Learning Physiotherapy through Virtual Action

Sandeep Kumar Dash[1], Partha Pakray[1], Alexander Gelbukh[2]

[1] National Institute of Technology Mizoram, Aizawl
India

[2] Instituto Politécnico Nacional, Centro de Investigación en Computación,
Mexico

{sandeep.cse,partha.cse}@nitmz.ac.in, www.gelbukh.com

Abstract. We describe a research framework for virtualizing documented physiotherapy instructions. Our approach bridges the gap between human understanding and the written manuals of instructions for physiotherapy. Techniques of Natural Language Processing involving semantic and spatial information processing are important in this approach. We have also explained the physiotherapy considerations that we employed in this research.

Keywords. Natural language processing, virtual action, physiotherapy.

1 Introduction

Readers tend to visualize a scenario while reading any document or text. Thus converting natural language text into virtual action is of great importance with respect to the complexity faced mainly in understanding of written manuals, simulation of production line for manufacturing units to make the task easier, error-free, and risk minimizing, story description for students as an aid in education and in other similar circumstances. In addition, the use of natural language text to describe a scene is much easier than complex manipulation interfaces through which objects are constructed and then precisely positioned within scenes.

The complexity in this case lies with natural language, in which both linguistic and real-world knowledge, in particular knowledge about the spatial and functional properties of objects, prepositions, and the spatial relations they convey is often ambiguous. Also, verbs and how they resolve to poses and other spatial relations [4] need to be analyzed as well in the input text for a proper scene building. Furthermore, implicit facts about real world have also to be analyzed from the text since they are rarely mentioned therewith.

Physiotherapy is an area where textual descriptions of the exercises are difficult to interpret. The various body parts, joints, and their angle along with the direction of movement have utmost importance while carrying out specific physiotherapies. This is one of the challenging areas, where interpretation of textual content has to be perfect or else consequences can be damaging for health.

In this paper, we introduce a structure for representing text describing different body part movements during physiotherapic exercises. We describe how the input textual description of a physiotherapy exercise can be converted into a virtual action describing it. The method is proposed only for some of the therapies, but we believe that it can be extended for other types of therapies, though further research is necessary for this.

2 Related Work

Research on the relation between images and text can be subdivided into three areas: (1) generating textual description of the given images, (2) extracting information from a combination of images and text, and (3) generating images from texts. The fist type is probably most common: for example, it includes tagging images or actions in the social media with natural text [1–3]. The second area includes, for example, multimodal sentiment

ISSN 1405-5546
doi: 10.13053/CyS-20-3-2450

analysis, a recent topic that aims to understand sentiment in the multimodal content [13, 15, 16]. In this paper, we address the third type of research: generating images from a given text.

There has been a considerable number of works that implemented natural language text as input for generating 3D scene as a virtual scenario. Researchers have utilized various text-processing methods in order to build structures that can be made to generate the scene.

Generally, such systems follow four basic steps. First, they parse the input text into a structure or template establishing spatial relation between the different objects mentioned in it. Secondly, using real world knowledge they identify implicit constraints that are not mentioned in the text. At the third step, the scene templates are converted into geometric 3D scene. Lastly, the system optimizes the placement of objects as per the templates.

This type of works influenced our research in the sense that the spatial information about the objects mentioned in the sentence carries importance, as there can also be different body parts placement during a physiotherapy along with the furniture and objects used of present during the activity, such as chairs, walls, floors, sitting mats, balls, etc.

The pioneering work in this area was carried out by Coyne et al. [4]. They developed an automatic text-to-scene conversion system named WordsEye. Their first system was designed with a large library of 3D objects to depict scenes from the input text. Their current system contains 2,200 3D objects, 10,000 images, and a lexicon of approximately 15,000 nouns. It supports language-based control of objects, spatial relations, and surface properties (e.g., textures and colors). In addition, their system handles simple co-reference resolution, allowing for a variety of ways of referring to objects. They utilized a self-designed lexical knowledge base called SBLR (Scenario-Based Lexical Resource), which consists of an ontology and lexical semantic information extracted from WordNet [5] and FrameNet [6].

Their system works by, first, tagging and parsing the input text, using Church's part-of-speech tagger [7] and Collins's parser [8]. The parser output is then converted into a dependency structure, which is processed to resolve anaphora and other types of co-reference. The dependency structure is utilized to create semantic nodes, from which the semantic relations of lexical items are formed with the help of information in the SBLR. These semantic relations are then converted into a final set of graphical constraints representing the position, orientation, size, color, texture, and poses of objects in the scene. This final structure is then used to create the 3D scene with the help of graphic rendering software.

Chang et al. [9] have developed an approach to finding implicit spatial relationships between objects described in a scene. They considered the output scene to be generated as a graph whose nodes are objects mentioned either implicitly or explicitly in the text and the edges as the semantic relationships among them. The semantics of a scene was described using a "scene template" and the geometric properties were described using a "geometric scene."

In their approach, a scene template is a triplet T = (O, C, Cs), which consists of object descriptions O, constraints C on the relationships between the objects, and a scene type Cs. Here, object description provides information about the object as category label, basic attributes such as color and material, and the number of its occurrences in the scene. Spatial relations between objects were expressed as predicates of the form supported_by(o_i, o_j) or left(o_i, o_j), where o_i and o_j are recognized objects. Geometric Scene represents specific geometric representation of a scene. It consists of a set of 3D model instances—one for each object—that capture the appearance of the objects. To position exactly the object, they also derived a transformation matrix that represents the position, orientation, and scaling of the object in a scene. A scene template was generated by selecting appropriate models from a 3D model database and determining transformations that optimize their layout to satisfy spatial constraints.

Jianqing Mo et al. [10] have created a visual action design environment that improves action design for intelligent toy robot. The environment automatically creates action sequence files in plain text or XML formats from given action sequences. Action sequences are just sets of arrays of actions aggregated as per their time of occurrences. These actions may be parallel actions with complete or in-part synchronous with time, linear action or

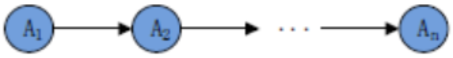


Fig. 1. Linear Action

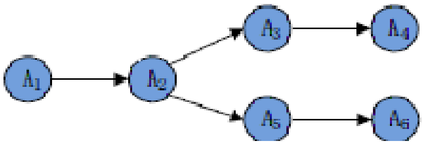
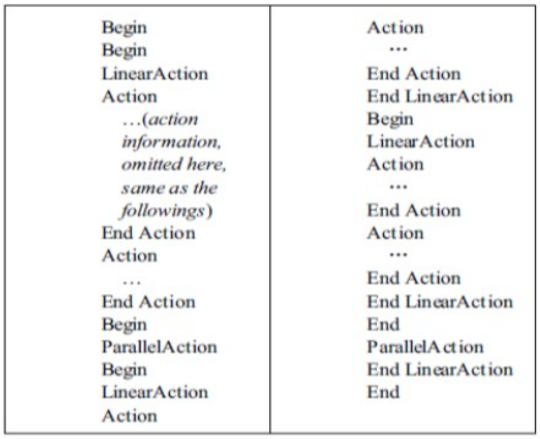


Fig. 2. Parallel Action

repeated serial actions. They are illustrated in Figures 1 and 2.

These action sequences can be created in the visualized design environment and the action sequence file can be automatically generated thereafter. These files contain action sequences in terms of tags and data blocks as shown in Figure 3.

In Figure 3, the tag “Begin LinearAction / End LinearAction” represents linear action, tag “Begin ParallelAction / End ParallelAction” represents parallel action, and the tag “Action / End Action” represents data block that contains information about a particular action. The intelligent toy robots are limited to only rotational freedom. Hence all actions are defined logically as A_i (Joint, Angle, Axis, Absolute, Time, StartTime, IsParallel). The parameters respectively represent joint that exerts action, angle of rotation, axis of rotation, absolute or relative value of the angle, and timeline for adjusting and harmonizing the coherent action sequence, order of size of action, and finally whether the node is an action node or substitute node. Further, they have used the queue data structure to store the action files as per their order of StartTime. Finally, virtual action was generated by matching action sequence files into corresponding joints of intelligent toy robot model. They used Eon studio for virtual simulation purposes.



```
Begin
Begin
LinearAction
Action
  …(action
  information,
  omitted here,
  same as the
  followings)
End Action
Action
  …
End Action
Begin
ParallelAction
Begin
LinearAction
Action
```

```
Action
  …
End Action
End LinearAction
Begin
LinearAction
Action
  …
End Action
Action
  …
End Action
End LinearAction
End
ParallelAction
End LinearAction
End
```

Fig. 3. Action Tags [10]

In another innovative research project funded by EU named MUSE (Machine Understanding for interactive Storytelling) [11], the aim was to bring texts to life by developing an innovative text-to-virtual-world translation system. It mainly evaluated two scenarios: story-telling and patient education materials. The idea involved recognition of semantic roles in a given sentence, spatial relations between objects, and chronological order of events.

3. Our Framework

Our framework for converting descriptive text of physiotherapies into virtual action describing the actions thoroughly proceeds as follows:

– Input: descriptive text for physiotherapic exercise;
– Identification of posture and joints involved in it, type of movement, angle of movement, and the duration of halting in the pose;
– Forming logical representation of the instruction;
– Forming action representation;
– Converting all the action representations sequentially to virtual action of the therapy.

In our framework, the knowledge base incorporates the knowledge about types of posture each therapy generate by involving the joints of the respective parts, the prescribed angle between joints, type of movements and the necessary time duration of halting in the pose as well as the number of times the action is to be repeated. At the current stage, we incorporate in our system the information described in the following subsections.

3.1 Type of Body Planes

For describing the structural positions and directions of functional movement of the body, the standard posture is that of the body facing forward, the hands at the sides of the body, with the palms facing forward, and the feet pointing straight ahead.

Body planes are derived from dimensions in space and are oriented at right angles to one another. The median sagittal plane, also called the *midsagittal* plane, divides the body into right and left halves. The *coronal plane* is vertical and extends from side to side. The transverse plane is a horizontal plane and divides a structure into upper and lower components. We use these three types of planes for representing different postures.

3.2 Axes of Movements

An axis represents a line around which movement is to occur. We consider three types of movements possible around each joint: *rotation*, *translation*, and *curvilinear* motion. All movements that occur around an axis are considered rotational, whereas linear movements along an axis and through a plane are called translational. Curvilinear motion occurs when a translational movement accompanies rotational movements.

3.3 Type of Joints

The three types of joints we consider are mainly based upon how much movement they allow. A *synarthrosis* joint permits no movement, as, for example, the gomphoses that connect the teeth to the skull. An *amphiarthrosis* joint allows a slight amount of movement at the joint, for example, as in pubic symphysis of the hips. The third type is the *diarthrosis* joint. Diarthroses have the highest range of motion of any joint and include the elbow, knee, shoulder, and wrist.

3.4 Type of Movements

Some of the specific movements we consider around joints are as follows:

Flexion: Bending parts at a joint so that the angle between them decreases and the parts come closer together (bending the lower limb at the knee).

Extension: Straightening parts at a joint so that the angle between them increases and the parts move farther apart (straightening the lower limb at the knee).

Hyperextension: Excess extension of the parts at a joint, beyond the anatomical position, such as bending the head back beyond the upright position.

Eversion: Turning the foot so that the sole faces laterally.

Inversion: Turning the foot so that the sole faces medially.

Protraction: Moving a part forward, such as thrusting the chin forward.

Retraction: Moving a part backward, such as pulling the chin backward.

Elevation: Raising a part, such as shrugging the shoulders.

Depression: Lowering a part, such as drooping the shoulders.

4 Dataset

We introduce the LAS (Logical Action Structure) that describes each pose for a particular therapy. It involves information such as which body plane, which axes, what movement, and which joints are involved in the described text. A LAS has the following format:

{b_plane, axes, no_of_joints {$joint_1$, $joint_2$, …., $joint_k$}, m_type, min_and_max_angle_between_individualjoint_i, duration_of_halt_in_pose}

where:

b_plane is body plane;

axes are the axes of movements;

no_of_joints are the numbers of joints mainly involved in the therapy;

m_type is the type of movements;

minmax angle is the appropriate angle of
movement of joints; and
duration_of_halt_in_pose is the time gap between
consecutive actions.

Through efficient text-processing methods along with implicit information incorporation in our database of different therapy poses, each structure can represent its logical action that can lead sequentially to generate the exact virtual action prescribed in the written manuals.

5 Conclusion

Our method of transforming text to virtual action considers only the physiotherapic aspect in which both the semantic and spatial information from text are to be retrieved. In addition, the newly introduced logical structure of exercise description from text is unique of its kind. While initially we limit the number of therapies and its associated movements, the structure is extensible to large number of such items.

Our future work will focus on the use of common sense-based reasoning [12, 17, 19, 24], reflecting emotions present in the text [14, 18] and personality traits [20] for better rendering of the human face. Textual entailment-based techniques [21–23] will be an important part of our reasoning scheme.

Acknowledgements

This research is supported by DST-DAAD Project Based Personnel Exchange Programme: an Indo-German Joint Research Collaboration (No. INT/FRG/DAAD/P-15/2016).

References

1. **Cambria, E., Poria, S., Bisio, F., Bajpai, R., & Chaturvedi, I. (2015).** The CLSA model: a novel framework for concept-level sentiment analysis. *International Conference on Intelligent Text Processing and Computational Linguistics, CICLing,* Lecture Notes in Computer Science, Springer, Vol. 9042, pp. 3–22. DOI: 10.1007/978-3-319-18117-2_1.
2. **Chikersal, P., Poria, S., Cambria, E., Gelbukh, A., & Siong, C.E. (2015).** Modelling public sentiment in Twitter: Using linguistic patterns to enhance supervised learning. *International Conference on Intelligent Text Processing and Computational Linguistics, CICLing*, Lecture Notes in Computer Science, Springer, Vol. 9042, pp. 49–65. DOI: 10.1007/978-3-319-18117-2_4.
3. **Chikersal, P., Poria, S., & Cambria, E., (2015).** SeNTU: Sentiment analysis of tweets by combining a rule-based classifier with supervised learning. *SemEval 2015*, p. 647.
4. **Coyne, B., Sproat, R., & Hirschberg, J. (2010).** Spatial relations in text-to-scene conversion. *Computational Models of Spatial Language Interpretation,* Workshop at Spatial Cognition.
5. **Miller, G. & Fellbaum, C. (1998).** *Wordnet: An electronic lexical database*.
6. **Baker, C., Fillmore, C., & Lowe, J. (1998).** The Berkeley FrameNet Project. *COLING-ACL*.
7. **Church, K.W. (1988).** A stochastic parts program and noun phrase parser for unrestricted text. *Proceedings of the Second Conference on Applied Natural Language Processing,* pp. 136–143. DOI: 10.3115/974235.974260.
8. **Collins, M., Ramshaw, L., Hajič, J., Tillman, C. (1999).** A statistical parser for Czech. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics.* Association for Computational Linguistics, pp. 505–512. DOI: 10.3115/1034678.1034754.
9. **Chang, A.X., Savva, M., & Manning, C.D. (2014).** Learning Spatial Knowledge for Text to 3D Scene Generation. *EMNLP,* pp. 2028–2038.
10. **Mo, J., He, H., & Zhang, H. (2012).** Virtual simulation of intelligent toy robot driven by the action sequence files. *International Conference on Computer Science and Automation Engineering, CSAE*, Vol. 1, IEEE, pp. 401–404. DOI: 10.1109/ CSAE.2012.6272625.
11. **De Mulder, W., Do Thi, N.Q., van den Broek, P., & Moens, M.F. (2013).** Machine understanding for interactive storytelling. *Proceedings of KICSS 2013: 8th international conference on knowledge, information, and creativity support systems, Springer,* pp. 73–80.
12. **Poria, S., Agarwal, B., Gelbukh, A., Hussain, A., & Howard, N. (2014).** Dependency-based semantic parsing for concept-level text analysis. *International Conference on Intelligent Text Processing and Computational Linguistics, CICLing*, Lecture Notes

in Computer Science, Springer, Vol. 8403, pp. 113–127. DOI: 10.1007/978-3-642-54906-9_10.

13. **Poria, S., Cambria, E., & Gelbukh, A. (2015).** Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. *EMNLP,* pp. 2539–2544.
14. **Poria, S., Cambria, E., Gelbukh, A., Bisio, F., & Hussain, A. (2015).** Sentiment data flow analysis by means of dynamic linguistic patterns. *IEEE Computational Intelligence Magazine,* Vol. 10, No. 4, pp. 26–36. DOI: 10.1109/MCI.2015.2471215.
15. **Poria, S., Cambria, E., Howard, N., Huang, G.B., & Hussain, A. (2016).** Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing,* Vol. 174, pp. 50–59. DOI: 10.1016/j.neucom.2015.01.095.
16. **Poria, S., Cambria, E., Hussain, A., & Huang, G.B. (2015).** Towards an intelligent framework for multimodal affective data analysis. *Neural Networks,* Vol. 63, pp. 104–116. DOI: 10.1016/j.neunet.2014.10.005.
17. **Poria, S., Cambria, E., Ku, L.W., Gui, C., & Gelbukh, A. (2014).** A rule-based approach to aspect extraction from product reviews. *Proceedings of the Second workshop on natural language processing for social media, SocialNLP*, pp. 28–37.
18. **Poria, S., Cambria, E., Winterstein, G., & Huang, G.B. (2014).** Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems,* Vol. 69, pp. 45–63. DOI: 10.1016/j.knosys.2014.05.005.
19. **Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., & Bandyopadhyay, S. (2013).** Enhanced SenticNet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems,* Vol. 28, No. 2, pp.31–38. DOI: 10.1109/MIS.2013.4.
20. **Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., & Howard, N. (2013).** Common sense knowledge based personality recognition from text. *Mexican International Conference on Artificial Intelligence, MICAI*, Lecture Notes in Artificial Intelligence, Vol. 8266, Springer, pp. 484–496. DOI: 10.1007/978-3-642-45111-9_46.
21. **Pakray, P., Neogi, S., Bhaskar, P., Poria, S., Bandyopadhyay, S., & Gelbukh, A. (2011).** A textual entailment system using anaphora resolution. *System Report, Text Analysis Conference, Recognizing Textual Entailment Track (TAC RTE) Notebook,* Vol. 2011.
22. **Pakray, P., Pal, S., Poria, S., Bandyopadhyay, S., & Gelbukh, A. (2010).** JU_CSE_TAC: Textual entailment recognition system at TAC RTE-6. *System Report, Text Analysis Conference, Recognizing Textual Entailment Track (TAC RTE) Notebook,* Vol. 2010.
23. **Pakray, P., Poria, S., Bandyopadhyay, S., & Gelbukh, A. (2011).** Semantic textual entailment recognition using UNL. *Polibits,* Vol. 43, pp. 23–27.
24. **Poria, S., Gelbukh, A., Cambria, E., Hussain, A., & Huang, G.B. (2014).** EmoSenticSpace: A novel framework for affective common-sense reasoning. Knowledge-Based Systems, Vol. 69, pp. 108–123. DOI: 10.1016/j.knosys.2014.06.011.

Sandeep Kumar Dash is Assistant Professor and PhD (pursuing) at Dept. of CSE, NIT Mizoram. His area of research is text processing and NLP.

Dr. Partha Pakray received his Ph.D. degree in Computer Science and Engineering from the Jadavpur University, India. He is currently Head and Assistant Professor at the Department of Computer Science and Engineering of the National Institute of Technology Mizoram. He received fellowship from European Research Consortium for Informatics and Mathematics (ERCIM) two times and worked at the Norwegian University of Science and Technology, Norway, and the Masaryk University, Czech Republic, as a postdoctoral fellow. He also worked at the Xerox Research Centre Europe (XRCE) as a research intern. He has published 45 research publications in various areas of natural language processing.

Dr. Alexander Gelbukh received his M.Sc. degree in mathematics from the Lomonossov Moscow State University, Russia, and his Ph.D. degree in computer science from VINITI, Russia. He is currently a Research Professor and Head of the Natural Language Processing Laboratory of the Center for Computing Research (Centro de Investigación en Computación, CIC) of the Instituto Politécnico Nacional (IPN), Mexico. He is a former President of the Mexican Society of Artificial Intelligence (SMIA), a Member of the Mexican Academy of Sciences, and a National Researcher of Mexico (SNI) currently at excellence level 2. He is author or co-author of more than 500 research publications in natural language processing and artificial intelligence.

Article received on 03/05/2016; accepted 20/08/2016.
Corresponding author is Sandeep Kumar Dash.