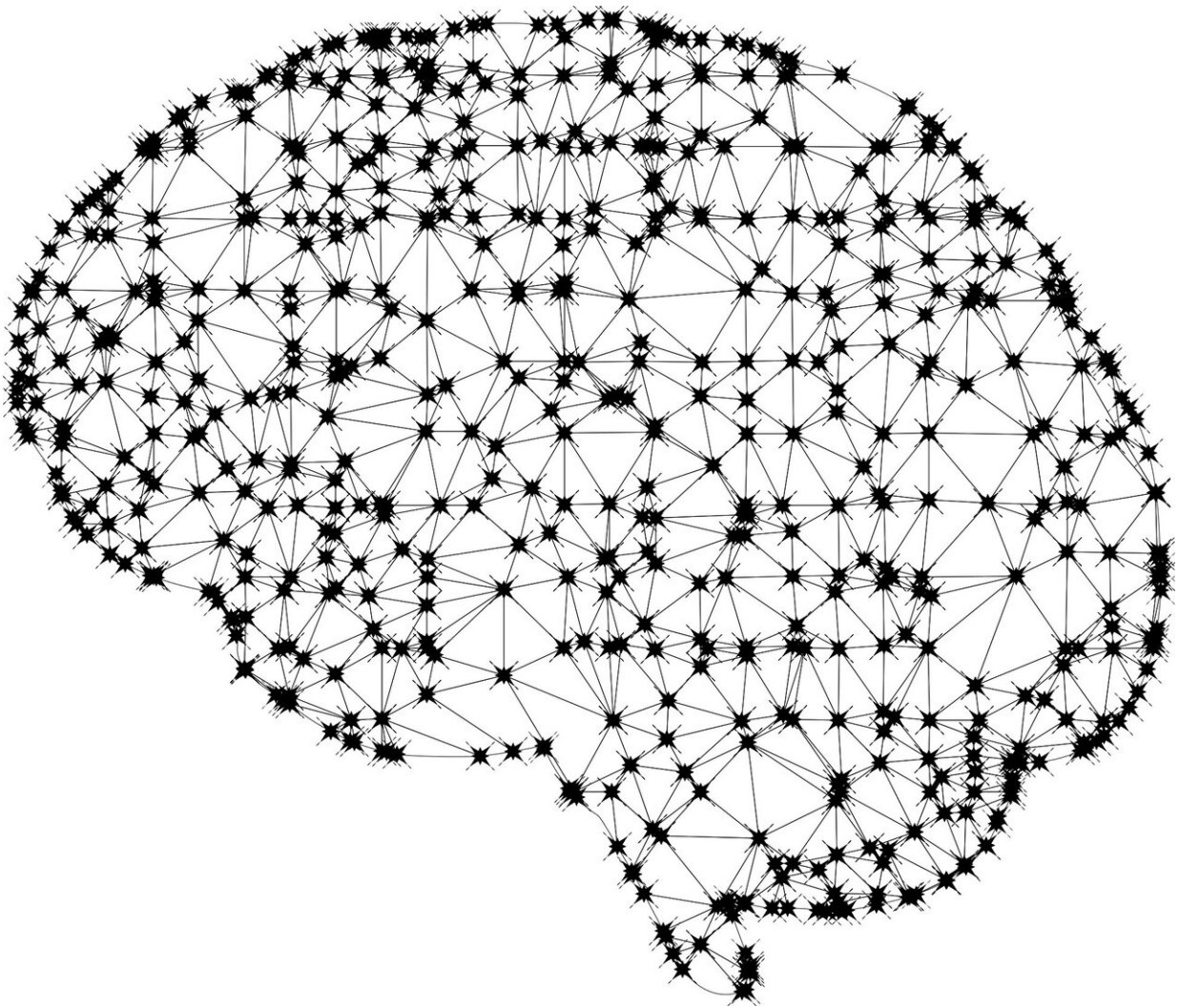


How well can computers connect symptoms to diseases?

January 10 2020, by Rob Matheson



Credit: CC0 Public Domain

A new MIT study finds "health knowledge graphs," which show relationships between symptoms and diseases and are intended to help with clinical diagnosis, can fall short for certain conditions and patient populations. The results also suggest ways to boost their performance.

Health knowledge graphs have typically been compiled manually by expert clinicians, but that can be a laborious process. Recently, researchers have experimented with automatically generating these knowledge graphs from [patient data](#). The MIT team has been studying how well such graphs hold up across different diseases and patient populations.

In a paper presented at the Pacific Symposium on Biocomputing 2020, the researchers evaluated automatically generated health knowledge graphs based on real datasets comprising more than 270,000 patients with nearly 200 diseases and more than 770 symptoms.

The team analyzed how [various models](#) used electronic health record (EHR) data, containing medical and treatment histories of patients, to automatically "learn" patterns of [disease](#)-symptom correlations. They found that the models performed particularly poorly for diseases that have high percentages of very old or young patients, or high percentages of male or female patients—but that choosing the right data for the right model, and making other modifications, can improve performance.

The idea is to provide guidance to researchers about the relationship between dataset size, model specification, and performance when using electronic health records to build health knowledge graphs. That could lead to better tools to aid physicians and patients with [medical decision](#)-making or to search for new relationships between diseases and symptoms.

"In the last 10 years, EHR use has skyrocketed in hospitals, so there's an

enormous amount of data that we hope to mine to learn these graphs of disease-symptom relationships," says first author Irene Y. Chen, a [graduate student](#) in the Department of Electrical Engineering and Computer Science (EECS). "It is essential that we closely examine these graphs, so that they can be used as the first steps of a diagnostic tool."

Joining Chen on the paper are Monica Agrawal, a graduate student in MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL); Steven Horng of Beth Israel Deaconess Medical Center (BIDMC); and EECS Professor David Sontag, who is a member of CSAIL and the Institute for Medical Engineering and Science, and head of the Clinical Machine Learning Group.

Patients and diseases

In health knowledge graphs, there are hundreds of nodes, each representing a different disease and symptom. Edges (lines) connect disease nodes, such as "diabetes," with correlated symptom nodes, such as "excessive thirst." Google famously launched its own version in 2015, which was manually curated by several clinicians over hundreds of hours and is considered the gold standard. When you Google a disease now, the system displays associated symptoms.

In a 2017 Nature Scientific Reports paper, Sontag, Horng, and other researchers leveraged data from the same 270,00 patients in their current study—which came from the emergency department at BIDMC between 2008 and 2013—to build health knowledge graphs. They used three model structures to generate the graphs, called logistic regression, naive Bayes, and noisy OR. Using data provided by Google, the researchers compared their automatically generated health knowledge graph with the Google Health Knowledge Graph (GHKG). The researchers' graph performed very well.

In their new work, the researchers did a rigorous error analysis to determine which specific patients and diseases the models performed poorly for. Additionally, they experimented with augmenting the models with more data, from beyond the emergency room.

In one test, they broke the data down into subpopulations of diseases and symptoms. For each model, they looked at connecting lines between diseases and all possible symptoms, and compared that with the GHKG. In the paper, they sort the findings into the 50 bottom- and 50 top-performing diseases. Examples of low performers are polycystic ovary syndrome (which affects women), allergic asthma (very rare), and prostate cancer (which predominantly affects older men). High performers are the more common diseases and conditions, such as heart arrhythmia and plantar fasciitis, which is tissue swelling along the feet.

They found the noisy OR model was the most robust against error overall for nearly all of the diseases and patients. But accuracy decreased among all models for patients that have many co-occurring diseases and co-occurring symptoms, as well as patients that are very young or above the age of 85. Performance also suffered for patient populations with very high or low percentages of any sex.

Essentially, the researchers hypothesize, poor performance is caused by patients and diseases that have outlier predictive performance, as well as potential unmeasured confounders. Elderly patients, for instance, tend to enter hospitals with more diseases and related symptoms than younger patients. That means it's difficult for the models to correlate specific diseases with specific symptoms, Chen says. "Similarly," she adds, "[young patients](#) don't have many diseases or as many symptoms, and if they have a rare disease or [symptom](#), it doesn't present in a normal way the models understand."

Splitting data

The researchers also collected much more patient data and created three distinct datasets of different granularity to see if that could improve performance. For the 270,000 visits used in the original analysis, the researchers extracted the full EHR history of the 140,804 unique patients, tracking back a decade, with around 7.4 million annotations total from various sources, such as physician notes.

Choices in the dataset-creation process impacted the model performance as well. One of the datasets aggregates each of the 140,400 patient histories as one data point each. Another dataset treats each of the 7.4 million annotations as a separate data point. A final one creates "episodes" for each patient, defined as a continuous series of visits without a break of more than 30 days, yielding a total of around 1.4 million episodes.

Intuitively, a dataset where the full patient history is aggregated into one data point should lead to greater accuracy since the entire patient history is considered. Counterintuitively, however, it also caused the naive Bayes model to perform more poorly for some diseases. "You assume the more inpatient information, the better, with machine-learning models. But these models are dependent on the granularity of the data you feed them," Chen says. "The type of model you use could get overwhelmed."

As expected, feeding the model demographic information can also be effective. For instance, models can use that information to exclude all male patients for, say, predicting cervical cancer. And certain diseases far more common for elderly patients can be eliminated in younger patients.

But, in another surprise, the demographic information didn't boost performance for the most successful model, so collecting that data may be unnecessary. That's important, Chen says, because compiling data and training models on the data can be expensive and time-consuming. Yet,

depending on the model, using scores of data may not actually improve performance.

Next, the researchers hope to use their findings to build a robust model to deploy in clinical settings. Currently, the health knowledge graph learns relations between diseases and symptoms but does not give a direct prediction of disease from symptoms. "We hope that any predictive [model](#) and any medical knowledge graph would be put under a stress test so that clinicians and machine-learning researchers can confidently say, "We trust this as a useful diagnostic tool," Chen says.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: How well can computers connect symptoms to diseases? (2020, January 10) retrieved 9 April 2024 from <https://techxplore.com/news/2020-01-symptoms-diseases.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
