24th International Conference of the European Federation for Medical Informatics
Quality of Life through Quality of Information – J. Mantas et al. (Eds.)
MIE2012 / CD / Short Communications (Poster)

The Role of Unstructured Data in Gait Analysis

Vladimir ERGOVIC[a1], Stanko TONKOVIC[b], Vladimir MEDVED[c]
[a] *IBM Software Group, Zagreb, Croatia*
[b] *Faculty of Electrical Engineering and Computing/ZESOI, Zagreb, Croatia*
[c] *Faculty of Kinesiology/ZOPK, Zagreb, Croatia*

Abstract. In today's world, most of medical information is still being kept in unstructured format. Over the last few decades, knowledge stored in unstructured free-text format has become precious for clinicians and researchers. We introduce a recommendation engine for gait analysis that describes signal with words by leveraging existing data-mining techniques used in time series. Such words are validated against documents that contain knowledge that comes from clinicians.

Keywords. Text mining, gait analysis, time series

Introduction

Gait report is lengthy and complex. It includes patient's medical history, physical examination, processed data from the gait study and clinical interpretation [1, 2]. We describe a solution that utilizes different kind of data sources in order to allow rapid production of a preliminary gait interpretation and report. Knowledge stored in existing documents can be utilized by means of advanced linguistic analysis performed on gait signal descriptions and possibly other structured data from previous trials in order to provide gait interpretation [3].

1. Method

We built a custom dictionary that contains specific knowledge domain terms for gait analysis. We can apply a descriptive approach to a signal and describe signals using various statements (e.g. stance L pelvic obliquity degrees are below range, swing L pelvic obliquity degrees are below ranges, stance R pelvic obliquity degrees are extremely above, swing R pelvic obliquity degrees are extremely above, etc.). Words from statements that match words in the custom dictionary are indexed, stored in a database and associated to a document or gait report. We defined 5 major categories (Normal, Below, Above, Extremely below, Extremely above) [4, 5]. In order to avoid a lot of irrelevant data, we utilized content-based ranking. Gait reports are being scored and the ones with highest score are returned first, as they are supposed to have higher confidence ratio. Applied scoring metric included: word frequency (the number of

[1] Corresponding Author. Vladimir Ergovic; IBM Software Group, Zagreb, Croatia; Email: vladimir.ergovic@hr.ibm.com

times the words in the query appear in the document helps to make report more relevant), location in the document (the main subject of a report, such as normal and pathological gait, that was the case with our BTS Elite Clinic generated reports) and word distance (multiple words in the query should appear close together in a document). Typically, the first section of a report contains demographic data, date of trials, pathology, followed by temporal and distance parameters and interpretation. The last section contains kinematics and kinetics with associated graphs and optionally colorification charts and EMG [5, 6]. The method used is shown in Figure 1. Gait data is being fed into a signal description process in order to generate human readable description of the signal. Input is merged with patient history and a new report is created. That document is then being scored against patient reports and other documents that are stored in databases.

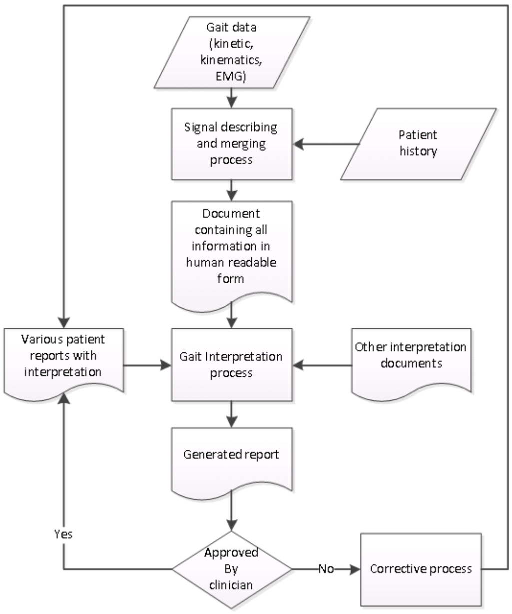

Figure 1. Process of text-mining gait reports and unstructured data

The result is a generated report that is given to a clinician for validation. If the clinician approves the document, the database is being updated and the ranking of most significant contributor (report) is being updated. If not, the clinician updates a report and modification of the newly created report is stored in the database.

2. Results

Prototype was tested with 12 cases from the database hosting 30 different gait reports obtained in the clinical process [3]. Most of reports in the database contained descriptions of normal gait patterns. 12 reports contained descriptions of pathological gait (cerebral palsy, and pre-surgery meniscus inquires). LanguageWare tool was used to build a dictionary and to annotate documents as shown in Figure 2 [7]. The system correctly identified 10 out of 12 cases making 83.33% success. In one case, to the system failed because of additional surface forms not having been entered in initial custom dictionary, and in another case, this happened due to the fact that our database was actually too small to host a sufficient number of reports.

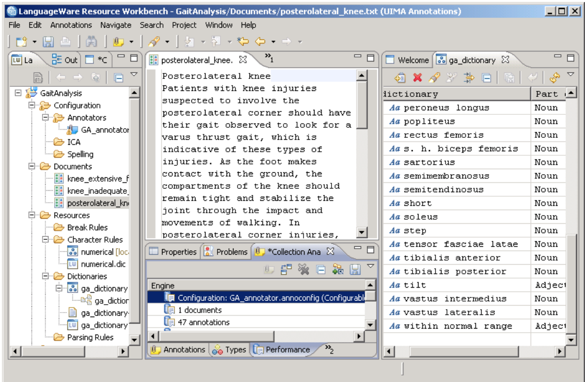

Figure 2. LanguageWare screenshot with dictionary and custom document

3. Discussion

The main advantage of our descriptive approach followed by text analytics is the possibility to validate and compare different methods of measurement. Conclusions and interpretations provided with this approach are written in human understandable form with documented and easy to follow reasoning [4, 5]. It is feasible to leverage text analytics in gait analysis. However, there is still a long way for researchers and practitioners to lay down foundations for those kinds of systems before commercial usage. One of the challenges to be addressed is the use of unstructured researched papers. Research papers typically cover more that just individual cases. Therefore, additional mechanisms need to be leveraged or article must be properly annotated. Further research should orient on dictionary unification or creation of individual and language specific dictionaries which is common practice in the Internet age.

References

[1] Sheldon S. Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems. Journal of Biomechanic 2004;37:1869-1880

[2] Medved V, Ergovic V, Tonkovic S. Towards a High Performance Expert System for Gait Analysis, IFMBE Proceedings of 4th European Congress for Medical and Biological Engineering 2008,. 2105-2108

[3] Kasovic M et al. Biomechanical diagnostics of knee joint condition in sport. Proceedings Book of 5th International Scientific Conference in Kinesiology 2008, ;242-245

[4] Ergovic V et al. Human Gait Data Mining by Symbol Based Descriptive Features. World Congress on Med. Phys. & Biomed. Eng. 2009;25:460-463

[5] Ergovic V, Tonkovic S, Medved V. Symbolization and Coding of Colored Signal Representation in Gait Analysis, IFMBE Proceedings 2011, 822-825

[6] Heimer Z. Automated clinical measurement of biomechanics and kinesiology of gait (Master Thesis). Zagreb: FER; 2005.

[7] http://www-01.ibm.com/software/ebusiness/jstart/textanalytics/infocenter.html (last accessed 08.01.2012)