided answers for many decades. He asks the IE community to understand how fundamental NLP technology has been used for the creation of a task-perhaps the most comparable goal, information recovery. Cost is an important obstacle to the development of IE systems. For example, CONSTRUE took 9.5 years of work (Hayes & Weinstein, 1991). Portability and scalability for IE systems, which rely significantly on domain knowledge, are also two key difficulties.

In extracting information from Web pages, the virtues of NLP and wrapper technologies. Unlike NLP, wrapper induction works regardless of particular domain knowledge. Wrapper technique discovers appropriate information based on the linguistic properties that surround the requested data, instead of analysing the meaning of the speech at phrase level. On the surface, wrappers operate aspects, which characterise text exemplary training.

1.4 Information Retrieval:

Feldman argued that NLP approaches, in conjunction with other technologies like as visualisation, smart agents, and voice recognition should be employed to achieve success in IR[5]. Pirkola has shown that the morphological features of languages differ considerably. However, there are two variables for each language that characterise the complexity of morphological elements: a synthesis index (IS) which defines the quantity of affixation in a given language, i.e. the average quantity of morphemes per word in the language, and a fusion index (IF). Pirkola observed that language computation of ISs and IFs is a very straightforward operation, and once established, these may be used in empirical IR research and system development successfully. Changes in the subject matter presentation have a major impact on IR, and hence the linguistic variety of textual documents is one of the major issues in this area.

Khoo, Myaeng and Oddy examined if, in comparison with keywords alone, information derived through matching the relation of cause-effect, given in documentation, improved outcomes in the retrieval of documents without addressing the relationships. Their experimentation in the complete text database of the Wall Street Journal showed that looking for the reason or effect as wildcard might enhance the efficiency of information recovery, as a result of the corresponding weight. The authors emphasised, however, that the results of this study were not as great as they anticipated.

Chandrasekar and Srinivas argued that consistent text has considerable latent information, such as grammatical structure and language usage patterns, and that the information may be exploited to improve information recovery systems performance. The main objective of this study is to show that sophisticated

NLP strategies for indexing and looking for text documents are superior than basic keywords and string-based approaches for full text statistical recovery.

Rabin advocated the use of random irreducible polynomic-based hash algorithms to construct very long strings of short fingerprint representations. The Hash function had the wonderful feature that the fingerprint of two identical strings had the same fingerprints, while distinct strings had extremely tiny chance of colliding different fingerprints. First LSH introduced by Broder. He suggested using mini-independent functions to build fingerprints that maintain the Jaccard similarity between all vectors. For instance, these approaches are being employed to remove duplicate web pages.

2. DISCUSSION

Even if a real attempt is made to trace the major research work in the field of natural language machine comprehension, even in this very concentrated area of AI we have not been able to grasp the vast array of research difficulties. Contemporary systems which we were unable to include might differ from those discussed above in terms of styles and low levels of functioning, but we have been able to highlight some methods developed over the years:

- Language processing and problem text analysis.
- Information storage and access, rules and domain knowledge.
- Machine comprehension and intermediate representation of the problem inside the domain.
- Solution strategy implementation.

While the previous systems had to apply routines of language processing, mostly through format matching approaches, simultaneous innovations in the field of NLP were benefiting from the later systems. In many circumstances the approach or the weak search technique to adhoc and heuristic programming is used, it seems to be very beneficial in typical circumstances, but it reduces a scope to properly comprehend and to represent different situations. In general, the examination of these systems offers us a decent sense of the computer modelling challenge of the comprehension and resolution of mathematical issues by human intellect, knowledge and thinking skills.

Finally, it is evident that more systematic study is needed in numerous fields of science and technology to build natural language problem solvers. The importance of the research is not only because computer-aided problem solving is novel but also that it may enrich alliance-related study fields such as machine intelligence and machine education, NLPs and KR as well as computer-based instruction (CBT) and cognitive science.

3. CONCLUSION

Some NLP experiments showed good results in this chapter. One must nonetheless not forget that the majority of these experimental systems remain in the laboratory; very few real systems or products are transformed to experimental systems. The unavailability of big test sets and reusable experimental methodologies and tools was one of the main obstacles to NLP research, as in disciplines such as data retrieval research. Recent assessment studies have been positive.

Some NLP experiments showed good results in this chapter. One must nonetheless not forget that the majority of these experimental systems remain in the laboratory; very few real systems or products are transformed to experimental systems. The unavailability of big test sets and reusable experimental methodologies and tools was one of the main obstacles to NLP research, as in disciplines such as data retrieval research. Recent assessment studies have been positive 3. The usage of computer-friendly web-based encoding of natural language documents is symbolic of the revolutionary trend towards human knowledge computerisation.

Although machine translations are not always flawless and do not generate as excellent as human translations, the outcomes and evidence of concern for enhancing MT systems performance are quite encouraging. The natural language question answering system has been one area of application that has attracted a lot of research attention but where the outcomes have yet to deliver an acceptable degree of efficiency for the public. Natural language text processing research has major hurdles in scalability and portability. For large-scale NLP applications, sophisticated NLP approaches such as idea extraction are prohibitively costly. However, there are constant attempts by the research community. There may be less to do with the inefficiencies of the systems or researchers than the intricacies and idiocy of human conduct, or the pattern of communication, since NLP-systems do not function reliably with a high level of performance with great sophistication.

The tremendous quantity of knowledge available on the web has recently being used by NLP researchers. Many applications ranging from word meaning ambiguation, question answering, to mining semantic resources have already benefitted from searching on IR engines for basic surface patterns. However, most techniques for language analysis can't work on the web. One instance is that the lists of similitudes are generated using co-occurrence statistics, which have a quadratic time to run on the input. We overcome this challenge in this research by giving a random algorithm which linearizes this work and reduces memory needs.

REFERENCES

[1] C. L. Borgman, "Multi-media, multi-cultural, and multi-lingual digital libraries: Or how do we exchange data in 400 languages?," *D-Lib Magazine*. 1997.

[2] G. G. Chowdhury, "Template mining for information extraction from digital documents," *Libr. Trends*, 1999.

[3] D. G. Bobrow, "Natural Language Input for a Computer Problem Solving System," 1964.

[4] E. D. Liddy, T. Diamond, and M. McKenna, "DR-link in tipster III," *Inf. Retr. Boston.*, 2000, doi: 10.1023/A:1009986331526.

[5] S. Feldman, "NLP meets the jabberwocky natural language processing in information retrieval," *Online (Wilton, Connect.*, 1999.

[6] M. X. Goemans and D. P. Williamson, "Improved Approximation Algorithms for Maximum Cut and Satisflability Problems Using Semidefinite Programming," *J. ACM*, 1995, doi: 10.1145/227683.227684.

[7] B. DasGupta, G. A. Enciso, E. Sontag, and Y. Zhang, "Algorithmic and complexity results for decompositions of biological networks into monotone subsystems," *BioSystems*, 2007, doi: 10.1016/j.biosystems.2006.08.001.

[8] M. Roux and V. Ledoray, "Understanding of medico-technical reports," *Artif. Intell. Med.*, 2000, doi: 10.1016/S0933-3657(99)00034-2.

[9] M. S. Charikar, "Similarity estimation techniques from rounding algorithms," in *Conference Proceedings of the Annual ACM Symposium on Theory of Computing*, 2002, doi: 10.1145/509907.509965.

[10] J. Goldstein, M. Kantrowitz, V. Mittal, and J. Carbonell, "Summarizing text documents: Sentence selection and evaluation metrics," in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1999*, 1999, doi: 10.1145/312624.312665.