

July 27, 2015

Melissa Haendel, Nicole Washington, Maricel Kann

Co-chairs, *PSB '16: Innovative approaches to combining genotype, phenotype, epigenetic, and exposure data for precision diagnostics*

Dear Dr. Haendel, Dr. Washington, Dr. Kann,

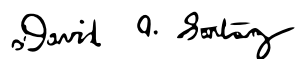
Enclosed please find our paper entitled “Learning Low-Dimensional Representations of Medical Concepts”, which we submit to the PSB as an original research article for consideration in the PSB 2016 session “Innovative approaches to combining genotype, phenotype, epigenetic, and exposure data for precision diagnostics”. The submitted paper contains original, unpublished results, and is not currently under consideration elsewhere. All co-authors concur with the contents of the paper.

Our paper shows how to learn low-dimensional representations for a wide range of medical concepts (e.g., UMLS CUI's, medications, labs, ICD9 codes), putting these into a common space. Our algorithms construct these embeddings across different modalities using the temporal data of millions of patients. We think this will be broadly useful as a means of combining different phenotypic information for precision diagnostics. For example, using these one can obtain a low-dimensional representation of each patient, which would be useful for computing patient similarity. It may also be very helpful for specifying inclusion criteria for electronic phenotyping and for cohort selection, allowing users to query to find other diagnosis codes and lab test results that might be relevant to their desired query.

Thank you for your consideration of our paper.

On Behalf of all the authors:

Sincerely,



David Sontag

Assistant Professor,

Department of Computer Science

715 Broadway, 12th floor, Room 1204

New York University

New York, NY 10003

Phone: (+1) 212-998-3498

Email: dsontag@cs.nyu.edu