Building Knowledge Graphs for Diagnostic Medicine Cardiovascular

Hao Lin (hl3307) Xinpei Ma (xm2225)

Research Proposal

2020-02-12

Abstract: Cardiovascular diseases have become one of the major killers of health due to its own characteristics, so the diagnostic medicine of these diseases showed great social significance. As a product of the Internet era, the knowledge graph has a clear advantage in dealing with the correlation between large amounts of information and nodes. In this research, a knowledge graph will be built based on data coming from different sources. The result will be used for the diagnostic medicine of cardiovascular diseases.

1. Background (Review of Related Literature):

The cardiovascular system consists of three interrelated components: blood, the heart, and the blood vessels. It's an important part for supporting the normal functioning of the human body. However, the diseases related to the cardiovascular system are numerous and could not be underestimated. According to epidemiological surveys around the world, for every 100 people on average, 3 to 5 people have some degree of heart failure, and one in five people have varying degrees of cardiovascular disease, and the ratio is even higher in the elderly[1, 2]. In addition, Cardiovascular disease usually progresses from a small illness. For example, heart dysfunction may eventually develop into heart failure. Thus, it becomes particularly important for the development of diagnostic medicine of cardiovascular system diseases, especially when a country enters an aging society.

Several ways could be used for the diagnostic. Considering that humans and some animals have similar physiological characteristics, many scholars will use animal experiments to figure out the diagnostic medicine of cardiovascular system diseases. Although animal experiments have many advantages, their limitations are also obvious: it is difficult to control the initial conditions of animals in animal experiments, and large individual differences make it difficult for scholars to obtain better statistical laws. The extracorporeal simulated circulation system (Mock Circulatory System, MCS) can simulate the pulsating flow environment of the human cardiovascular system and reproduce the hemodynamic process of the cardiovascular system under various physiological states[3, 4]. However, there are physiological characteristics that are difficult to achieve on the extracorporeal simulated circulation system, such as the regulation of ventricular elasticity.

The Knowledge Graph is a knowledge base used by Google and its services to enhance its search engine's results with information gathered from a variety of sources. As a technology to connect and present related information, knowledge graph gives a way to significantly enhance the value of information[5]. Its advantages include flexibility, and easy visualization. Prior to this, knowledge graphs have been used in many industries to establish links between different events and to predict possible events. Thus, this time, the knowledge graph method is used for the diagnostic medicine of cardiovascular system.

2. Introduction to the Project:

First, in order to have a sufficient amount of information for building the knowledge graph, data mining is needed. At this stage, literatures will be collected, and the hospitals' and research institutes' databases will be accessed and recorded. After collecting enough data, the data needs to be cleaned up to remove irrelevant elements and merge similar elements. In this process, in order to make the subsequent data processing more targeted, the physiological meanings behind different data should be learned.

Second, a knowledge graph is planned to be built, where entities are cardiovascular system diseases and diseases related symptoms. In this process, the Natural Language Process (NLP) method will be used to extract key information from text materials and parameters. The model will be trained to make the results more precise. After that, the knowledge graph will be generated, and further study will be based on this graph. The final goal is to generate a knowledge graph where all cardiovascular diseases and disease related symptoms could be linked together and then reveal the relationship between them.

3. Introduction to the Dataset:

The data mainly comes from literature, hospital and research institutes, and detailed data information has not been obtained.

The obstacle is that we might not be able to extract all entities and relationships precisely and clearly because of the limitation of today's NLP technology. Thus, some preprocessing work on the data, like making it more standardized, might be needed.

4. Plan:

Milestone1: Focusing on accumulation of basic knowledge, a series of questions need to be answered: what's cardiovascular diseases and related symptoms? what's Knowledge Graph? How to build it? The data will be prepared during this stage, and preprocessing work on data will also be done.

Milestone2: Focusing on data processing, data from literature, hospital and research institutes will be mined.

Milestone3: Focusing on the final knowledge graph, the knowledge graph will be built by implementing NLP. First, a raw version knowledge graph will be produced. Then some improvement work could be conducted to make it clearer, and more precise.

Reference:

[1] D. M. e. al., "Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association," Circulation, vol. 133, no. 4, 4, pp. 38-360, Jan 26 2016.

[2] K. K. K. e. al., "The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-fifth Adult Heart Transplantation Report-2018; Focus Theme: Multiorgan Transplantation," J Heart Lung Transplant, vol. 37, no. 10, pp. 1155-1168, Oct 2018.

[3] D. Timms, M. Hayne, K. Mcneil, and A. Galbraith, "A Complete Mock Circulation Loop for the Evaluation of Left, Right, and Biventricular Assist Devices," Artificial Organs, vol. 29, no. 7, pp. 564-72, 2005.

[4] D. V. Telyshev, A. A. Pugovkin, and S. V. Selishchev, "A Mock Circulatory System for Testing Pediatric Rotary Blood Pumps," Biomedical Engineering, journal article vol. 51, no. 2, pp. 83-87, July 01 2017, doi: 10.1007/s10527-017-9689-4.

[5] J. Pujara, H. Miao, L. Getoor, and W. Cohen, "Knowledge Graph Identification," Berlin, Heidelberg, 2013: Springer Berlin Heidelberg, in The Semantic Web – ISWC 2013, pp. 542-557.