Visualization of Sentences for Children with Intellectual Challenges

Rugma R[1], Sreeram S[2]

[1]M.Tech Student, Dept. of Computer Science & Engineering, MEA Engineering College, Perinthalmanna
[2]Associate professor & HOD, Dept. of Computer Science & Engineering, MEA Engineering College, Perinthalmanna

---***---

Abstract - *Children with intellectual challenges face difficulties in thinking, communication and socialization. However, advances in computer technologies have facilitated their learning and social activities. Computer-assisted teaching approaches provide a flexible learning platform for those children. Delay in language acquisition is one of the major problems faced by those children and it is one of the main reasons for their lack of academic success. Visualizing the verbal content present in their learning materials will improve language comprehension. This paper proposes a simple text-to-scene conversion system for assisting the education of intellectually challenged children.*

Key Words: Natural Language Processing (NLP), Computer Assisted Language Learning (CALL), Text-to-Scene Conversion (TTS).

1. INTRODUCTION

There exist a number of assistive technologies for supporting the needs of intellectually challenged children. With recent advances in technologies, there has been a strong interest in the use of computer-assisted teaching approaches in the education of such children, since they offer great opportunities for them to have an enhanced and enjoyable learning process. According to studies, the main reason for the lack of their academic success is delayed language development [1], and there is a greater chance that these individuals may better understand what they see than what they hear. Words are abstract and rather difficult for the brain to retain, whereas visuals are more permanent and easily remembered. The use of visual representation makes it easier for the child to understand the abstract ideas present in the sentences. So a tool to convert text into corresponding visual representation will have positive impact on their learning process.

A text-to-scene conversion system consists of following three modules:

- Linguistic Analysis
- Semantic Representation
- Scene Generation

i.e. analysing the whole text, extracting the meaningful elements from it and modelling the scene corresponding to the semantic content.

2. RELATED WORK

Text-to-scene conversion is likely to have a number of important impacts because of the ability of an image to convey information quickly. However, relatively little research has considered the conversion from text to visual representations. Any implementation is however limited because of the semantic ambiguities present in the sentence, data set limitation or the lack of context and world knowledge. This section discusses some of those existing text-to-scene conversion systems.

One of the earliest systems was the SHRDLU (even though it did not have a graphics component), a natural language understanding computer program, developed by Terry Winograd at MIT in 1968-1970. In it, the user carries on a conversation with the computer, moving objects, naming collections and querying the state of a simplified "blocks world", essentially a virtual box filled with different blocks.

NALIG (Natural Language Driven Image Generation) by Adorni G, Manzo M.D and Giunchiglia F (1984) is another early work which was aimed at recreating static 2D scenes. NALIG considered simple phrases in Italian of the type subject, preposition, object etc. One of the major goals of the work was to understand relationships between space and prepositions.

Another early program was the Put system by Clay and Wilhelms (1996), which studied spatial arrangements of existing objects on the basis of an artificial subset of English consisting of expressions of the form Put(X P Y), where X and Y are objects, and P is a spatial preposition.

Later, many other text-to-scene conversion systems have been developed. S2S [2], a system for converting Turkish sentences into representative 3D scenes, allowed intellectually challenged people establish a bridge between linguistic expressions and the concepts these expressions refer to via relevant images. The system used SYNSEM (SYNTAX-SEMANTICS) feature structure representation to store information and generated scene from this feature structure representation.

Another system was AVDT (Automatic Visualization of Descriptive Texts) [3], which stored POSIs (Parts of Spatial

Information) as a directed graph and used this directed graph representation for scene generation.

Carsim system [4] converted written car accident reports in to animated 3D scenes. Information from accident reports was stored as a template structure and the system then animated them.

ScriptViz [5] is another system which allowed users to visualize their screenplays in real time via animated graphics. It made use of a Parameterized Action Representation (PAR) that specifies the steps to carry out for generating animations.

WordsEye [6] is one of the famous text-to-scene conversion systems in the world which is developed by AT&T laboratory, Semantic Light Co.Ltd. It contains a large database of linguistic and world knowledge about objects, parts, and other properties. The text input is represented as a dependency structure, semantic information are extracted from it and scene is modeled with the help of large database.

Another recent work [7] discusses scene modeling using a Conditional Random Field (CRF) formulation where each node corresponds to an object, and the edges to their relations. They generate scenes depicting the sentences' visual meaning by sampling from the CRF.

Most of these existing systems successfully convey the meaning of the natural language input sentence. Efficiency of these systems varies for various factors. For example, the system will be more efficient when it is capable of generating images that are more realistic. However, considering children who are not capable of grasping complicated configurations, abstract scenes are highly effective in simply conveying the semantic information. Here the paper proposes a simpler text-to-scene conversion system, taking intellectually challenged children in to consideration.

3. PROPOSED METHODOLOGY

This section discusses the development of a simple and efficient text-to-scene conversion system that generates abstract scenes from input sentence. The system can be divided into following three modules.

- Linguistic Analysis
- Semantic Representation
- Scene Generation

Linguistic analysis involves basic natural language processing steps such as tokenization, lemmatization, tagging etc. In semantic representation step, the input sentence is converted into a dependency structure representation. Then important semantic elements are extracted from this dependency representation. Scene generation module converts these semantic contents into corresponding abstract scenes.

3.1 Linguistic Analysis

Basic natural language processing techniques such as tokenization, lemmatization, Part of Speech (POS) tagging etc. are performed in this step. The system used Stanford CoreNLP library for performing NLP tasks. The Figure 1 shows the linguistic analysis module output corresponding to the example input sentence "A boy is sitting under the tree".

a	:a	:DT
boy	:boy	:NN
is	:be	:VBZ
sitting	:sit	:VBG
under	:under	:IN
the	:the	:DT
tree	:tree	:NN

Fig -1: Linguistic Analysis

When the input text is entered, the linguistic analyzer first converts it into tokens, list of the words present in the sentence. The first part shows the tokenization process. These tokens are then converted in to their lemma form. For example 'sitting' is converted to its root form 'sit' and 'is' is converted into 'be'. Each of these tokens is then tagged with its part-of-speech. In figure, 'NN' stands for singular or mass noun, 'VBZ' stands for 3rd person singular present verb and so on. Part-of-speech tagging helps the system to keep visually relevant words such as nouns, verbs etc. Determiners like 'a', 'the' are not important in visual representation, so they can be omitted in further processing.

3.2 Semantic Analysis

After analyzing the whole text, the meaningful elements have to be extracted from the input sentence. Here text is converted into a dependency structure representation, and this dependency structure is then semantically interpreted and semantic representation is generated.

Figure 2 shows the dependency structure for the given example input sentence. 'Sitting' is the main root verb. 'Boy' and 'tree' are the two nouns dependent on the root verb. 'Under' is the preposition dependent on the noun 'tree'. All these semantically important elements can be extracted from this dependency structure. The dependency structure representation is more convenient for semantic analysis.

-> sitting/VBG (root)

-> boy/NN (nsubj)

-> a/DT (det)

-> is/VBZ (aux)

-> tree/NN (nmod:under)

-> under/IN (case)

-> the/DT (det)

Fig -2: Dependency Structure Representation

It is possible to generate dependency structure for large complex sentences. But this paper focuses only on simple sentences, which are easy for intellectually challenged children to understand. So the work is restricted to simple subject-verb-object sentences. If there are any propositions related to position, the system considers them too.

Next step includes the conversion of dependency structure into semantic representation. From the given dependency structure, system extracts meaningful semantic elements ie. root verb, subject, object and preposition if any. Figure 3 shows the semantic representation that the system extracts out from the dependency structure.

In the given example, 'sit' is the main action, 'boy' is the subject performing the action, 'tree' is the object and 'under' is the positional relation. This semantic representation is used for the scene generation process.

3.3 Scene Generation

The semantic elements extracted from the previous step are converted into corresponding visual representation. The scene generation relies on the database which contains a number of images and location information for various relations. If the noun present in the input sentence is a human being, the database provides different poses and facial expressions too.

Action	:sit
Subject	:boy
Object	:tree
Relation	:under

Fig -3: Semantic Representation

4. EXPERIMENTAL RESULTS

Database for the system is created with the help of abstract scene data set provided by [8]. Some of the poses and facial expressions given in the database are shown in figure 4.

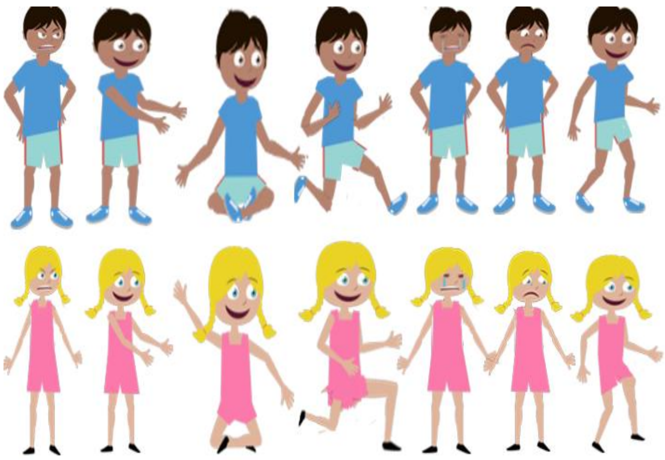

Fig -4: Poses and Facial Expressions

Image corresponding to the subject, object and their actions are retrieved from the database and scene is generated by positioning them according to the location information.

Figure 5 gives the output scene generated for the given sentence "The boy is sitting under the tree". The scene is generated using attractive clipart objects. Other two example output scenes are depicted in figure 6 and 7.

Fig -5: Output for "The boy is sitting under the tree"

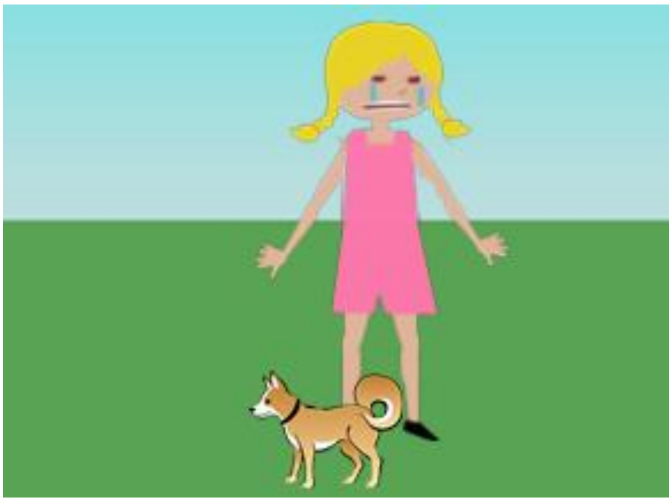

Fig -6: Output for "The girl is crying on seeing the dog"

Fig -7: Output for "Airplane is flying over the clouds"

5. CONCLUSION AND FUTURE WORK

The field of text-to-scene conversion is a very promising area of computer science. It is clear that text-to-scene conversion systems have a number of important impacts because of the ability of a picture to convey information quickly. A text-to-scene conversion system, as an assistive tool for the education of intellectually challenged children will have high social impact. The system can contribute much to the special education field, since visual representation may make it easier for those children to understand the abstract ideas in the verbal expressions.

To the best of our knowledge, S2S is the only system which had implemented the concept of text-to-scene conversion in the field of special education. But the system is restricted to positional relation representation. WordsEye and scene modeling using CRF has many advantages over other existing systems, since they use comparatively high quality models and generate scenes with various object features such as poses, facial expressions etc.

The proposed system also models the scene using various parameters such as facial expressions, poses and positional information. In this work, relatively simple and attractive clip art objects are used for scene generation. Those objects are highly effective in simply conveying the semantic information present in the input sentence for children with intellectual challenges. The dependency structure used in this work is very efficient in semantic analysis. The system now considers only simple sentences with subject-verb- object structure. However it can be modified for complex sentences too, because dependency structure representation is capable of dealing with large complex sentences.

This technique is not restricted to special education domain. It can also be used for other scene generation purposes. A small database of limited set of object and related information is used for the implementation of this work. Defining poses, expressions and location information for each relation was a very challenging task. A large dataset requirement is a limitation of the system. Developing an efficient database is an important area for future research. Learning from a trained set, computing the probability and making the system capable of generating the scene for a new given sentence is another area of future work. The possibility and efficiency of retrieving images from the internet is also has to be studied.

REFERENCES

[1] U. E. Kilicaslan Y, Ucar O and G. E.S., "Visualization of Turkish for autistic and mentally retarded children," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 144-147, Jan. 30-Feb. 1, 2008.

[2] O. U. Yilmaz Kilicaslan and E. S. Guner, "An nlp-based 3d scene generation system for children with autism or mental retardation," Proceedings of the 9th International Conference on Artificial Intelligence and Soft Computing, ICAISC, pp. 929-938,June 2008.

[3] H. D. Christian Spika, Katharina Schwarz and H. P. A. Lensch, "Avdt - automatic visualization of descriptivetexts," Proceedings of the Vision, Modeling, and Visualization Workshop, October 2011.

[4] P. N. Richard Johansson and D. Williams, "Carsim: A system to convert written accident reports into animated 3d scenes," Proceedings of the 2nd Joint SAIS/SSLS Workshop Artificial Intelligence and Learning Systems, AILS-04, pp. 76-86, April 2004.

[5] Z.-Q. Liu and K.-M. Leung, "Script visualization (scriptviz): a smart system that makes writing fun," Soft Computing, vol. 10, pp. 34{40, January 2006.

[6] B. Coyne and R. Sproat, "Wordseye: An automatic text-to-scene conversion system," Proceedings of the 28th annual conference on Computer Graphics and interactive techniques, pp. 487-496, August 2001.

[7] P. D. Zitnick C.L. and V. L, "Learning the visual interpretation of sentences," IEEE International Conference on Computer Vision (ICCV), pp. 1681-1688, December 2013.

[8] C. Lawrence Zitnick and Devi Parikh, "Bringing semantics into focus using visual abstraction," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3009-3016,2013.

BIOGRAPHIES

Rugma R is pursuing her M.Tech Course in Computer Science & Engineering at MEA Engineering College, Perinthalmanna, Kerala, India

Sreeram S is the HOD in the Dept. of Computer Science & Engineering at MEA Engineering College, Perinthalmanna, Kerala, India