“Decoding · Encoding”

– an exploration of data narrative in Tibetan characters

Anqi Song, Xintong Song, Yuhao Chen, Guangyu Luo, Qiansheng Li *

Shanghai University

Abstract—Tibetan，as a communicative tool of the Tibetan people, it has a long history second only to Chinese in China. Tibetan is a beautiful and expressive language. It is difficult for us to come into contact with this ancient and mysterious writing in our daily life. How can we overcome the time-space distance and language barriers to establish cultural communication in the ever-changing modern society? This is the research background of our project. We created an interactive website, which driven by data narrative visualization and interactive experience, foucs on Tibetan characters. We tried to decode Tibetan artifacts, the writing patterns of Tibetan calligraphy and Tibetan poetry by data, reconstructing our understanding of the Tibetan culture, promoting and improving the mutual connection between different cultures.

Index Terms—data visualization, data narrative, Tibetan calligraphy, Tibetan

INTRODUCTION

Text is the precipitation of history, which can help people know a nation's past with its unique characteristics. Text is a tool for people to communicate with each other. Words are also the smallest unit of connecting cultures, their manifestations in our lives can be language, writing and reading. Our project tells a data story from Tibetan characters to Tibetan culture in an innovative and experiential way and bring the Tibetan language into the public's eye. We want to use data visualization to build a bridge between Tibetan and Chinese cultures, to promote the spread of Tibetan culture.

1 RELATED WORK

In this project, we collected 30 letters of Tibetan from the Internet, books and signboard images in Tibet region. A comparative experiment was designed in writing: We invited a Tibetan calligraphy inheritor from Guoluo, Qinghai and a student who wrote Chinese calligraphy to join our experiments. They were asked to write the Tibetan and Chinese versions of diffrents kinds of contents. During the process of writing, we collected their brain activity data and EMG data in order to compare and analyze participants' psychological and physiological activities. On the other hand, we disassemble the written Tibetan and Chinese poems and compare the morphology of characters by using parametric analysis of their "ink ratio" and "literal" data.

1.1 Brain Activity Data (EEG data)

In exploring the EEG data collection of writing movement, we used Mindwave Mobile equipment to collect the EEG of Tibetan calligraphy inheritors and students who write Chinese calligraphy.
Mindwave Mobile uses a sensor placed on the forehead and a reference electrode contact placed on the ear to measure brain wave signals, collecting data in eight dimensions as well as eSense index values. The eSense index value is ThinkGear's amplification of the original brainwave signal, and in the process of filtering out the interference caused by environmental noise and muscle movement, so that the collected brainwave data is more reliable. Finally, the processed EEG signals were calculated using eSenseblem algorithm to represent people's concentration and relaxation,which used to analyze the psychological changes of Tibetan calligraphy inheritors and Chinese calligraphy students during writing exercise.

1.2 Electromyogram Data (EMG data)

In the experiment, MYO bracelet was used to collect the surface electromyography signals of the right arm of Tibetan calligraphy inheritors and students studying Chinese calligraphy, which measure the changes of muscles when the wearer's gestures change.
Myo is mainly composed of 8 bioelectrical sensor units with different sizes and thicknesses, with built-in three-axis accelerator,
It collects data on four dimensions, EMG, acceleration, gyroscope and direction, to analyze the differences between the forces used to write two different characters.
Since both the myo bracelet and Mindwave Mobile are connected via bluetooth, we collect them by two computers to avoid interference with each other. And when the equipment is interrupted during the experiment, we will ask the collector to stop writing and do it after the equipment is connected to ensure the accuracy of the data.

1.3 Other Data

In addition to the data collection of brain waves and electromyography, we also collected and stored images, photographed the writing process as video data, and recorded everyone's writing time. Because bamboo pens are used in Tibetan calligraphy writing, it is necessary to dip the ink bottle frequently. The number of times the inheritor has to dip the ink bottle to write the poem can be counted through the video. The collected information can be stored permanently, at the same time, people can share information anytime and anywhere, and the data can be improved and modified continuously in the later period according to different needs of use.
We collected the high-resolution images of signboards with Tibetan characters, and sorted out the Tibetan, Chinese and English versions of “Tsangyang Gyatso Love Poems”, analyzed the emotional values of 64 poems and 270 sentences.

2 DATA VISUALIZATION DESIGN

After the data collection, we obtained the following data: experimental process images, written text scan image, writing duration data, EEG data and EMG data, ink ratio of words, literal analysis data, Tibetan signboard pictures and poetry texts. We have processed the above and make visual design, produced a data visual narrative web page with the theme of deconstructing Tibetan.

2.1 Visual Analysis of EEG and EMG data

In the experiment, we choose the concentration and relaxation in the brain activity data, analyzed them in the form of histogram, and visualized them with polar coordinates. A dark yellow histogram with a value greater than zero indicates concentration, while a light yellow histogram with a value less than zero indicates relaxation.

* nick_lee@shu.edu.cn

A larger scale on the EEG indicates either concentration or relaxation.We find that Chinese characters takes a short time to write and it is easier to concentrate, while the Tibetan word formation takes a long time due to its complex structure. In this process, which requires a high degree of concentration, and also shows a high degree of relaxation.

In terms of EMG data processing, we used variance formula to calculate the variance value of EMG data, which was used to express the intensity of muscle activity at this time, and then processed it into a smoother curve to observe the variation trend of EMG data when writing different characters. Due to the use of soft head brush in Chinese calligraphy, the writing dynamic is soft, and it can be seen that the muscle data changes little and the lines are relatively gentle. The Tibetan calligraphy uses bamboo pens, which requires better control and greater force, so the muscle data changes greatly and the lines fluctuate violently.

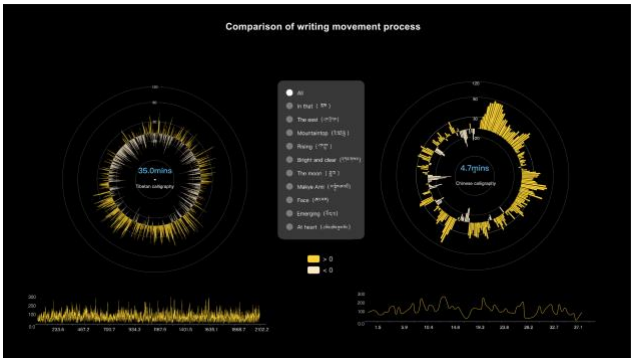

Fig. 1. Visualization of brain wave data and Electromyogram

2.2 Visual Analysis of Tibetan and Chinese characters

In the comparative analysis of the data of ink ratio and literal block size, we need to ensure that the writing standard and writing range of Tibetan and Chinese characters are the same, and then use the text analysis tool to get the ink ratio and literal block of the corresponding words. The amount of ink used is presented in the form of the size of the rectangular area. In the case of a certain length, by compareing the height of the width, it can be concluded that the ink ratio written in Tibetan words is mostly larger than that written in Chinese characters. The literal analysis of Tibetan and Chinese characters is presented with parameterized geometric figures; Chinese characters are irregular polygons close to squares, while Tibetan characters are more slender.

2.3 Visual Analysis of Textual data

We made the visualization analysis and emotional interpretation of 64 poems in “The Love-songs of 6th Dalai Lama Tshangyang Gyatsho", And refered to the form of word cloud diagram, relationship diagram and tree diagram to design a new visualization form. The graph consists of three layers, the outermost ring is the first sentence of the 64 love poems. We can view the entire poem through touching the sentence by the mouse pointer, It visually presents the relationship between the sentences before and after the poem. We put the emotional value of poetry on a scale from zero to one hundred. On the one hand, the lower the number, the more negative the emotion and the lighter the color; on the other hand, and the higher the number, the more positive the emotion and the darker the color. In the middle color ring, each color block represents a sentence, which is helpful for us to understand the emotional orientation of each poem. The most frequent key words in poetry are in the innermost layer of the ring. If we touch a certain word, the corresponding verse with it will present. Words such as lover, heart, woman and so on throughout the whole poetry anthology, fully expressed Tsangyang Gyatso's eager and sincere pursuit of love life.

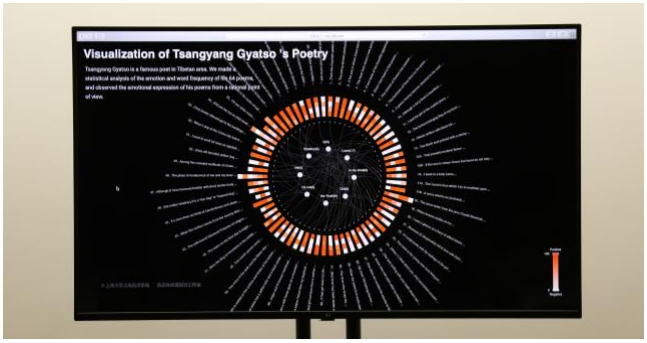

Fig. 1. Text visualization of Tsangyang Gyatso 's love poems

3 Data visualization website design

We processed and visualized the collected data. The webpage is divided into three parts: deconstructing Tibetan characters; deconstructing Tibetan calligraphy through data visualization; deconstructing text.

This webpage allows users to navigate to view the information, with greater autonomy and exploratory nature. Reasonable use of visual interactive functions can improve users' cognition of data. Using game interaction, hand-written Tibetan text, etc. to enhance the audience's sense of participation, interest and in-depth understanding of data insights in the display of Tibetan calligraphic data stories.

4 Conclusion

Our practice means proved that it is feasible and effective to collect data in the process of writing Tibetan calligraphy by equipment. In the current experiment, we only obtained data from one calligraphy inheritor of Tibetan and one calligraphy student of Chinese, the date comparison may have some deviation. In the future, we need to collect more data for analysis and comparison, using data visualization to build a bridge between Tibetan and Chinese cultures, and promote the spread of Tibetan culture.

References

IEEE Transactions on Visualization and Computer GraphicsVolume [J]16, Issue 6. 2010. PP 1139-48

Gibbs, Mark. NeuroSky MindWave Fun with brainwaves[J]. Network World,2012,29(3).

Qiu Qingju. Research on feature extraction and pattern classification of surface EMG signal[D]. Shanghai Jiaotong University, 2009.

Cai Zhijie, Cai Rang Zhuoma. Research on the Distribution of Tibetan Characters[J]. Chinese Journal of Information, 2016, 30(04): 98-105.