Language Relationship Model for Automatic Generation of Tamil Stories from Hints

Rajeswari Sridhar, Anna University, Department of Computer Science and Engineering, Tamil Nadu, India
V. Janani, Anna University, Tamil Nadu, India
Rasiga Gowrisankar, Anna University, College of Engineering, Guindy, Tamil Nadu, India
G. Monica, Anna University, Tamil Nadu, India

ABSTRACT

In this paper, we propose to develop a Story Generator from hints using a machine learning approach. During the learning phase, the system is fed with stories which are POS tagged and are converted into a Language Relationship model that is represented as a conceptual graph. During the synthesis phase, the input hints which are delimited using hyphen and converted to a conceptual graph. This graph is matched with the conceptual graph of the corpus and probable words, its sequences along with the relationship are determined using three proposed methods namely Randomized selection, Weighted Selection using Bigram Probability of hint phrases and Weighted Selection using product of Bigram Probability of Conceptual Graph and Bigram Probability of hint phrases. Using the words, sequences and relationships, a sentence assembler algorithm is designed to position the words to form a sentence. To make the story complete and readable, suffixes are added using Tamil grammar to the assembled words and a story is generated which is syntactically and semantically correct.

KEYWORDS

Conceptual Graph, Morphological Generator, Tamil Story, Weighted Selection Bigram Probability

INTRODUCTION

Natural Language Processing and Generation is a field of Computer Science which deals with the processing and generation of human languages by a computer that is done either automatically or semi-automatically (Chowdhury, 2003). Natural Language Processing and Generation is considered to be a sub-field of Artificial Intelligence (Downey & Charles, 2015). A computer is made to understand, interpret and generate text in natural languages similar to how humans do, using techniques like Machine Learning and Pattern Recognition. Machine Learning is related to the design of algorithms which lets the computer to learn from data automatically (Grace & Williams, 2016). The learning can be supervised (system is trained with a training set), unsupervised (system operates without prior knowledge of input or output) or reinforcement (system learns based on feedback given) (Downey & Charles, 2015; Grace & Williams, 2016). Natural language processing finds applications in the areas of games, spam filtering of emails, recommendations and other variety of fields (Downey & Charles, 2015; Bourara, Hamou & Amine, 2015; Bouraga, Jureta, Faulkner & Herssens, 2014). Our system uses the techniques of Natural Language Processing and Generation to process and generate meaningful sentences from hints and present that in a comprehensible form to the users. We have developed this story generation system from hints in Tamil language. The fundamental aim of this work is to teach the machine intelligence and making it behave like humans. In a small family society,
the system when integrated could engage children by narrating stories when integrated with the text-to-speech system.

The aim is to build a Story Generator from hints in Tamil language which takes hint phrases as input and generates a syntactically and semantically correct story, utilizing the data corpus previously fed to the system. The system should represent the input hints and the story corpus as conceptual graph and match them to determine the most probable words and use them to generate a syntactically and semantically correct story as output.

This system is the first Story Generator from Hints in Tamil language. The usage of Conceptual Graph to represent the story corpus in intermediate form makes a contribution to the applications of Conceptual Graph. The readability of the generated output shows the efficiency of the Morphological Generator that we have developed. Though Tamil is a free-word order language, we have made an attempt to capture all the sentence patterns possible in Tamil language. We have designed our Conceptual graph algorithm in such a way that we retain all possible relevant information (including the adjectives and adverbs present in the corpus) and we put them to good use while developing the stories when new hint phrases are given as input. We have also developed rules and conditions specific to Tamil language such that all parts of speech such as adjectives, adverbs, etc are automatically learnt from the story corpus and not tagged using a look-up. Pronoun resolution has been implemented to the extent of identifying and resolving a pronoun by backtracking technique and with the help of suffixes.

This paper is organized as follows: Section 2 discusses the sources from which the ideas for different parts of the project were derived. The advantages and disadvantages of some of the related work are also discussed. Section 3 explains the overall system architecture and the design of various modules along with their complexity. Section 4 gives the input and output details of each module, Section 5 elaborates on the results of the implemented system and gives an idea of its efficiency. It also contains information about the dataset used for testing and other the observations made during testing. Section 6 concludes the paper and gives an overview of its criticisms. It also states the various extensions that can be made to the system to make it function more effectively.

**RELATED WORK**

Developing stories based on corpus has already been done in English. But these systems generally take a topic sentence or the characters and a set of pre-defined actions as the input. Those papers have been researched upon to serve as the base idea of the project. In addition, additional concepts that are used in this project have been derived from various other projects. This Chapter discusses the literature survey that has been carried out to derive idea for our Story Generation from Hints project.

**Conceptual Graphs**

A Conceptual Graph (CG) is a graph representation for logic based on the semantic networks of artificial intelligence and the existential graphs (Sowa, 2008). A Conceptual Graph is a typed or sorted version of logic. Each concept has a type or label which denotes the type of entity the concept refers to. A comparison between different graphical representations like dependency graph, existential graph, relational graph and conceptual graph has been made by John Sowa (Sowa, 2008). Conceptual Graph was first developed to serve as an intermediate language for mapping natural language questions and assertions to relational database. In our system, we have used conceptual graphs to represent the conceptual relations in the story corpus. In our conceptual graph, nouns form the nodes while the verbs are the connecting edges. We have also represented cause-effect relationships in the CG to ensure the coherence of adjacent lines of the generated story and also to expand the hints into a complete and meaningful story.
Related Content

Towards Stable Model Bases for Causal Strategic Decision Support Systems
www.igi-global.com/article/towards-stable-model-bases-causal/2424?camid=4v1a

Semiotic Landscapes in Commercial Communication: A Preliminary Reading of Greek-Cypriot Shop Signs
www.igi-global.com/article/semiotic-landscapes-in-commercial-communication/185498?camid=4v1a
Epidemic Estimation over Social Networks using Large Scale Biosensors
www.igi-global.com/chapter/epidemic-estimation-over-social-networks-using-large-scale-biosensors/140458?camid=4v1a

Neutrosophic Soft $\Gamma$-Semiring and Its Ideals
www.igi-global.com/article/neutrosophic-soft-gamma-semiring-and-its-ideals/201561?camid=4v1a