ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

Text-to-Scene Conversion System for Assisting the Education of Children with Intellectual Challenges

Rugma R[1], Sreeram S[2]

M.Tech Student, Department of Computer Science &Engineering, MEA Engineering College, Perinthalmanna, Kerala, India[1]

Associate Professor, Department of Computer Science &Engineering, MEA Engineering College, Perinthalmanna, Kerala, India[2]

ABSTRACT:Children with intellectual challenges face serious problems in thinking and communicating with linguistic structures. Software technologies offer great opportunities for such children to communicate and socialize. 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 their language development skills. This paper proposes a simple text-to-scene conversion system that can be used as an assistive tool for the learning process of intellectually challenged children. The system first convert the natural language input sentence in to a dependency structure representation and then extract the meaningful contents from it. The semantic content is then mapped with the image objects and scene corresponding to the sentence is rendered.

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

I. INTRODUCTION

Children with intellectual challenges often have problems in thinking, communication and socialization. With the advent of information and communication technologies (ICT), new hopes are emerging for those children. Today, 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 special education field. Software technologies provide a flexible learning platform for children with intellectual challenges. According to studies, the main reason for the lack of their academic success is delayed language development [1]. For example, an intellectually disabled child hearing the word `cat' may not be able to connect that word to an actual cat that he or she sees. Words are abstract and rather difficult for the brain to retain, whereas visuals are more permanent and easily remembered.So there is a greater chance that these individuals may better understand what they see than what they hear. We have come to accept the saying “a picture is worth a thousand words” as truth in our culture because of the ability of an image to quickly convey so much meaning with so little explanation. 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. This paper discusses the development of a simple text-to-scene conversion system for assisting the education of intellectually challenged children.

Two major steps involved in the conversion of natural language sentence into corresponding visual representation are natural language understanding and scene generation. First, basic natural language processing techniques such astokenization, lemmatization, Part of Speech (POS) tagging etc. are performed. Next is the syntactic analysis part, which gives the structural representation of the input. The most important step in natural language processing is the

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

extraction of meaningful elements from the input sentence. Finally these semantic contents are mapped with the database objects and scene corresponding to the sentence is generated.

II. 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.

S2S [2], a system for converting Turkish sentences into representative 3D scenes, allows intellectually challenged people establish a bridge between linguistic expressions and the concepts these expressions refer to via relevant images. The system uses SYNSEM (SYNTAX-SEMANTICS) feature structure representation to store information and generated scene from this feature structure representation. Another system is AVDT (Automatic Visualization of Descriptive Texts) [3], which stores POSIs (Parts of Spatial Information) as a directed graph and uses this directed graph representation for scene generation.Carsim system [4] converts written car accident reports in to animated 3D scenes. Information from accident reports is stored as a template structure and the system then animates them.ScriptViz [5] is another system which allows users to visualize their screenplays in real time via animated graphics. It makes use of a Parameterized Action Representation (PAR) that specifies the steps to carry out for generating animations.Paper entitled 'Preliminary Implementation of Text-to-scene System[6]' proposes sentence pattern library concept- a small database to store the frequently used sentence patterns and grammar. The input sentence can be matched to the database sentence patterns and grammar, the computer can make a quick conversion, and this can save much time. There exists another system called Write a Picture [7], an educational program intended to offer a web-based text-to-scene interface which can familiarize its users with vocabulary as well as with spatial relations in a newly-acquired language. Another work [8] discusses about a text to scene generation system which integrate learned lexical groundings with a rule-based scene generation approach. In this paper, they introduce a dataset of 3D scenes annotated with natural language descriptions and learn from this data how to ground textual descriptions to physical objects. Extraction of information about scene layout from text descriptions and the conversion of text into scene are discussed in [9].

WordsEye [10] 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 [11] 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.

III. TEXT-TO-SCENE CONVERSION SYSTEM (TTS)

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

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".

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

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

Fig. 1: Tokenization, lemmatization and POS tagging outputs for the sentence "A boy is sitting under the tree".

Given an input text, tokenization is the task of chopping it up into pieces, called tokens, perhaps at the same time throwing away certain characters, such as punctuation. These tokens are then converted in to their lemma form. The goal of lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma. Each of these tokens is then tagged with its part-of-speech. 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.

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 tree obtained for the sentence "A boy is sitting under the tree"

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 is any proposition 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. the root verb, subject, object and preposition if any.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

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

Fig. 3: Semantic elements extracted from the sentence "A boy is sitting under the tree"

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.

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. Database for the system is created with the help of abstract scene data set provided by [12]. Image corresponding to the subject and object are searched in the database and those with higher probability are retrieved. Scene is generated by positioning the retrieved images according to the location information. Figure 5 gives the output scene generated for the given sentence "The boy is sitting under the tree".

Fig. 5: Output scene generated for the sentence "A boy is sitting under the tree"

IV. EXPERIMENTAL RESULTS

The developed text-to-scene conversion (TTS)system has many advantages over other existing systems. S2S is an existing text-to-scene conversion system with the same objective as our proposed system. It uses feature structure for semantic representation. Extraction of meaningful contents from the text is comparatively efficient in S2S. But the system is restricted to positional relation representation. No actions or verbs can be visualized using this system. These drawbacks are resolved in the developed system. The proposed methodology can visualize various object features such as actions, emotions etc. Among the existing text-to-scene conversion systems, WordsEye and scene modeling using CRF has many advantages, since they use comparatively high quality models and generate scenes with various parameters.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

When considering children with intellectual challenges, the wordseye system's scene generation module is less efficient. But the developed system produces comparatively simple and attractive abstract scenes. The figure 6 gives the output generated by three different systems for the same input sentence “The boy is sitting under the tree”.

Fig. 6: Output generated by three different systems for the sentence “A boy is sitting under the tree”

The scene generation component is efficient in scene modeling using CRF too. But the system experience limitations in semantic representation. The figure 7 provides a comparison between the developed text-to-scene conversion system and two of the best existing technologies. The predicate tuple extraction used by the system sometimes produces incorrect tuples. The dependency structure representation used in TTS is very efficient in semantic analysis. It helps the system to find relations between the words in a given sentence. Extraction of meaningful elements from this dependency structure representation is more efficient and easy. The graph is generated with the help of the results of a human study asking which scenes better convey the meaning present in the sentence. Scene generated by TTS is more attractive and understandable, which are important parameters when considering children with intellectual challenges.

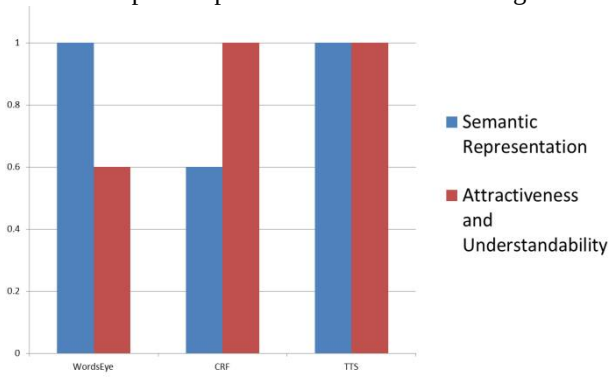

Fig. 7: Performance comparison graph

V. CONCLUSION

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

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

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.

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. YilmazKilicaslan 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 descriptive texts," Proceedings of the Vision, Modeling, andVisualization 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] J. S. Fuping Yang and Z. Huang, “Preliminary implementation of text-to-scene system,” International Conference on Information Sciences, Machinery, Materials and Energy (ICISMME 2015), pp. 1295-1299, June.

[7] J. Roux, “Exploring text-to-scene feedback as an alternative for second language acquisition,” Master's Thesis at Grenoble Institute of Technology, 2013.

[8] W. M. Angel Chang and M. Savva, “Text to 3d scene generation with rich lexical grounding," Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics 32 and the 7th International Joint Conference on Natural Language Processing, p. 53-62, July 26-31 2015.

[9] Y. Z. Fuping Yang and X. Luo, “Scene layout in text-to-scene conversion," 2014 2nd International Conference on Systems and Informatics (ICSAI 2014), pp. 891-895.

[10] 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.

[11] 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.

[12] Z. C.L and P. D, “Bringing semantics into focus using visual abstraction,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3009-3016,2013.