Supporting trial recruitment and design by automatically interpreting eligibility criteria

Milian, K.

2014

document version
Publisher's PDF, also known as Version of record

SUMMARY

This thesis presents a small contribution to a greater goal of connecting heterogeneous medical data sources. Since clinical research is often delayed due to low participation in clinical trials, the main focus of the presented work was on supporting patient recruitment and design of eligibility criteria for new studies.

Each clinical trial aims to recruit a specific type of patients, defined by a set of eligibility criteria. This work is aimed at automatic interpretation of such criteria for the purpose of supporting verification of patient eligibility and creating a database of correlated criteria. We focused on clinical trials related to breast cancer, but many of our methods can be reused in different domains.

This thesis investigates four main research questions. The first one asks To what extent can we capture the content of medical texts such as eligibility criteria of clinical trials?. The second question goes one step further: Can we generate queries from the free text of trial eligibility criteria in order to assess patient eligibility?. Such evaluation requires the presence of corresponding data items in patient records. Since medical applications often need to deal with incomplete data, we investigated the impact of this issue on our particular task. The next question asks: Does the evaluation of incomplete data help in determining patient eligibility?. Finally, we investigated the possibility of creating a library of correlated criteria and addressed the last question: To what extent can we capture the content of a corpus of eligibility criteria, to support querying for relevant trials and criteria? We approached these questions by analyzing a large set of criteria. We utilized the observed specificity of their language, medical terminologies, healthcare standards and semantic reasoning in order to investigate

the possibility of automating the evaluation of patient eligibility for clinical trials. This research led to the following findings. First, we demonstrated a pipeline of processing steps for transforming textual eligibility criteria to a structured and finally computable representation - allowing to automatically verify patient eligibility. The structured representation consists of a combination of language patterns, concepts from standard medical terminologies and normalized measurements. The evaluation of the expressivity of this structuring approach, performed manually with a domain expert, demonstrated a nearly complete coverage on the considered set of trials. It is an automatic process that is naturally more challenging and interesting from the computer science perspective, as it would allow to scale the solution. The program which detects patterns achieves satisfactory performance. However, the detection of their correlations is more challenging, and this aspect requires further investigations.

Next, we studied the subsequent step in the pipeline. We demonstrated how to transform the structured criteria to queries. To enhance the reusability of our approach, we assumed that the data are represented according to one of the health care data representation standards (openEHR). Further, we described the experiments performed with patient data from the Maastro clinic in the Netherlands. Evaluation of the performance of the query generation from a few specific types of textual criteria showed satisfactory results but also revealed a need for further investigations. An interesting outcome of this experiment was the observation that evaluating of even an incomplete set of criteria can significantly reduce the effort required to manually inspect the remaining ones. Finally, we aimed at the automatically building of a library of structured criteria that enables semantic search for relevant criteria and trials. In order to achieve this, we applied the previous algorithm, which processes textual criteria, on a corpus of breast cancer clinical trials. The fraction of automatically structured criteria indicates that the pattern detection algorithm requires an extension, in order to be applicable for criteria composed of multiple patterns. Nevertheless, the obtained results allowed us to further process the criteria. We designed a method which compares the criteria on their restrictiveness and orders them accordingly. This algorithm can be used to support criteria relaxation during the trial design phase. The subsequent

evaluation of the approach from the medical perspective encouraged further research. The resulting library of criteria can facilitate the reuse of the structured representations, searching for related but broader criteria and relevant trials with the fine grained queries.

This thesis presents an interesting use case which demonstrates how the specificity of a domain and integrated results from many areas may be utilized to deliver methods relevant for specific users. From a biomedical clinical research perspective this work might potentially lead to a more efficient recruitment of patients for clinical trials, and faster finalization of trials and delivery of scientific evidence.