IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.005, LIIR277
DOI: https://doi.org/10.52731/liir.v005.277

Extraction of Emotional Patterns in Novels and Analysis of Their Transitions

Masako Furukawa [*], Yoshitomo Yaginuma [†]

Abstract

As part of research to analyze novels, this paper extracts emotions from novels and performs clustering to classify the patterns in which emotions appear in novels. Since transitions in emotions are important in novels, we clarify how emotions transition from the preceding parts of the novel to the ending parts, which are important for concluding the story. We also examine whether the patterns of appearance of emotions differ depending on the authors.

Keywords: Novels, Emotion analysis, Clustering, Emotional transition

1 Introduction

Research on the use of AI for creative tasks like generating images and writing documents is widespread [1-4]. In Japan, there's an award called the Hoshi Shinichi Prize, named after a well-known short story writer, Shinichi Hoshi [5]. This award accepts AI-created entries, but so far, no entry made solely by AI has won. For AI to write novels, it's necessary to understand the features of novels written by humans in a way that computers can manage.

Novels have structure, and AI needs to understand this structure in order to make advanced use of novel data. There has been much research to identify patterns in the structures of novels. Propp analyzed and categorized Russian fairy tales [6]. Murai focused on the actions of characters to describe the structure of stories [7]. Furukawa et al. clarified how well a computer can distinguish between the ending parts of novels and their preceding parts using random forest [8]. Muhammad compared fairy tales and novels and revealed the differences in emotional expression between them [9]. Nanyun et al. proposed a framework for analyzing story corpora for generating stories [10]. Reagan et al. studied the emotional arcs in stories and identified six types, including a tragic one where the emotions continue to decline [11].

This paper focuses on emotions as part of the structural analysis of novels. Clarifying the changes in emotion in novels can not only be used as a basis for novel generation, but also help in understanding novels more deeply, and in language learning, it can also be used to generate learning materials such as dividing a novel into parts based on individual emotions. As an analysis of emotions, we extract emotions from novels and use clustering to identify patterns in how emotions are presented. Since the transition of emotions is important in novels, we analyze how emotions evolve from the beginning to the end, which is crucial for wrapping up the story. We also explore whether the patterns of emotional representation vary among different authors.

[*] National Institute of Informatics, Tokyo, Japan
[†] The Open University of Japan, Chiba, Japan

2 Data Acquisition

The novel data from Aozora Bunko [12] were used for analysis. Aozora Bunko is an initiative that aims to collect free e-books on the Internet. Similar to a library, anyone can access it. It includes works whose copyrights have expired and works that are "free to read," digitized in text and HTML formats.

First, we downloaded all document data registered as of October 2022 from Aozora Bunko. As a result, we obtained 17,319 documents. However, this total includes some works whose copyrights have not expired. It also includes documents written in old kana orthography and documents other than literary works. For this reason, we only extracted documents whose copyrights had expired, that were written in modern kana orthography, and whose NDC was in the 910 range. NDC stands for Nippon Decimal Classification, and the 910s represent Japanese literature. As a result, we obtained 7,119 works.

Looking at these documents, some had notes about notation methods at the beginning of the text. Additionally, some documents included source information at the end, and some included information about the pronunciation of words. Therefore, we created a parser to remove these elements. The header section was deleted by setting conditions such as the header being separated by "---". Source information was removed by setting conditions such as following the word "source". Information on pronunciation and so on was deleted under conditions such as being enclosed in special parentheses.

After these preprocessing steps, we calculated the number of sentences in each document. Generally, "." is used as a delimiter between sentences. Analyzing the number of sentences extracted, we found that many documents had a small number of sentences. When performing emotion extraction, to exclude documents that were too short or too long, we only included documents with between 100 and 1,200 sentences. The 3,711 files obtained in this way were the targets for subsequent processing.

3 Extraction of Emotion

Pymlask [13] was used for emotion extraction. Pymlask is a Python version of ML-Ask [14], released under the BSD 3-Clause License. By pattern matching using a dictionary of 2,100 words, 10 types of emotions are estimated: joy, anger, sorrow, fear, shame, liking, dislike, excitement, relief, and surprise. Table 1 shows examples of words expressing each emotion in the dictionary.

The purpose of this paper is to classify the patterns in which emotions appear in novels and to clarify the transitions between the ending parts and the preceding parts of novels. For this reason, emotion extraction was performed by dividing the novels into ending parts and preceding parts. The novels were divided into a ratio of 8:2 from the beginning, with the latter being the ending parts and the former being the preceding parts. Since the 3,711 files were divided in two, emotion extraction was performed on 7,422 files.

To extract emotions, we first counted whether each sentence in the documents contained any of the 10 types of emotions. That is, if a sentence contained words expressing a certain emotion, it was counted as 1, and if it did not, it was counted as 0. For example, if a sentence contains words

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

expressing joy and anger, the expression of emotion is (1,1,0,0,0,0,0,0,0,0). Next, if a certain document consists of N sentences, the number of emotions in the N sentences is added up. For example, if the number of sentences containing joy is 3 and the number of sentences containing anger is 2, the expression of emotion of the document is (3,2,0,0,0,0,0,0,0,0).

Since the size of document data varies, to eliminate this influence, we normalized it so that the sum of emotions equals 1, and used it as the final emotion expression of the document. This represents the proportion of each emotion among all emotions.

Table 1: 10 emotions and examples of words that express those emotions

ID	Emotion	Example of words
1	Joy	Happiness, Satisfaction, Smile, Pleasant, Enjoy
2	Anger	Scold, Glare, Displeasure, Exasperate, Criticize
3	Sorrow	Sadness, Tears, Loneliness, Crying, Tragedy
4	Fear	Horrible, Anxiety, Hesitate, Eerie, Cowardice
5	Shame	Insult, Blushing, Humiliation, Shyness, Embarrassment
6	Liking	Cute, Love, Admire, Passion, Familiarity
7	Dislike	Pain, Suffering, Malice, Dissatisfaction, Annoyance
8	Excitement	Excited, Exploding, Confused, Nervous, Shaking
9	Relief	Calm, Safe, Peaceful, Relaxed, Relieved
10	Surprise	Unexpected, Amazement, Panic, Daze, Bewilderment

4 Clustering of Emotion

We perform emotion clustering to find typical emotional patterns in novels. Here, the input emotion vectors were standardized using the mean value and standard deviation before clustering was performed using the k-means method. In the k-means method, it is necessary to specify the number of clusters in advance. The silhouette method was used to determine the number of clusters, and k = 6, which had the largest silhouette coefficient in the range from k = 2 to 14, was selected. These six clusters were named A, B, C, D, E, and F, respectively. The labels are arranged in descending order by the number of data points. The number of data points in each cluster is 2686, 1364, 1218, 1005, 699, and 450, respectively.

Figure 1 shows the centroid position of each cluster by radar chart. It shows how much of the 10 emotions in Table 1 are included in the six clusters. The further away from the center, the stronger the emotion.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Looking at the characteristics of each cluster, Cluster A has peaks in Anger and Dislike, but the values are relatively low, suggesting a cluster of almost flat emotions. Cluster B has peaks in Sorrow and Excitement. Cluster D has peaks in Joy and Relief. Similarly, Cluster C has a peak in Liking, Cluster E in Fear, and Cluster F in Shame. These results revealed several typical characteristic emotional patterns that appear in novels.

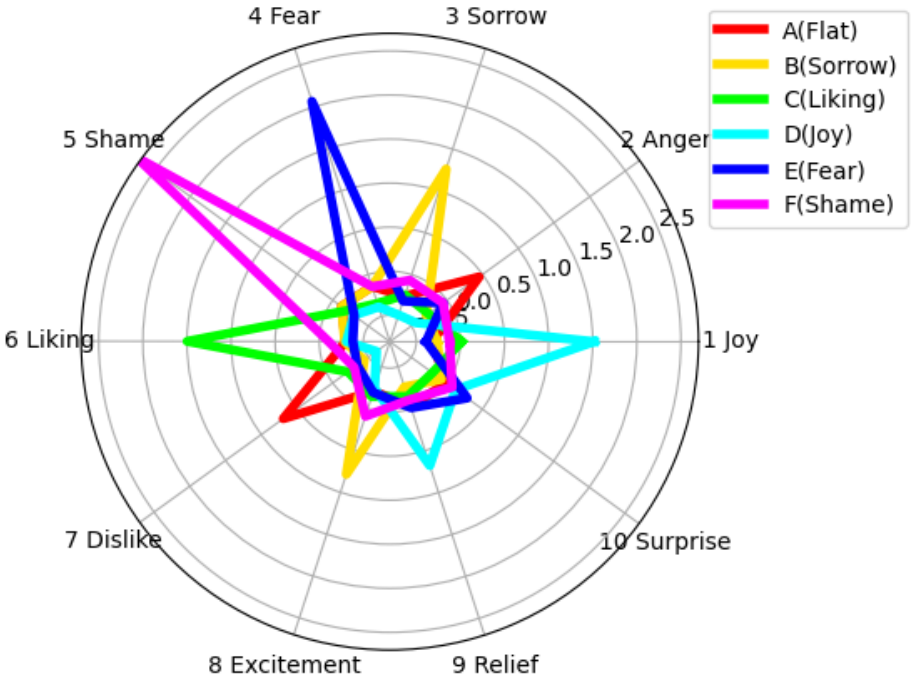

Figure 1: Clustering of emotion

5 Emotional Transitions in Novels

Next, we examine the transition of emotions in novels. Figure 2 shows a Sankey diagram representing the flow of the transition from six emotions in the preceding parts to six emotions in the ending parts.

Regarding the preceding parts, it can be seen that there are many instances of the Flat emotion, and there are a few instances of Fear and Shame. On the other hand, in the ending parts, it can be seen that the Flat emotion decreases and emotions such as Shame and Fear increase.

Regarding transitions in emotions, a relatively high proportion of the emotions remained the same in both the preceding and the ending parts. However, only about 1/3 of the most common Flat emotion transitions to Flat, and about 2/3 transition to other emotions.

Figure 3 shows the results of extracting only Osamu Dazai's works and creating a Sankey diagram to see whether the pattern of emotional transition depends on the author. Looking at this result, it can be seen that there are fewer instances of Fear and more of Shame compared to Figure 2. Dazai is known for his dark style of portraying his inner self, and this result is thought to be a reflection of that. Structuring emotions by author would be a useful perspective for analyzing similarities between authors and as a basis for AI to understand and use the structure of the

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

work.

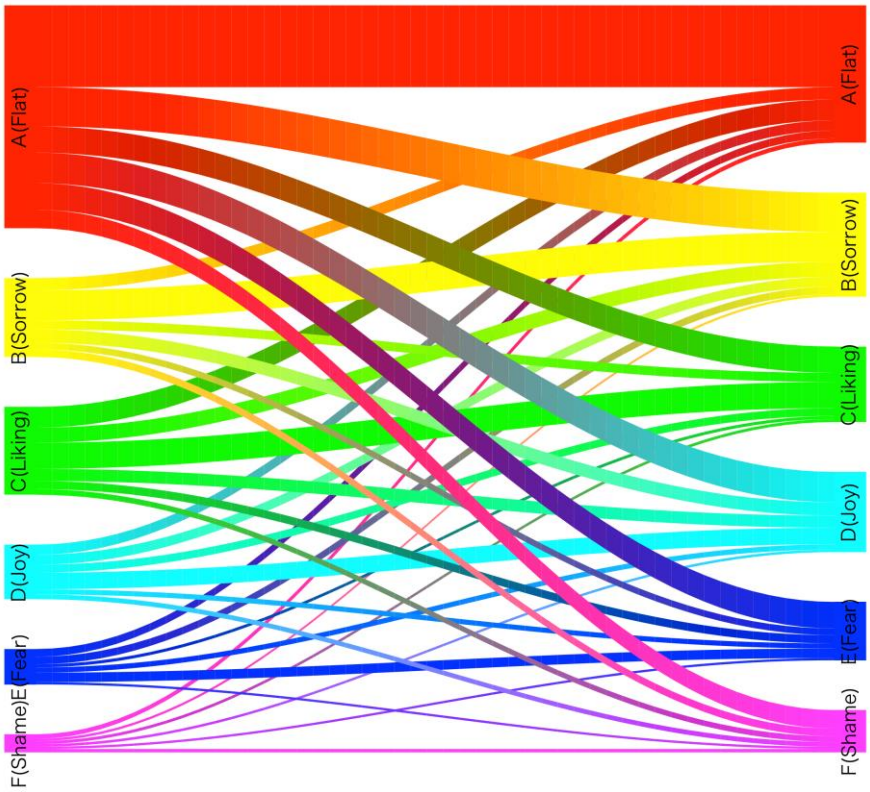

Figure 2: Emotional transitions

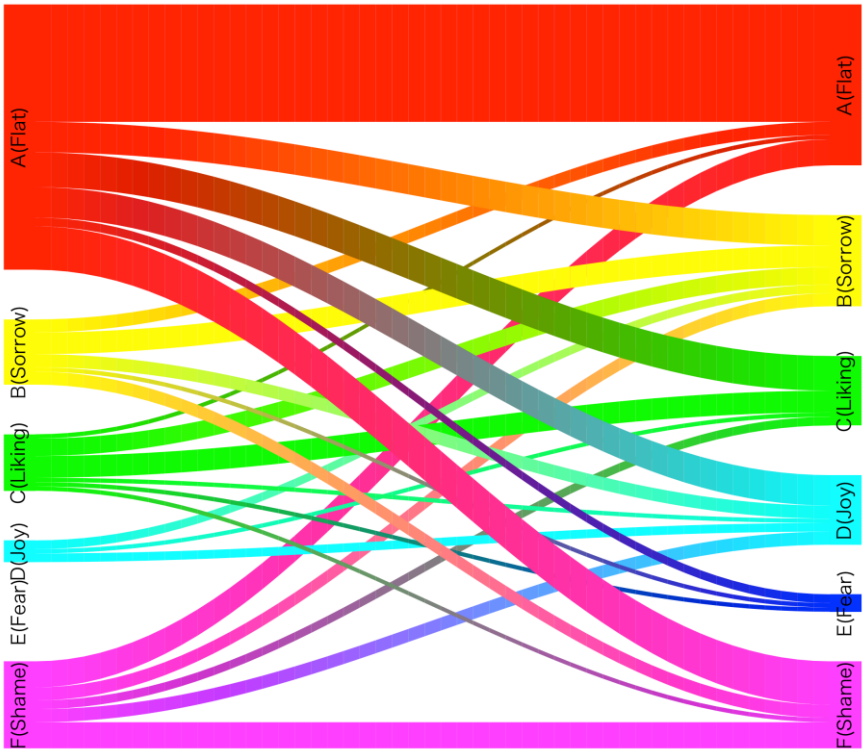

Figure 3: Emotional transitions of Dazai works

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

6 Conclusion

In this paper, we extracted emotions from novels and performed clustering to classify the patterns in which emotions appear in novels. We also clarified how emotions transition from the preceding parts of the novel to the ending parts. Detailed evaluation of emotion analysis methods and the automatic generation of novels based on this analysis are future challenges.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number JP24K06300.

References

[1] Stable Diffusion, https://stablediffusionweb.com

[2] DALL-E2, https://openai.com/product/dall-e-2

[3] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, Sequence to Sequence Learning with Neural Networks, Advances in Neural Information Processing Systems 27 (NIPS 2014), 2014.

[4] ChatGPT, https://openai.com/blog/chatgpt

[5] Hoshi Shinichi Awards, https://hoshiaward.nikkei.co.jp

[6] Vladimir Propp, Morphology of the FolkTale, trans. Laurence Scott, revised Louis A. Wagner. Austin, University of Texas Press, 1968.

[7] Murai, Hajime, Plot analysis for describing punch line functions in Shinichi Hoshi's microfiction, OpenAccess Series in Informatics, 41, pp.121-129, 2014.

[8] Furukawa, M. and Yaginuma, Y., Detection and Clustering of Ending Parts of Novels, 14th IIAI International Congress on Advanced Applied Informatics, pp. 738-739, 2023.7.

[9] Mohammad S., From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales, Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pp.105–114, 2011.

[10] Nanyun Peng, Marjan Ghazvininejad, Jonathan May, Kevin Knight, Towards Controllable Story Generation, Proceedings of the First Workshop on Storytelling, pp.43-49, 2018.

[11] Reagan, A.J., Mitchell, L., Kiley, D. et al., The emotional arcs of stories are dominated by six basic shapes, EPJ Data Sci., 5, 31, 2016.

[12] Aozora Bunko, https://www.aozora.gr.jp

[13] pymlask, https://github.com/ikegami-yukino/pymlask

[14] ML-Ask, http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/repository/mlask.htm

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.