Visualization of Text Data for Human Health

Zekun Zhao

Abstract—In modern world, there is no doubt that data visualization has become an indispensable and important means in the field of data analysis. Everytime when we think about to present our data into computer screen by softwares like powerpoint, Prezi or Tableau, we need to find or design a better way to transform our message by data visualization techniques. However, data visualization is a huge topic, and the methods of visualization vary widely depending on the data objects and the purpose of presentation. In this paper, we want to focus on the visualization of text data for human health field.

Index Terms— Human Health   Text Data Analysis   Visualization   Web Application

---

1 MOTIVATION

Everything in the era of big data can be digitized, and the sudden emergence of artificial intelligence is also attacking cities in various industries. Aside from pure rational AlphaGo, in terms of sensibility, DeepDream has created a large number of works of art, and Xiao Bing has published poetry and so on. It seems that under the offensive of artificial intelligence, the barriers that human beings are proud of are falling apart.

After all of this, biomedical field seems to be one of the ultimate lines of defense for human beings. The volume and complexity of health information continues to increase, but the methods in which its communicated has not changed in centuries. AI can recognize people's illness through medical images, history data, and can express quite effective analysis results in different ways. However, whether such an analysis based on the implicit complex logic relationship can be equated with the true status of human beings is indeed difficult to draw conclusions.

Can the health status of human beings be visualized?

The answer is basically yes. Although the current AI technology is not enough to fully simulate human body, great progress has been made in identifying body structure and quantifying illness.

Data visualization is an intuitive display of all quantitative information by humans, and visualization can become possible when health status can be approximately quantified as numbers by artificial intelligence. Just as we can describe our health status in words, we can also use data and charts to approximate emotions.

2 RELATED WORK

2.1 From Emotion Analysis Filed

In 1980, psychologist Robert Plutchik proposed the concept of the famous emotional ring, using graphics and color to show the eight basic emotions that humans have: love, trust, fear, surprise, sadness, disgust, anger, expectation.

For each of the emotions to expand to three degrees, a total of 24 main emotions are formed, and the closer to the inner circle, the more intense the emotion. The closer you are in the emotional ring, the more similar the emotions are.

With the help of these basic psychological models, emotional quantification has become possible in both academic and engineering circles. More and more emotional models emerge in an endless stream, further strengthening the development of emotional visualization.
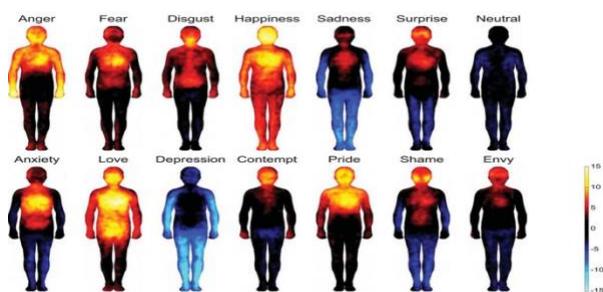
2.2 Human Health Visualization

Figure 1 *Maps of body locations where they feel basic emotions (top row) and more complex ones (bottom row). Hot colors show regions that people say are stimulated during the emotion. Cool colors indicate deactivated areas.*

Interaction has played an important role in the field of visualization and visual analysis. But the health state of the user is rarely reflected in these interactive technologies. In many scenarios, if the health status of oneself or others can be visualized, it often helps the user to make better decisions.

With the help of health data, the visualization system can deepen people's inner level and explore more possibilities. Hailed as the equivalent of Google Maps for the human body, the BioDigital Human is a scientifically accurate cloud based virtual body that empowers everyone to learn about health and medicine in an entirely new visual format. Anatomy, disease and treatments - all in an engaging, interactive 3D format that resembles life itself.

Figure 2 *Human Body Muscle Visulization*

Many researchers focused on the growing textual information as the Internet evolved. Although human health data are particularly complex, the basic dimensions can be simplified into data of different human body parts. For a large number of text comments on the Internet, such as DinXiang Doctor Inc, we can clearly get the data expressed by the comments.

Figure 3 *Interactive Human Body Platform from the BioDigital Human*

People's behavior on the Internet, such as comments or Twitters, often reflects their recent status, including the mood at the time. In order to get a data dataset about the disease, I crawled 1.5 million disease quiz data and used python to do data analysis. Through scraAn interactive health data visualization analysis system based on the text data is presented to extract the information of Internet users.

3 PROPOSED RESEARCH DIRECTIONS

The project idea is to use data analysis techniques to let users to improve their understanding about human body, reducing workload for data processing. The project objective is to explore a view for seeing health information which hided in tons of data from users' comments.

| 16 | 5 | Gender: Female | 2009-11-27 | Health consultation description: The first half of the night is itchier than the daytime. There is only one hand with a finger on the board. There is a piece of red. It has been caught for a while and it is better. Just 3 days. The first is to sleep for a few hours. [Evening] Itching | The condition may be related to eczema. You can try to apply eczema cream. Guidance: Your child's condition may be related to eczema. You can watch for a while and then you can Use pediatric eczema cream to smear the effect, then

you need to go to the hospital dermatology to see if you can open a drug treatment to see the effect ah life care: pay attention to keep dry, reduce the moist environment 13445134 | Itchy on the board. What is it? Thank you |

Figure 4 *Original Data Crawled from Web*

After we stored all of our data crawled from web into mySQL database, we did some simple analysis for the data, just to get a bigger picture how the data looks like.
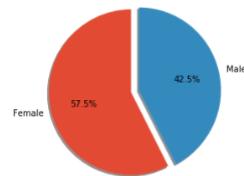

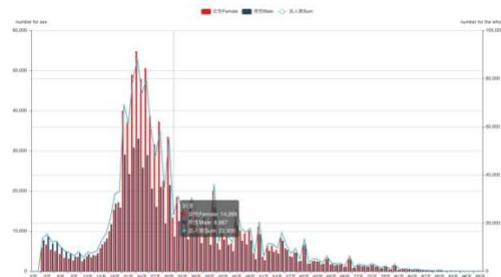**Figure 5** *the percentage of different sex in data*

Figure 6 *the percentage of different age and sex in data*

It shows that there is more female asking question in our dataset, especially for the age range between 20 to 30. Also, people in these ages seem more likely to ask questions than other ages. I assume that it is because people in these ages are busier at working, they are also better at using online services for their life.

3.1 Beginning Stage

Initialy, a basic framework for text data analysis is necessary. It should be smiple, clear and understanadable.

["breathing", 145479], ["spirit", 98481], ["cough", 85610], ["dizziness", 64465], ["tension", 61847], ["fever", 53182], [" Abdomen", 45828], ["brain", 40897], ["brain", 36356], ["nausea", 35548], ["insomnia", 35046], ["eyes", 34963], ["fatigue ", 28471], ["Neck", 27051], ["memory", 19557], ["gastrointestinal", 16715], ["chest", 16325], ["belly", 13288], ["back" , 12118], ["Teeth", 11728], ["Heartbeat", 10976], ["leg", 10920], ["menstruation", 8664], ["reproductive", 8079], ["tongue", 6519], ["nose", 6463], ["breath", 5070], ["ear", 4847], ["hair", 4736], ["shoulder", 4373], ["Sneeze", 4206], ["Scalp", 4123], ["Face", 3760], ["Masturbation", 3148], ["Left Hand", 2869], ["Right Hand", 2702] , ["Throat", 2535], ["Wrist", 2451], ["Head", 2284], ["Face", 975], ["Foot", 891], ["Mouth", 863], ["Left Arm", 724], ["Ass", 585], ["Lip", 557], ["Right Arm", 306], ["Foot" , 139]

Figure 7 *Human Body Nouns Mentioned in Text Data*

The most frequent used nouns are a good way to visualizing text information first. In our dataset, ror each noun, we can find out how many questions are concerned about this part and how did people describe about their health status.

Also, according to medical dictionary, we also build our model of frequency analysis for different illness from tons of text data. Using builded-In dictionary as tokenizes, we select different range of information from our dataset, like illness frequency in different ages, sex and time. Here is wordcloud picture we generate from people whose age is bigger than 60.

Figure 8 *Left pic is generated from ill_detail; Right pic is generated from docto_answer.*

3.2 Main Page

Then, we want to apply all the users' results to 2d human body structure in order to add spatical domain information. The reason for doing this is that we want to have an intuitive picture for the actual data. Form visualization in 2d human shape, I believe users can get more ideas and understanding for the data.
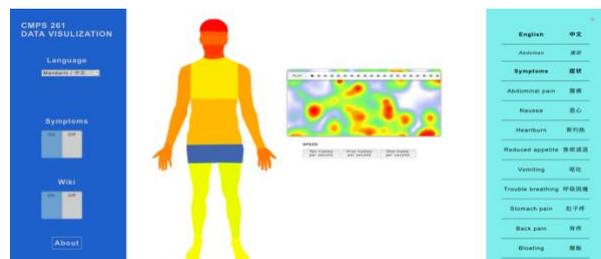

Figure 9 *Main Page for Visualization of illnss; left part is control panel; milddle one is the frequency visualization in human body domain and time domain; right part is some descriptions for certain part of human.*

The initial idea of this project is to help people who dose not know how to express their feelings and need quick info about some illness. Some senior people cannot type which actually block them from lots of resources online. But our website info is only base on click operation, which minimize the effort for users to learn how to collect info they need. The human body

picture is one SVG image which allows we set value for different part of human body.

Since diferent colors represent the frequency of the question concerning about one specific human body part in our project, we want to use the same things generate a different view for user to get the ideas about the relationships between ages or time and other health information. In order to keeping thoes great features in our works, I used heatmap to visualize user infomation in 2d image.

Figure 10 *The Number of illness changes over different age level in regular heatmap*

The heatmap in Fig 10 gives us some useful ideas that we can see few illness are actually become more serious when we become older, but some are more often occured when we are younger. It is great for us to notice this trend before certain age, so that we can do some exercise or health care services to minimize the possibility of surffering this illness.

Figure 11 *The Number of illness changes over different age level in animation style*

Same idea shows in Fig 11, but I also consider animation effect for the work to illustrate the trend that how the value grows in different time period. Putting animation effect to our heatmap is valuable for lots of aspect, we can get more dimensional info from it.
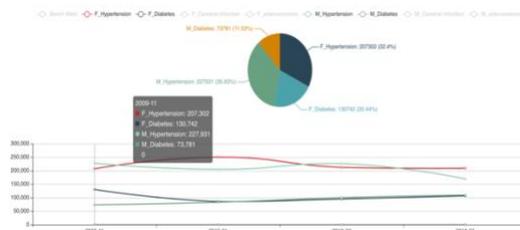

Figure 12 *The Number and precentage of illness changes over time for certain part of human body(head)*

There are lots of efforts for processing statistic data. In Fig 12, we show 4 major illness which cover 60% of the whole number of illness coccured in our dataset for people whose age is bigger than 60. We can see trend and percentage for female and male in different time period.
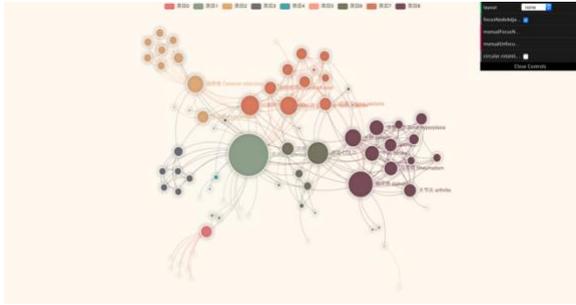
Figure 13 *The relationship between different illness in our dataset, different node color represents different part of human body, node size represents how frequency is this illness, edge represents two connected illness could happen in one type of symptom.*

Different ways to visualize our data can let us see different aspects of data. The method in Fig 13 is useful of medical organization to design more effective method to help people's health status. The way in Fig 10 and 11 is great for individual to understand more about illness in different ages. In addition, we also explored some relationship between illness. Fig 14 illustrates our idea in a graph base model. In current version, the position of each node is fixed, but we want to design and implement one way to generate position for each node, when we get all info about nodes and edges.

4 FUTURE WORK

In this project, we considered frequency as our main evaluation method for different illness, but ignore other info like how serious it is, and how soon usually does this illness can be cued. I believe that all these infomation could give us more ideas about the insight of data.

More, we only have data from user's age, sex and time but no any info about doctors.

Thus, it is hard to evaluate how confidence the answer could be.

I also want to use more visualization techniques to shows people how the big picture from our data are generated, which could also make the result become more believeable for our users.

REFERENCES

[1] Deqing, L., Honghui, M., Yi, S., Shuang, S., Wenli, Z., & Junting, W., et al. (2018). Echarts: a declarative framework for rapid construction of web-based visualization. *Visual Informatics*.

[2] Chen, Z., Wang, Y., Sun, T., Gao, X., Chen, W., Pan, Z., ... & Wu, Y. (2017). Exploring the design space of immersive urban analytics. Visual Informatics, 1(2), 132-142.

[3] Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What Do We Talk About When We Talk About Dashboards? IEEE transactions on visualization and computer graphics, 25(1), 682-692.

[4] Lu, Y., Wang, F., & Maciejewski, R. (2014). Business intelligence from social media: A study from the vast box office challenge. IEEE computer graphics and applications, 34(5), 58-69.