

# A Comparative Study to Understanding about Poetics Based on Natural Language Processing

Lingyi Zhang<sup>1</sup>, Junhui Gao<sup>2</sup>

<sup>1</sup>Wuxi No. 1 High School, Wuxi, China

<sup>2</sup>American and European International Study Center, Wuxi, China

Email: jhgao68@163.com

**How to cite this paper:** Zhang, L. Y., & Gao, J. H. (2017). A Comparative Study to Understanding about Poetics Based on Natural Language Processing. *Open Journal of Modern Linguistics*, 7, 229-237. <https://doi.org/10.4236/ojml.2017.75017>

**Received:** August 27, 2017

**Accepted:** October 6, 2017

**Published:** October 9, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

This paper tries to find out five poets' (Thomas Hardy, Wilde, Browning, Yeats, and Tagore) differences and similarities through analyzing their works on nineteenth Century by using natural language understanding technology and word vector model. Firstly, we collect enough poems from these five poets, build five corpus respectively, and calculate their high-frequency words, by using Natural Language Processing method. Then, based on the word vector model, we calculate the word vectors of the five poets' high-frequency words, and combine the word vectors of each poet into one vector. Finally, we analyze the similarity between the combined word vectors by using the hierarchical clustering method. The result shows that the poems of Hardy, Browning, and Wilde are similar; the poems of Tagore and Yeats are relatively close—but the gap between the two is relatively large. In addition, we evaluate the stability of our approach by altering the word vector dimension, and try to analyze the results of clustering in a literary (poetic) perspective. Yeats and Tagore possessed a kind of mysticism poetics thought, while Hardy, Browning, and Wilde have the elements of realism combined with tragedy and comedy. The results are similar comparing to those we get from the word vector model.

## Keywords

Poets, Natural Language Processing, Word Vector Model, Similarity, Cluster Analysis

---

## 1. Introduction

Deep Learning is a new field in machine learning, a learning method based on the representation of data. The concept is derived from the study of artificial

neural networks. By combining low-level features to form a more abstract high-level representation of attributes, categories, or features, the aim is to discover the distribution of data.

The earliest neural network in deep learning originated from the MCP artificial neuron model in 1943 (Bryant, 2016), which was used to simulate human neuronal responses by computers at that time. In 1958, Rosenblatt invented the perceptron algorithm that used MCP for machine learning (Rhys, 2017).

The deep learning in natural language began in 2006 when Hinton proposed the concept of Deep Belief Network (DBN) (Imagination Tech, 2017). Previously, the neural network was a complex one that was difficult to train, and only studied as a mathematical theory. In addition, Word vector model is the most common model used in natural language deep learning process. The core idea of this model is to symbolize the language into 1 and 0, a mode that is suitable for machine learning.

Andrew L et al. used a probabilistic model of documents, which learns semantically focused word vectors, to learn the word representations to encode word meaning—semantics (Maas, Andrew, & Ng, 2011).

Mikolov et al. proposed two new model structures for computing continuous vector representations of words from very large data sets to measure the similarity between syntactic and semantic words, and the results are compared to the previously techniques based on different types of neural networks (Mikolov, Chen, Corrado, & Dean, 2013).

Attabi et al. studied the effectiveness of anchor models to solve multiple emotion recognition problems from speech, based on the FAU AIBO Emotion Corpus—a database of spontaneous children’s speech. Compared with generative model such as the Gaussian Mixture Models, the anchor models improve significantly the performance of GMMs by 6.2 percent relative in such problems (Attabi & Dumouchel, 2013).

Sreeja et al. discussed the automatic recognition of emotions in English poems, which included Love, Sad, Anger, Hate, Fear, Surprise, Courage, Joy and Peace, by using the Vector Space model with a total of 348 poems of 163 poets mined from the web (Sreeja & Mahalakshmi, 2016).

Zhou Yingying et al. conducted experiments on the Chinese Quora—Zhihu—by using the topic2vec vector model in Chinese corpora. They found out that the convolutional neural network (CNN) with topic2vec gained an accuracy of 98.06% for long content texts, 93.27% for short time texts and an improvement comparing with other word embedding models (Zhou & Fan, 2016).

According to a series of previous study in deep learning of natural language, we can find that some have studied the syntax and semantics of text on the basis of word vector models. Some, based on the study of their predecessors, compared the efficiency of different models applied to the similar task. Others did detailed research such as using plenty of poems as corpora to carry out emotion recognition. Based on the study above, we will use the traditional word vector

model for comparative poetics study.

## 2. Materials and Method

We will describe them from data, word vector calculations, and comparative approaches among poets in the following content.

### 2.1. Materials

Four of the five selected poets are from England, including Thomas Hardy, Wilde, Browning, and Yeats. The one left is Tagore, a poet from India. We selected a total of 257 poems from Thomas Hardy (Poemhunter, 2017), 96 poems from Oscar Wilde (Poemhunter, 2017), 63 poems from Browning (Blackcatpoems, 2017), nearly 400 poems from (Yeats, 1951; Blake, 2002), and 86 poems from Tagore (Tagore, 2011).

The main reason for selecting the five poems is to avoid the errors caused by all sorts of differences. Firstly, the origin version of poems of their works are all in English. In this way, we do not need to translate their works in which we get the second-hand poems containing the translation errors in order to get accuracy results from analysis. Secondly, the gaps between their living years are very small since nearly all of their works are produced in early nineteenth Century to mid twentieth Century, which was the golden years of the development of European poetry. Thus, the problems which may be caused by the differences between archaic words and modern words can be effectively avoid. For example, in old English, poets used “thou” in lieu of “you” to express you’s nominative form and “thee” in lieu of “you” to express you’s accusative form. The five poets who all gathered during nineteenth Century and twentieth Century almost eliminate the use of old English, although some old words may also appear in their poems rarely. In other words, we will not choose to compare Beowulf with Mark Twain’s The Million Pound Note because they do not belong to different language systems at various times.

### 2.2. Word Vector Calculation

Although the research of natural language has already existed, traditional natural language study is a basic bottom-up study, from words, sentences, and paragraphs, and finally to the structures of text, but still can not let the computers understand the natural language well. One of the obstacles is the poor understanding of semantics. Before word2vec occurred, the research of semantic in NLP was mainly based on the understanding of latent semantic (LSA, Latent Semantic Analysis), and then its subsequent model (topic model) was introduced (Niketim, 2016).

Word2vec and topic models are completely different things. In the topic model, the basic granularity is still the word, and the topic is a probabilistic combination of words.

The semantics mined from the topic model of the article is at high level. In

word2vec, however, the word “fundamental granularity” has a new expression, which is called the word vector (word embedding).

Before the occurrence of word vector, we often used the method called 1-of-N (or one-hot). In this representation, the great majority of elements is 0, and only one dimension is 1. This dimension represents the current word.

Suppose that we have five words in our table: King, Queen, Man, Woman, Child. If we want to represent ‘Queen’, we can express it in 1-of N, as shown in **Table 1**.

This simple method has two drawbacks. One is the curse of dimensionality. Another is a phenomenon called “lexical gap”, namely the isolation between any two words, and is unable to judge a synonym like “microphone” and “Mike”.

The new method of word representation is called Distributed Representation. This method in representing word uses the position of a real vector to represent a word such as [0.792, -0.177, -0.107, 0.109, -0.542, ...], as shown in **Table 2**.

For each poet, we combine all the poems we collected, and construct the corpus by NLTK. Then, the corresponding word vectors are generated by Word2vec.

Natural Language Toolkit, referred to as NLTK, is a Natural Language Processing kit and a often used Python library in NLP, which was developed by Steven Bird and Edward Loper in the information science department at University of Pennsylvania (Baïke, 2017).

### 2.3. Comparative Approaches among Poets

For each poet, we find the common high-frequency words of him and other poets, and assume that each high-frequency word is a 100 dimensional vector, and finally combine all the vectors into one corresponding to the high-frequency words.

Then, we calculate the distance between the five vectors by cosine method. The cosine similarity is derived by the cosine value of the angle between the two vectors in the vector space to measure the difference between the two individu-

**Table 1.** Expression in 1-of-N.

0	1	0	0	0
King	Queen	Man	Woman	Child

**Table 2.** Distributed Representation.

	King	Queen	Woman	Princess	...
Royaling	0.99	0.99	0.02	0.98	...
Mascalining	0.99	0.05	0.01	0.02	...
Feminin	0.05	0.93	0.999	0.94	...
Age	0.7	0.6	0.5	0.1	...
...	...	...	...	...	...

als. The closer the cosine is to one, the more the angle is closer to zero degrees, namely the close resemblance between the two vectors. This is called “cosine similarity” (Yuhushangwei, 2016).

After we get the distance between the five poets, the value is subtracted by 1, and we consider this value as the similarity between the five poets. Afterwards, we employ cluster analysis to analyze the relationship between the five poets.

The difference between clustering and classification is that the classes divided by clustering are unknown. Clustering is a process that classifies data into different classes or clusters, so the objects in the same cluster have great similarity, while objects between different clusters have great diversity. From the point of view of statistics, clustering analysis is a way to simplify data through data modeling.

There are many kinds of clustering methods, and here we use hierarchical clustering. This method decomposes the given data set as a hierarchical level until reaching a certain condition. Concretely, it can be divided into two programs: condensed and split. Hierarchical agglomerative cluster is a bottom-up strategy. Firstly, take each object as a cluster, and then combine these clusters into bigger clusters until all the objects are in one cluster, or a certain condition is reached. The great majority of the hierarchical clustering method belongs to this class, and only the definitions of the similarity between clusters are different. Split level clustering is opposite to hierarchical agglomerative cluster, by using strategy of top-down. It will first put all the objects into one cluster, and then gradually subdivide them into smaller clusters until each object forms a cluster, or a certain condition is reached.

### 3. Results

We will show our results from three aspects: statistics of high-frequency word, similarity calculation, and cluster analysis.

#### 3.1. Statistics of High-Frequency Word

The statistics of the high-frequency words of the five poets are shown in **Table 3**. This table is arranged from left to right, and from top to bottom. The word in the upper left corner has the highest number of occurrence, which is 1225; The word in the lower right corner has the minimum number of occurrence, which is 392.

#### 3.2. Similarity Calculation

We set the word vector dimension to 100, then calculate the word vector, and finally compare the similarity between the five poets, as shown in **Table 4**.

#### 3.3. Cluster Analysis

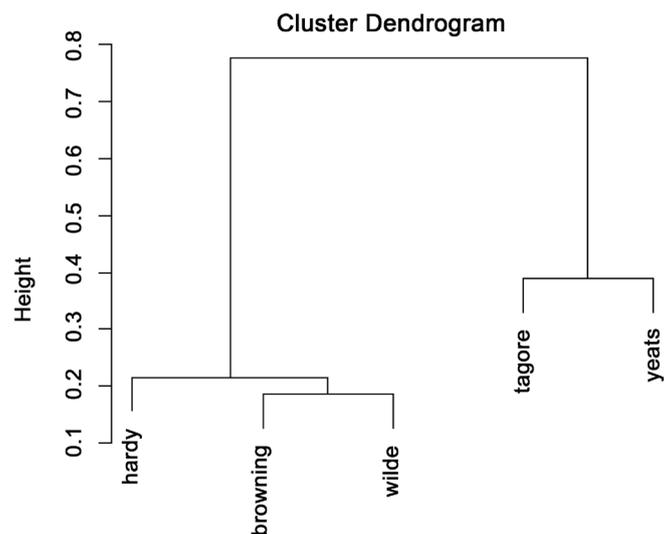
Based on **Table 4**, we use hierarchical clustering, and the results are shown in **Figure 1**.

**Table 3.** Public High-Frequency Words (first 20).

word	times	word	times	word	times	word	times
one	1225	come	715	know	530	shall	432
would	999	day	661	time	504	upon	420
like	941	life	623	king	480	never	416
said	821	man	584	old	442	night	399
heart	786	love	531	could	441	let	392

**Table 4.** Similarity between the Five Poets.

	browning	hardy	tagore	wilde	yeats
browning	1.00	0.79	0.22	0.81	0.26
hardy	0.79	1.00	0.54	0.78	0.54
tagore	0.22	0.54	1.00	0.34	0.61
wilde	0.81	0.78	0.34	1.00	0.40
yeats	0.26	0.54	0.61	0.40	1.00

**Figure 1.** A Hierarchical Clustering Map of Five Poets by a 100 Dimensional Vector Model.

In **Figure 1**, the abscissa is five poets. The ordinate is the distance between those poets. The shorter the distance between the poets, the higher the similarity. From **Table 1**, Hardy, Browning, and Wilde are similar, with the difference of about 0.2, especially the latter two. Tagore and Yeats are close to each other, with the difference of about 0.4, not as close as the first three poets. However, the difference between the group of Hardy, Browning and Wilde and the group of Tagore and Yeats is large, with the value between 0.7 and 0.8 (the largest difference is 1).

#### 4. Discussion

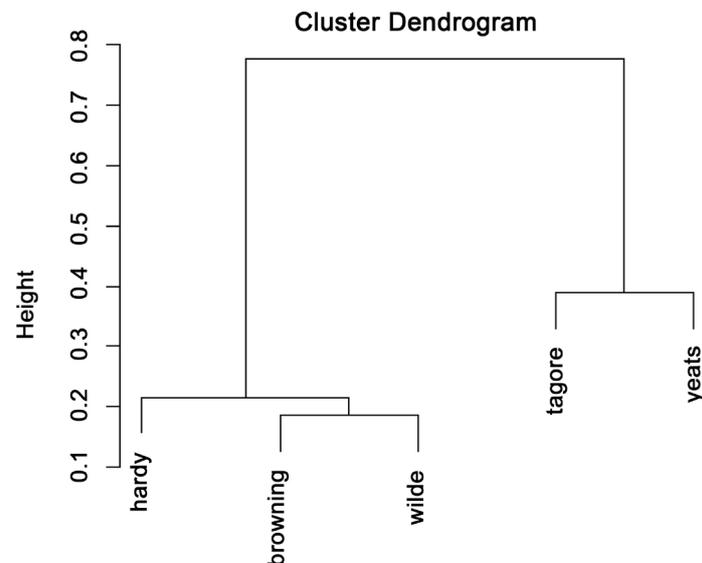
As mentioned earlier, we talked about the definition of 100 dimensional compu-

tational vector of word, and obtained the results in **Tables 1-4**. In order to test the stability of the results, we also use 80 dimension and 120 dimension to calculate the word vector, and the result we get from the calculation is very close to that of 100 dimension. Take 120 dimension as an example. The clustering result we obtain is shown in **Figure 2**. The results of **Figure 1** and **Figure 2** are very close to each other, indicating that our method is stable and reliable.

From a literary perspective, Tagore is a patriotic poet, and his works reveal his patriotism and the spirit of Democracy. Yeats showed the reverence to Aestheticism and Romanticism in his early years. After he experienced the nationalist political movement in Ireland in his forties, the style of his poetry gradually went close to realism.

Tagore and Yeats developed their friendship because of poetry. They shared many points of view in literature. First of all, Tagore and Yeats had direct contacts in life. In 1912, they met each other due to “Gitanjali”. Yeats admired Tagore’s talent very much, and helped Tagore publish this collection and made the preface of it. Second, both of them possessed a kind of mysticism poetics thought. Tagore’s belief is a mixture of religious philosophy while Yeat’s belief is derived from his natural disposition, which is personal philosophy. Third, although they are modern poets, they do not belong to Modernism since both of them criticize the modernist literature in their poems. Therefore, the results we obtained from literary appreciation are similar to those gained from the cluster analysis above (Wang, 2012).

Wilde is one of the representative poets of aestheticism, with fairy tales as the main characteristic. His poems are full of the elements of duality, which shows the simultaneism of aesthetics and tragedy. Wilde is good at describing the contradiction between characters and the cruel social background. His tragic beauty and death consciousness contain his understanding about life (Sun, 2012).



**Figure 2.** A Hierarchical Clustering Map of Five Poets by a 100 Dimensional Vector Model.

Likewise, the poems of Thomas Hardy also have tragic color, which is mostly the natural revelation of personal experience and emotion. Hardy believes that society is the root of pain; the personality of human beings leads to the suffering in the world; and the destiny is controlled by the universe. The analysis of these unique perspectives illustrates the ubiquitous tragedy and distress in his poems (Ma, 2009). Robert Browning is a British poet and a playwright. He creates a unique form of poetry, referred to as “dramatic monologue”, using a cinematic narrative technique—Montage—to restructure and integrate time and space. Browning loves to show the changes in characters’ psychological and story scenes through personal confession. Owning the color of the mixture of tragedy and comedy, His poems express the complexities of the characters and their attitudes of life. To sum up, although the styles of the three poets belong to different genres, all of them do well in depicting tragedies, and showing the irreconcilable contradictions between man and society (Zhang, 2007). Thus, the results we obtained from literature perspective are similar to those gained from the cluster analysis above.

The main contribution of our work is that this research is the first work to study different poets’ works by using the word vector model, which is pioneering and original. The drawback is that the number of the poets we used is limited. Also, the poet’s geographical distribution was not uniform enough since of the five poets, four of them came from England, and one left came from India. Finally, the dimensions we used are limited that we only employed 80, 100, and 120 the three dimensions to calculate their difference, but larger ones have not been used.

## 5. Conclusion

This paper uses vector model and hierarchical clustering in deep learning to investigate the similarities between the works of the five poets—Thomas Hardy, Oscar Wilde, Robert Browning, William Yeats, and Rabindranath Tagore—in the nineteenth Century. Our research contributes to the field which combines mathematical analysis and literary analysis together. High frequency words picked from the five poets are analyzed by the word vector model in 100 dimensions. The results show that the poems of Hardy, Browning, and Wilde are similar; the poems of Tagore and Yeats are relatively close. We also have employed other dimensions such as 80 and 120 to test the stability of our results, which have been proved reliable then. In addition, we have obtained the similar results by analyzing the works of the poets from a literary perspective which indicate their similarity in the interpretation of the tragedy, and the conflicts between men and the society.

## References

- Attabi, Y., & Dumouchel, P. (2013). Anchor Models for Emotion Recognition from Speech. *IEEE Transactions on Affective Computing, 4*, 1-11.  
<https://doi.org/10.1109/T-AFFC.2013.17>

- Baike (2017). *Natural Language Toolkit*.  
<https://baike.baidu.com/item/NLTK/20403245?fr=aladdin>
- Blackcatpoems (2017). *Robert Browning*.  
[http://www.blackcatpoems.com/b/robert\\_browning.html](http://www.blackcatpoems.com/b/robert_browning.html)
- Bryant, L. J. (2016). The History of Deep Learning. *CSDN Blog*.  
<http://blog.csdn.net/u012177034/article/details/52252851>
- Imagination Tech (2017). *The History and Problems of Deep Learning in Natural Language*. [http://www.sohu.com/a/161325083\\_468740](http://www.sohu.com/a/161325083_468740)
- Ma, L. (2009). *On the Topics of Tragedy, Love & Marriage, and Christianity in Thomas Hardy's Novels and Poetry*. M. Thesis in Aesthetics, Tianjing Normal University, 8-11.
- Maas, A. L., & Ng, A. Y. (2011). *A Probabilistic Model for Semantic Word Vectors*. 1-8.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *Computer Science*, 1-12.
- Niketim (2016). The Introduction of Word Vector. *CSDN Blog*.  
<http://blog.csdn.net/u013362975/article/details/53319002>
- Poemhunter (2017). *Oscar Wilde Poems*. <https://www.poemhunter.com/oscar-wilde>
- Poemhunter (2017). *Thomas Hardy Poems*. <https://www.poemhunter.com/thomas-hardy>
- Rhys (2017). *Rosenblatt's Perceptron Algorithm*.  
[http://blog.sina.com.cn/s/blog\\_166b82f8b0102xcu3.html](http://blog.sina.com.cn/s/blog_166b82f8b0102xcu3.html)
- Sreeja, P. S., & Mahalakshmi, G. S. (2016). Comparison of Probabilistic Corpus Based Method and Vector Space Model for Emotion Recognition from Poems. *Asian Journal of Information Technology*, 15, 908-915.
- Sun, C. W. (2012). *Literature Review of Research on Wilde's Works in the Past Thirty Years* (pp. 1-4). Shenyang: Liaoning University.
- Tagore, R. (2011). Gitanjali. *Annals of Neuroscience*, 18, 66.
- Wang, X. S. (2012). *The Comparison of Tagore and Yeats' Poetic Thoughts* (pp. 1-3). M.Sc. Thesis, Chongqing: Chongqing Southwest University.
- Yeats, W. B. (1951). The Collected Poems of W.B. Yeats. *Wordsworth Poetry Library*, 1, 118-134.
- Yuhushangwei (2016). *The Calculation Method and Application of Cosine Similarity*.  
<http://blog.csdn.net/yuhushangwei/article/details/48541891>
- Zhang, W. (2007). *On the Cinematic Narrative Feature of Robert Browning's Poetry* (pp. 5-7). M.Sc. Thesis, Hangzhou: Zhejiang University.
- Zhou Y. Y., & Fan, L. (2016). Deep Learning on Improved Word Embedding Model for Topic Classification. *Computer Science and Application*, 6, 629-637.  
<https://doi.org/10.12677/CSA.2016.611077>