Author’s response to reviews

Title: An open access medical knowledge base for community driven diagnostic decision support system development

Authors:

Lars Mueller (lamueller@ucsd.edu)
Rashmi Gangadharaih (rgangadh@gmail.com)
Simone Klein (simone.celia@alum.mit.edu)
James Perry (james.s.perry@gmail.com)
Greg Bernstein (teteae@gmail.com)
David Nurkse (dave.nurkse@gmail.com)
Dustin Wailes (dwailes@ucsd.edu)
Rishi Graham (rishig@soe.ucsc.edu)
Robert El-Kareh (relkareh@ucsd.edu)
Sanjay Mehta (srmehta@ucsd.edu)
Staal Vinterbo (staal.vinterbo@ntnu.no)
Eliah Aronoff-Spencer (earonoffspencer@ucsd.edu)

Version: 1 Date: 02 Jan 2019

Author’s response to reviews:

Please find also the attached cover letter explaining our changes and reactions to reviewer comments. I am copying the contained text:

The feedback was very helpful to improve our submission. We added a new limitations sections, strengthened the connection to ontologies and simplified the described evaluation algorithm. In the following, we outline the specific changes for each reviewer comment:
Reviewer 1:

1. Why don't you consider health ontologies and semantic terms instead of 'terms'?

We have added more references and language to refer to ontologies. Our system is based on the UMLS Metathesaurus which provides interoperability and access to most medical ontologies. The critical mapping of semantic disease terms to observations does not exist yet. Our paper, is a first step to close this gap.

2. What is the meaning of "the incremental nature of the diagnostic process"?

We added a sentence to explain how a diagnosis is often incrementally refined to narrow the differential diagnosis.

3. I should a deeper review of the background, e.g. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2211452/.

This is an excellent paper highlighting issues both in ontologies and in clinical support tools that do not account for problems such as negation. The absence of key findings is often the lynch pin of medical decision making and is poorly accounted for in modern CDSS. In our work findings are coded with paired conditional probabilities that can account for negation. We emphasized these aspects in the new limitations section.

4. Which diseases did you work? I think your paper need to consider that every disease has its own peculiarity and details.

As noted in section 2, we focused on infectious diseases. A full list of resources can be found in the supplements. The new limitations section includes a critical discussion of our modelling approach and the current size of the knowledgebase.

5. I should recommend a second statistic study please check https://www.sciencedirect.com/science/article/pii/S1071581985710816

This is an excellent reference and we are aware of Grubers work. We believe a statistical calculation on the ontology is beyond the scope of this work. This paper is primarily meant to demonstrate the feasibility of using an UMLS based ontology, using preferred terms, to act as an open platform for prototyping DDSS. Ultimately we hope the knowledge base can grow and take more advantage of the UMLS structure (and related ontologies) by utilizing mappings such as synonyms and related terms.

6. I think you need a limitation section, Disease are so much different to treat as one DSSS, I think we need to consider how to combine the study independently by the disease.
We added a limitations section to clarify the open challenges in our approach and how we hope to address them in the future.

7. What is the meaning of DSS, AI, ML, ML-based, knowledge-based?

The first paragraph introduces indeed many of these acronyms, which is hard to read. We added a sentence to explain the main purpose of DDSS. The first two paragraphs introduce AI as a category of systems that include machine learning (ML) and knowledge-based systems.

8. "Error! Reference source not found". I hope it could be for the Figure 1

We fixed this.

Reviewer 2:

9. I don't know what "used symptoms" means here. Did you intend to say "usual" symptoms?

We intended to say that the symptom was used in the query interface. We removed the word “used” to avoid confusion.

10. I am not following the text in lines 166-173. Please provide more detailed explanation here, as this seems to be an essential part of your method. Why are you doing this only for sufficiently large q and what do you mean by sufficient?

The described algorithm (inspired by Reggia) was only used to evaluate the knowledge base in the first steps of the development. We avoided a longer description to avoid distracting from our key contribution, the knowledgebase. “q” refers to the number of possible explanations. With the growth of the knowledgebase, q became larger resulting in the usage of the greedy algorithm. We checked that in our evaluation only the greedy algorithm was used. Therefore, we removed the detailed description of Reggia’s algorithm.

11. Minor revisions 1-7

Thank you for pointing out these mistakes. We corrected them in this revision.