Data & Predictive Analytics – Application in Natural Language Processing

TVVV Prasad[1], Raghu B Korrapati[2]

Rayalaseema University, Kurnool – 518 002, A.P., India

Abstract: *Today's advancements in digital and Artificial Intelligence technologies have paved a new way for the study of natural language based applications using techniques such as predictive analytics. For example, in business applications one can predict the behavior of a customer, product performance, product sales, predictions in compliance and regulatory aspects given the past business trends and sentiment analysis using natural language processing. Now, Linguists can use similar applications to analyze the behavior of a poet, their poetic style, grammar, vocabulary, syntax and semantics etc. This paper is a comprehensive study of those applications in various industry segments. This study helps to appreciate such examples so that several business segments can leverage these techniques.*

Keywords: Data, Decision making, Artificial intelligence, predictive modeling, data science, natural language processing, syntactic and Symantec analysis, morphological techniques

1. Introduction

Data is a set of values that can be measured, reported, and analyzed. The data trends can be drawn using graphs or images. In the digital world data is both structured and unstructured that is drawn from business applications, social sites, natural language applications etc. Today's technology advancements paved the way for usage of artificial intelligence techniques to get meaningful insights from various data sources, thereby attributing cognitive behavior with data synthesis that is like human analysis. This coupled with predictive analytics, the branch of the advanced analytics, which is used to make predictions about unknown future events, using techniques such as data mining, statistics, modeling, machine learning, natural language processing etc., has benefited not only the business industry but also literature and language science. There by it has become one of the most important elements in digital enterprise along with mobile, social, analytics, cloud and artificial intelligence. Few such examples are usage of Natural language processing (NLP) applications in predicting compliance and regulatory needs or predicting product sales or predicting the behavior of a poet / poetic style etc. NLP is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction. One such example is " Predictive Modeling System for Telugu Language Poetry"

2. Predictive Modeling System for 'Telugu' Language Poetry

Intelligent predictive modeling system is used to find grammar ('Chandassu') of a given poem and enables right prediction of words, while building a new poem in Telugu literature. The key steps in this algorithm include – Parsing, building a lexicon, syntactic analysis, with the help of predefined rule base, determine ' chandassu (grammar) of a given poem. It also, builds a lexicon of all the words derived from a poem. The prediction techniques include leveraging statistical techniques to help the poet in finding an

appropriate word, while composing a new poem. An intelligent hashing function is used for faster searching. These techniques and algorithms will enable linguists to analyze ancient texts, poetic patterns, language study, civilization and culture.

3. Background to the Problem

The development of Language technology and its growth leads to the need for the detailed study of computational aspect of language and especially for those who are specialists in the field of technology. Computational Linguistics is an interdisciplinary of linguistics. The very term infers that it is a bridge for language and technology while these two fields in the earlier stage were defined as different areas of study. In the present work, a study is made on how the theory of Natural Language Processing, predictive analytics and morphological techniques [2] can be applied to a specific field of Telugu literature with a detailed example.
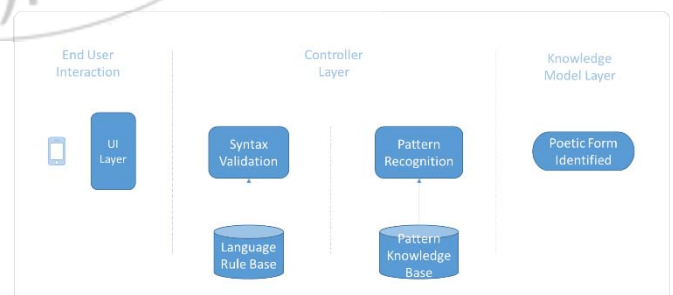
4. Design Approach

Figure 1: Overall Architecture

In the initial phase, we determine the 'Chandassu (grammar)' for a given input (i.e. given Telugu poem). We will parse the inbound data feed, calculate the complexity of the word and store it in the Lexicon. A letter code is assigned based on complexity - simple with 'S', medium with 'M', and complex with 'C'. And the syntactical analysis will be conducted simultaneously for the same feed based on the

predefined set of rules, whereas the syntax for the parsed string can be generated from the rules database. The rules database has been designed by using the set of rules based out of 'Telugu' grammar. The notation is similar to the one used to generate the grammar (i.e. chandassu) for a given poem in Telugu literature.

Once the rules database is formed, we need to update the generated syntax into the words database formed using the lexicon. The next step would be to generate the Unicode for the grammar. Here in this step if an error occurs in the syntax, then the error will be sent to the error log.

5. System Outputs

The following snapshots interpret the various steps of the process involved in this application.
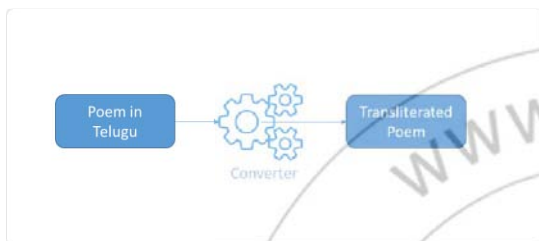

Figure 1: Language Transliteration
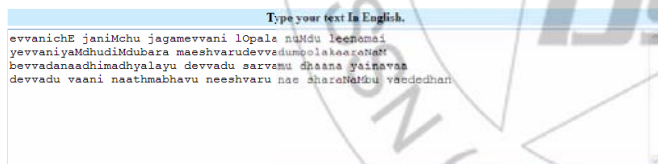

Figure 1: Poem in Native Language ('Telugu') [4]

Figure 2: Transliterated form of the Poem

Figure 3: Poetic Pattern

Figure 4: Poetic Analysis – Grammar Display

The system is developed and deployed in such a way that it's universally available and accessible.

6. System Implementation

The system is implemented using Natural Language Processing techniques. A knowledge base is created and is dynamically updated based on the input database. A heuristic

searching and sorting technique is applied using hashing technique to implement system performance.

This system can run on industry standard cloud based platforms like Amazon cloud (AWS) etc. The system flow chart is depicted in Fig. 5 for reference [5].
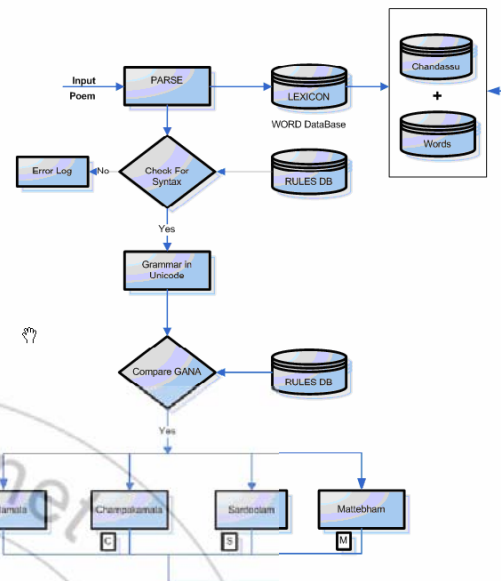

Figure 5: System Process Flow - Poetic Analysis

7. Conclusion and Summary

This paper describes the data and predictive modeling techniques for an application in natural language processing. An example along with sample screen shots is presented for reference. Further study includes algorithm analysis in terms of complexity and system performance aspects, applicability to other forms of Telugu poetry like prose, sonnets etc., and applicability to other languages in the World such as 'Sanskrit', to analyze morphological and linguistic aspects. Similar concepts can be extended to business applications such as "Compliance and Regulatory" needs of an enterprise, legal processing systems, product behavior, consumer behavior, customer sentiment analysis, employee performance analysis, behavior analytics etc., These ideas have shown business benefits in financial services, insurance, retail, utilities, transportation industries, linguistic research etc.

References

[1] Krishna, N. S., & Murthy, H. A. (2004, October). A new prosodic phrasing model for indian language telugu. In *INTERSPEECH*.

[2] Madhavi Ganapathiraju and Lori Levin, 2006, TelMore: Morphological Generator for Telugu Nouns and Verbs. Proc. Second International Conference on Universal Digital Library, Vol Alexandria, Egypt, Nov 17-19, 2006.

[3] S. Rajendran, Ph.D. August 2006, PARSING IN TAMIL –PRESENT STATE OF ART –LANGUAGE IN INDIA, Strength for Today and Bright Hope for Tomorrow, Volume 6 : 8

[4] 'Andhra Mahabhagavatamu' – Bammera Pothana, Potti Sreeramulu Telugu University Press.

[5] Raghu Korrapati, TVVV Prasad, Rayalaseema University, A Natural Language Application Todetermine 'Chandassu' (Grammar) & Predict The Right Word In Telugu Poetry – Summer Internet Proceedings, Jun2010, Vol. 12, p70 June 2010, Allied Publishers

Author Profile

TVVV Prasad has received the M.Tech. degree in Computer Science and Technology from Andhra University in Visakhapatnam. He is a graduate of Executive Leadership Program at Ross Business School, University of Michigan, USA. He has worked for Multinational Companies such as Tata Consultancy Services, Motorola and has more than 25 years of experience in implementing computer based solutions. He is now a Research Scholar at Rayalaseema University. He has participated several international conferences in knowledge based computing and business excellence.