



Optimization of Window Size for Calculating Semantic Coherence Within an Essay

Kshitiz Srivastava^a, Namrata Dhanda^b and Anurag Shrivastava^c

^a Ph.D. scholar, Dr. APJ Abdul Kalaam technical university, Lucknow, Uttar Pradesh, India

^b Professor, Department of computer science, Amity University, Lucknow, Uttar Pradesh, India

^c Professor, Department of Computer Science & Engineering, Babu Banarasi Das Northern India Institute of Technology, Lucknow, Uttar Pradesh, India
kshitiz.sri@gmail.com, ndhanda@lko.amity.edu, anuragbdlko@gmail.com

KEYWORDS

automated essay evaluation;
semantic attributes;
semantic coherence;
semantic mining

ABSTRACT

Over the last fifty years, as the field of automated essay evaluation has progressed, several ways have been offered. The three aspects of style, substance, and semantics are the primary focus of automated essay evaluation. The style and content attributes have received the most attention, while the semantics attribute has received less attention. A smaller fraction of the essay (window) is chosen to measure semantics, and the essay is broken into smaller portions using this window. The goal of this work is to determine an acceptable window size for measuring semantic coherence between different parts of the essay with more precision.

1. Introduction

In the world of computer science, automated essay assessment is always seen as a challenge. It is both a natural language processing application and a statistics classification challenge. There are numerous automatic essay evaluation systems that have been designed and implemented, Psalmerosi, F. H. (2019). Some are already on the market, while others are still in the works. Ke, Z., and V. Ng (2019). Thousands of pupils in the United States are scored each year using systems like Project essay grader, which was developed in 1966 and is currently a product of Measurement Inc. Systems like C-rater, on the other hand, were unable to influence the marketplace with their performance, Y. Attali, (2011).

Kshitiz Srivastava, Namrata Dhanda, Anurag Shrivastava

Optimization of Window Size for Calculating Semantic Coherence Within an Essay



C-rater is now a standard feature of many other grading systems. With advancements in the field of natural language processing, the need for essay grades that are closer to human scores has grown.

All these systems work on attributes, these attributes denote the different features of the essays, Chandrasekaran, D. et al., (2020). Attributes are broadly classified into three categories: style, content, and semantics of the essay, Khatavkar, V., and Kulkarni, P. (2019). These attributes are subdivided into smaller attributes for example lexical sophistication, Readability measures, Lexical diversity, mechanics, content, etc. These sub-attributes may be a group of other smaller attributes for example lexical sophistication may include all or few of the following sub-attributes: number of characters, number of words, number of long words, number of short words, etc. Zupanc, K., and Bosnic, Z. (2014).

All of the proposed systems have prioritized style and content features, with semantic attributes receiving less attention, Azmi, A. M. et. al. (2019). Measuring the semantic of an essay becomes a huge difficulty due to differences in writing style and content length. A few features were proposed in 2014, and these qualities were integrated as an important component of SAGE Zupanc, K., and Bosnić, Z. (2014), allowing us to quantify the semantic coherence between the portions of the essay. These attributes were able to measure semantic coherence by splitting the essay into smaller sections using various reduction techniques, then placing these portions as points in a high-density semantic space and measuring attributes using various methodologies. In one experiment, a window of around 25% of the essay's entire size was taken, and a smaller corpus was identified by moving the window by 10 words Zupanc, K., and Bosnic, Z. (2016). In our work, we are trying to identify the appropriate window size that can increase the overall performance of the system. For the experiment, we have taken data from the Kaggle website, A. Mellor, (2011).

2. Ethics

The field of Automated Essay Evaluation is carried out in different problem areas, A. Mellor, (2011). In the following subsection, we have discussed these problem areas.

2.1. Automated Essay Evaluation Systems and their Approaches

Project Essay Grader was the first automated essay evaluation system, proposed by Elis B Page M. D. Shermis and J. Burstein (2003). It was developed in the mid-1960s. According to Page, the system was a better solution to manual grading. Because of the technology of that time his project was not much accepted by the community, as the operational cost was too high. With the advancement in technology, the internet, and text processing software, development in the field of automated essay evaluation systems took a great pace. Many systems and techniques were proposed and implemented. A few of the mentions systems are *Intelligent essay assessor* Srivastava K., Dhanda N. & Shrivastava A. (2020), *Intellimetric* M. T. Schultz (2013), *Project essay grader*, and *E-Rater* J. Burstein, J. Tetreault, and N. Madnani (2013).

One of the major obstacles in achieving success in this area is the non-availability of any open-source automated essay evaluation system. Most of the systems are either commercially available (PEG, E-rater, Intellimetric, etc.) or under development. The only system whose compiled code along with source code was publicly available was LightSide, developed by Mayfield and Rose Mayfield, E., and Rosé, C. (2010). It was designed to do various these, in which essay grading was one of the major features.

The quality of an essay can be measured by focusing on three basic attribute styles, content, and semantics. Style attribute focuses on the way the essay is written spelling, punctuations, grammar, etc. The content attribute is based on a comparison of the essay with the pre-graded essay. This attribute is used to compares two essays and finding the similarities between them. The semantic attribute is used to verify the correctness of the essay, Chandrasekaran, D., and Mago, V. (2020), D. Higgins, J. Burstein, D. Marcu, and C. Gentile (2004), Darwish, S. M., and Mohamed, S. K. (2019).

Numerous methodologies are used to extract attributes from the essay. Latent semantic analysis is one of the most popular methodologies used. Other methods that are used are pattern matching, sentence similarity networks, Generalized Latent semantic analysis, n-gram approaches, etc. For correctness and consistency semantic networks, ontologies, fuzzy logic, open information extraction, etc. are used Zupanc, K., and Bosnic, Z. (2016); Ferreira-Mello et al. (2019); T. K. Landauer et al.,(1998).

Most of the system uses Machine learning algorithms to create the prediction models like regression modeling, cosine similarity, rule-based expert system, etc, Romero, C., and Ventura, S. (2020). The idea is to use a pre-graded essay to create a prediction model, T. Kakkonen, et al., (2008). This prediction model is then used to grade the rest of the essays, Peng, C., et al. (2020). The performance of such a model is calculated by the comparing score obtained with the actual score, Gao, Y. et al., (2019); P. W. Foltz, et al. (2013); P. W. Foltz, et al. (2007). A brief comparison of a few of the available systems is given in Table 1 Srivastava K., Dhanda N. & Shrivastava A. (2020).

Table 1. A comparison of Some AEE systems

System	Attribute	Prediction Model	Methodology
Autoscore	Style and content	Machine Learning	Statistical
BETSY	Style and content	Bayesian Network	Statistical
Bookette	Style and content	Neural Networks	NLP
CRASE	Style and content	Machine Learning	NLP
E-rater	Style and content	Linear regression	NLP
IEA	Content	Machine Learning	LSA, NLP
Intellimetric	Style and content	Multiple Math metical model	NLP
Lightside	Content	Machine Learning	Statistical
Markit	Content	Linear regression	NLP, PMT
PEG	Style	Multiple linear regression	Statistical
PS-ME	Style	Linear regression	NLP
SAGE	Semantics	Random Forest	OIE, NLP
SAGrader	Semantics	Rule-based expert System	FL, SN
SEAR	Style and content	Linear regression	Statistical

2.2. Semantic Coherence

Coherence is defined as the flow of information from one part of the essay to other parts Mimno, D., Wallach, H., Talley, E., Leenders, M., and McCallum, A. (2011). A highly coherent essay has a low coherence movement whereas a low coherent essay has high movement. The system proposed till now either measures the coherence using supervised or unsupervised approaches, Darwish, S. M., and Mohamed, S. K. (2019).

In unsupervised approaches system usually measures the repetition of words or repetition of phrases assuming that a highly coherent essay has high repeating words and phrases Zupanc, K., and Bosnic, Z. (2016). According to Foltz, P. W., Laham, D., and Landauer, T. K. (1999). a highly coherent essay contains high semantically related words and essay. Hearst Hearst, M. A. (1997) has proposed that an essay can be subdivided into smaller parts and can be used in identifying the repetition of semantically related words and phrases. The most commonly used approach is latent semantic analysis.

In a supervised approach, pre-graded essays are used, D. Higgins, et al., (2004); M. D. Shermis and B. Hamner (2013). According to the centric theory Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995) «The extent a discourse adheres to centering constraints, its coherence will increase and the inference load placed upon the hearer will decrease». This phenomenon helps in locating topic shifts in the essay, Goulart et al., (2018), Foltz, P. W. (2007).

2.3. Semantic Coherence Measurement

The first step towards semantic coherence measurement is to identify the attributes upon which we measure semantic coherence, Injadat, M. et. al. (2020), Janda, H. K. et. al. (2019). Around 72 different attributes (including all types of attributes: style, content, and semantics) are considered by different AES systems Zupanc, K., and Bosnić, Z. (2017), J. Burstein, (2010). We have considered four semantic attributes upon which we will be calculating semantic coherence, Muangkammuen, P., and Fukumoto, F. (2020), D. Higgins et al., (2004). These attributes are attribute-1 (average distance) [8], attribute-2 (Minimum distance), attribute-3 (Maximum distance), and attribute-4 (average nearest neighbor), Romero, C. and Ventura, S. (2020).

To calculate these attributes first data is cleaned by: lowering, stemming, stop-word removal and correcting spelling, Misuraca, M. et al., (2021). Then data obtained is divided into smaller datasets by choosing a window of certain length and moving it by certain words to obtain sequential overlapping parts, J. Burstein, et. al. (2010). For this example, our window consists of 25% of the overall size of data. This window is moved by 10 words to have the next overlapping part Zupanc, K., and Bosnić, Z. (2017).

Each dataset obtained is represented as point in semantic space by first calculating Tf-idf (statistical method). Tf-idf provides the weights to the words in a particular document, Janda, H. K. et al., (2019).

$$tf-idf(t, d) = tf(t, d) * idf(t)$$

This two-dimensional state is obtained with the help of principal component analysis [Zupanc, K. and Bosnic, Z. (2016)]. It is a process of obtaining the principal components and using obtained components to measure the change in the data. PCA may choose the first few principal components ignores the other principal components, Bhatt, R. et al. (2020). From these points, different attributes are measured, calculating the semantic relatedness of the essay Zupanc, K., and Bosnić, Z. (2017). Our experiment is to find the optimal window size that can be moved to obtained sequential overlapping parts, for improving the overall performance of the systems, Azmi et al., (2019).

3. Experiment and Analysis

To find optimal window size a series of the experiments were conducted over different datasets of essays. Each essay has experimented using different window sizes ranging from 1/2 to 1/10th size of

essay, Injadat, M., et al. (2020). Four attributes were selected for the calculation of the window size. These attributes are average distance, minimum distance, maximum distance, and average nearest neighbor Zupanc, K., and Bosnić, Z. (2017). Following are the effects of various window sizes over different attributes.

3.1. Attribute 1 (Average distance)

The average distance attribute helps in identifying semantic relatedness between the sequential parts of the essay. According to Foltz, P. W. et. al. (1999) high semantically related essays have smaller movements whereas less semantic related essays have greater movements between parts of essay. To measure average distance following formula is considered. It helps us understand how well an idea persists within the essay

$$\text{Distance} = \text{distance} + \sqrt{dx*dx+dy*dy}$$

$$\text{Average distance} = 2.0*\text{distance}/(n*(n-1))$$

Where n = total numbers of points in semantic space.

For the experimental purpose, the value of the average distance calculated is normalized and then rounded off to two places to get normalized results. The sample of the results is as follows (Table 2).

Table 2. Average window size

Essay/Window	Size 2	Size 3	Size 4	Size 5	Size 6	Size 7	Size 8	Size 9	Size 10
Essay 1	0.19	0.16	0.15	0.14	0.13	0.12	0.12	0.11	0.11
Essay 2	0.2	0.17	0.16	0.14	0.13	0.13	0.12	0.11	0.11
Essay 3	0.18	0.15	0.14	0.13	0.12	0.12	0.11	0.1	0.1
Essay 4	0.28	0.24	0.22	0.2	0.18	0.17	0.16	0.15	0.15
Essay 5	0.29	0.25	0.23	0.21	0.19	0.18	0.17	0.16	0.16
Essay 6	0.26	0.23	0.21	0.19	0.17	0.16	0.15	0.14	0.14
Essay 7	0.23	0.2	0.18	0.16	0.15	0.13	0.12	0.12	0.12
Essay 8	0.42	0.36	0.33	0.29	0.26	0.26	0.23	0.23	0.23
Essay 9	0.45	0.38	0.34	0.29	0.28	0.24	0.23	0.23	0.23
Essay 10	0.38	0.34	0.3	0.28	0.25	0.24	0.22	0.22	0.22
Essay 11	0.21	0.18	0.17	0.16	0.15	0.15	0.14	0.14	0.14
Essay 12	0.25	0.22	0.2	0.18	0.17	0.16	0.15	0.15	0.15
Essay 13	0.19	0.17	0.15	0.14	0.13	0.12	0.11	0.11	0.11
Essay 14	0.18	0.16	0.14	0.12	0.11	0.11	0.1	0.1	0.1

The highlighted result in above table shows that with decreasing window size the value of the result becomes stable. The difference between the value of window size 2 and window size 3 is greater in comparison to window sizes 8, 9, or 10. The difference between the values of window size 8, 9, or 10 is either minimum or null. It can be concluded from this experiment that lower window sizes have stable results.

3.2. Attribute 2 (Minimum Distance)

The minimum distance between the neighboring points shows how well the idea is transferred from one sentence to the next sentence. Minimum distance measures the distance between all the points in the semantic spaces and finds out the minimum movement between the various parts of the essay. This distance helps in understand how well the idea flows between the various parts of the essay. The minimum distance is calculated using the following formula

$$\text{Distance} = \text{distance} + \sqrt{dx^2 + dy^2}$$

if(min>dist):
min = dist

The value of the minimum distance is normalized and rounded off to get a normalized value. Sample results are as follows (Table 3).

Table 3. Minimum Distance

Essay/Window	Size 2	Size 3	Size 4	Size 5	Size 6	Size 7	Size 8	Size 9	Size 10
Essay 1	0.05	0.01	0.01	0.02	0	0.02	0.01	0.01	0.01
Essay 2	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0	0
Essay 3	0.02	0.01	0.01	0.01	0	0.01	0.01	0.01	0.01
Essay 4	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01
Essay 5	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Essay 6	0.09	0.04	0.05	0.04	0.04	0.01	0.03	0.03	0.03
Essay 7	0.06	0.02	0.02	0.01	0.02	0.01	0.01	0.03	0.03
Essay 8	0.02	0.03	0.02	0.04	0.03	0.03	0.03	0.03	0.03
Essay 9	0.05	0.07	0.03	0.03	0.02	0.03	0.03	0.01	0.01
Essay 10	0.06	0.03	0.01	0.03	0.02	0.02	0.01	0.01	0.01
Essay 11	0.06	0.02	0.01	0.03	0.02	0.03	0.02	0.02	0.02
Essay 12	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0	0
Essay 13	0.08	0.05	0.05	0.03	0.03	0.03	0.03	0.03	0.03
Essay 14	0.01	0.04	0.01	0	0	0	0.01	0.01	0.01
Essay 15	0.02	0.01	0.01	0	0.01	0.01	0.02	0.02	0.02
Essay 16	0.07	0.03	0.04	0.01	0.01	0.02	0.01	0.01	0.01
Essay 17	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Essay 18	0.05	0.07	0.03	0.05	0.06	0.04	0.04	0.04	0.04
Essay 19	0.02	0.01	0	0.01	0.01	0	0.01	0.01	0.01
Essay 20	0.06	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
Essay 21	0.06	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0
Essay 22	0.02	0	0.01	0	0.01	0.01	0.01	0.01	0.01
Essay 23	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Essay 24	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0
Essay 25	0.04	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Essay 26	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Essay 27	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

From the above table, it can be easily seen that the minimum distance works better with smaller window size (Highlighted cells). Because of window size and overlapping windows sometimes we obtain zero because the corpus obtained is highly identical or the value obtained is very small or negligible. The large variation is not attained as attained in the average distance because a smaller window size causes the result to be more accurate in terms of a fraction. It can be concluded that a lower window size works well with smaller window size.

3.3. Attribute 3 (Maximum Distance)

The maximum distance between the neighboring points shows the breadth of the discussed concept. Maximum distance measures the distance between all the points in the semantic spaces and finds out the maximum movement between the various parts of the essay. This distance helps in understanding the breadth of the discussed concept. The maximum distance is calculated using the following formula.

```
Distance = distance + sqrt(dx*dx+dy*dy)
if(max<dist):
    max=dist
```

The value of maximum distance is normalized and rounded off to get a normalized value. Sample results are as follows (Table 4).

Table 4. Maximum Distance

Essay/Window	Size 2	Size 3	Size 4	Size 5	Size 6	Size 7	Size 8	Size 9	Size 10
Essay 1	1.49	1.23	1.02	0.94	0.89	0.83	0.75	0.78	0.78
Essay 2	1.01	0.7	0.59	0.57	0.52	0.5	0.49	0.49	0.49
Essay 3	0.86	0.64	0.55	0.51	0.48	0.46	0.45	0.44	0.44
Essay 4	0.85	0.65	0.55	0.52	0.5	0.48	0.48	0.48	0.48
Essay 5	0.71	0.57	0.51	0.46	0.45	0.43	0.41	0.41	0.41
Essay 6	0.7	0.56	0.5	0.48	0.44	0.43	0.43	0.4	0.4
Essay 7	1.58	1.2	1.11	1.01	0.87	0.9	0.88	0.88	0.87
Essay 8	1.36	1.2	1.1	0.9	0.95	0.94	0.82	0.74	0.74
Essay 9	1.5	1.32	1.15	1.02	1.08	1.04	1.04	0.92	0.92
Essay 10	1.43	1.19	1.05	1	0.92	0.86	0.81	0.79	0.79
Essay 11	1.18	1.09	0.99	0.86	0.82	0.75	0.75	0.7	0.7
Essay 12	1.23	0.92	0.81	0.71	0.69	0.63	0.61	0.61	0.61
Essay 13	1.12	0.89	0.78	0.72	0.68	0.66	0.64	0.64	0.64
Essay 14	0.81	0.61	0.53	0.48	0.48	0.45	0.43	0.41	0.41
Essay 15	0.64	0.54	0.49	0.48	0.47	0.45	0.44	0.43	0.43
Essay 16	0.71	0.55	0.52	0.48	0.49	0.46	0.44	0.43	0.43
Essay 17	0.58	0.47	0.38	0.37	0.36	0.34	0.32	0.31	0.31
Essay 18	0.69	0.53	0.45	0.43	0.41	0.41	0.39	0.37	0.37
Essay 19	0.55	0.45	0.35	0.32	0.3	0.3	0.29	0.28	0.28
Essay 20	0.68	0.5	0.42	0.38	0.34	0.33	0.32	0.31	0.31
Essay 21	0.75	0.64	0.59	0.56	0.57	0.53	0.51	0.54	0.54
Essay 22	1.04	0.72	0.63	0.55	0.51	0.52	0.51	0.51	0.51
Essay 23	0.81	0.6	0.51	0.55	0.52	0.51	0.47	0.47	0.47
Essay 24	0.76	0.6	0.51	0.47	0.46	0.45	0.43	0.43	0.43
Essay 25	0.71	0.57	0.51	0.46	0.45	0.43	0.41	0.41	0.41
Essay 26	0.68	0.52	0.46	0.41	0.37	0.35	0.34	0.34	0.34
Essay 27	0.70	0.56	0.5	0.48	0.44	0.43	0.43	0.4	0.4

Similar to the experiment of minimum value, the maximum value also works better with smaller window sizes (Highlighted cells). The large variation is not obtained as obtained in the average distance because a smaller window size causes the result to be more accurate in terms of a fraction. It can be concluded that a lower window size also works well with maximum distance.

3.4. Attribute 4 (Average Nearest Neighbor)

Average nearest neighbor measures how fast an idea develops across an essay. Higher the value slower the idea developed across an idea and lowers the value faster the idea is developed across the essay. To calculate the average nearest neighbor firstly nearest neighbor of each point is calculated. Then the average of the value received is calculated. This result will tell us how fast the idea is developed across the essay [T. Kakkonen et. al. (2008)]. The following formula is used to calculate the average nearest neighbor, P. J. Clark and F. C. Evans, (1954).

$$\bar{d} = \frac{\sum_{i=1}^N d_i}{N}$$

Where,

d_i is nearest neighbor of every point

N is total number of features

The value of the Average nearest neighbor is normalized and rounded off to get a normalized value. Sample results are as follows (Table 5).

Table 5. Average Nearest Neighbor

Essay/Window	Size 2	Size 3	Size 4	Size 5	Size 6	Size 7	Size 8	Size 9	Size 10
Essay 1	3.86	3.48	2.08	1.94	1.72	1.05	0.07	0.07	0.07
Essay 2	1.51	1.18	0.95	0.87	0.7	0.53	0.49	0.64	0.57
Essay 3	0.68	0.76	0.72	0.78	0.65	0.55	0.64	0.58	0.81
Essay 4	0.98	0.88	0.95	0.86	0.84	0.64	0.53	0.57	0.72
Essay 5	0.67	0.59	0.56	0.5	0.64	0.5	0.54	0.5	0.61
Essay 6	0.78	0.64	0.58	0.54	0.58	0.63	0.65	0.7	0.44
Essay 7	0.75	0.57	0.57	0.56	0.59	0.46	0.44	0.34	0.65
Essay 8	0.52	0.6	0.46	0.5	0.37	0.4	0.43	0.42	0.38
Essay 9	0.55	0.5	0.47	0.45	0.55	0.65	0.71	0.61	0.65
Essay 10	0.51	0.47	0.52	0.45	0.51	0.5	0.39	0.43	0.37
Essay 11	0.30	0.27	0.27	0.27	0.24	0.22	0.18	0.21	0.21
Essay 12	0.32	0.28	0.28	0.26	0.28	0.25	0.27	0.26	0.39
Essay 13	0.36	0.35	0.34	0.34	0.36	0.31	0.26	0.26	0.36
Essay 14	5	3.67	2.8	1.68	1.98	1.27	2.06	2.06	0.14
Essay 15	3.25	2.61	2.11	3.57	1.25	1.54	0.33	0.21	0.21
Essay 16	2.27	1.42	1.35	1.43	1.45	0.93	0.52	0.79	0.74
Essay 17	2.79	2.71	1.57	0.69	0.63	1.56	0.81	0.18	0.18
Essay 18	0.28	0.26	0.35	0.33	0.28	0.33	0.33	0.4	0.48

(continued)

Table 5. Average Nearest Neighbor (continued)

Essay/Window	Size 2	Size 3	Size 4	Size 5	Size 6	Size 7	Size 8	Size 9	Size 10
Essay 19	0.34	0.26	0.22	0.19	0.23	0.21	0.2	0.24	0.26
Essay 20	0.33	0.29	0.25	0.27	0.23	0.23	0.3	0.32	0.38
Essay 21	0.28	0.27	0.24	0.23	0.23	0.24	0.3	0.26	0.28
Essay 22	0.35	0.25	0.22	0.24	0.2	0.22	0.17	0.2	0.18
Essay 23	0.33	0.25	0.25	0.23	0.24	0.21	0.24	0.21	0.22
Essay 24	0.29	0.21	0.23	0.2	0.26	0.29	0.31	0.29	0.37
Essay 25	0.29	0.24	0.23	0.2	0.21	0.18	0.28	0.3	0.32
Essay 26	2.57	2.14	2.57	1.8	1.35	1.21	0.88	0.56	0.56
Essay 27	3.70	4.12	3.23	2.44	1.26	0.43	0.43	0.01	0.01

The average nearest neighbor does not show any significant variation with changing window size (Highlighted cell). This means that no effect of varying window size has on an average nearest neighbor.

4. Conclusion and Future Scope

The findings of our experiment show that choosing window size for different qualities has a substantial impact. The value of Attribute 1 (average distance) becomes stable as the window size decreases. Similarly, with lower window sizes, the results for Attribute 2 (minimum distance) and Attribute 3 (maximum distance) are steady. However, when the window size is changed, the property 4 (average nearest neighbour) does not exhibit any significant fluctuation. The window size for four semantic qualities that quantify the coherence between the parts of the essay is proposed in this paper. For attribute 1, attribute 2, and attribute 3, the smaller window size displays the most accurate result. The result obtained is unaffected by the four window sizes used for attribute. To get a better outcome, use a window size that is less than 1/7th the size of the total essay.

In this experiment we have focused on the window size that will help to subdivide the essay into parts, these parts are represented as a point in semantic space from where we can measure different attributes. In our future work, we will focus on the movement of these windows and choosing some appropriate movement techniques to obtain better results for various attributes.

References

- A. Mellor, (2011). «Essay Length, Lexical Diversity and Automatic Essay Scoring», *Memoirs of the Osaka Institute of Technology*, vol. 55, no. 2, pp. 1–14.
- Azmi, A. M., Al-Jouie, M. F., and Hussain, M. (2019). AAEE–Automated evaluation of students’ essays in Arabic language. *Information Processing & Management*, 56(5), 1736-1752.
- Bhatt, R., Patel, M., Srivastava, G., and Mago, V. (2020). A Graph Based Approach to Automate Essay Evaluation. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 4379-4385.

- Chandrasekaran, D., and Mago, V. (2020). Evolution of Semantic Similarity--A Survey. arXiv preprint arXiv:2004.13820.
- D. Higgins, J. Burstein, D. Marcu, and C. Gentile, (2004). «Evaluating Multiple Aspects of Coherence in Student Essays», in Proceedings of HLT-NAACL, Boston, MA.
- Darwish, S. M., and Mohamed, S. K. (2019). Automated Essay Evaluation Based on Fusion of Fuzzy Ontology and Latent Semantic Analysis. In International Conference on Advanced Machine Learning Technologies and Applications, pp. 566-575.
- Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., and Romero, C. (2019). Text mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.
- Foltz, P. W. (2007). Discourse coherence and LSA. Handbook of latent semantic analysis, 167-184.
- Foltz, P. W., Laham, D., and Landauer, T. K. (1999). The intelligent essay assessor: Applications to educational technology. Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 1(2), 939-944.
- Gao, Y., Davies, P. M., and Passonneau, R. J. (2018). Automated content analysis: A case study of computer science student summaries. In Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications, pp. 264-272.
- Goulart, H. X., Tosi, M. D., Gonçalves, D. S., Maia, R. F., and Wachs-Lopes, G. A. (2018). Hybrid model for word prediction using naive bayes and latent information. arXiv preprint arXiv:1803.00985.
- Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse.
- Hearst, M. A. (1997). Text Tiling: Segmenting text into multi-paragraph subtopic passages. Computational linguistics, 23(1), 33-64.
- Injadat, M., Moubayed, A., Nassif, A. B., and Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. Knowledge-Based Systems, 200, 105992.
- Janda, H. K., Pawar, A., Du, S., and Mago, V. (2019). Syntactic, semantic and sentiment analysis: The joint effect on automated essay evaluation. IEEE Access, 7, 108486-108503.
- J. Burstein, J. Tetreault, and S. Andreyev, (2010). «Using Entity-Based Features to Model Coherence in Student Essays», in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, no. June. Los Angeles, California: Association for Computational Linguistics, pp. 681-684.
- J. Burstein, J. Tetreault, and N. Madnani (2013), «The E-rater Automated Essay Scoring System», in Handbook of Automated Essay Evaluation: Current Applications and New Directions, M. D. Shermis and J. Burstein, Eds. New York Routledge, 2013, ch. 4, pp. 55-67.
- Khatavkar, V., and Kulkarni, P. (2019). Trends in Document Analysis. In Data Management, Analytics and Innovation (pp. 249-262). Springer, Singapore.
- Ke, Z., and Ng, V. (2019). Automated essay scoring: a survey of the state of the art. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (pp. 6300-6308). AAAI Press.
- M. D. Shermis and J. Burstein (2003) «Introduction», in Automated essay scoring: A cross-disciplinary perspective, M. D. Shermis, and J. Burstein, Eds. Manwah, NJ: Lawrence Erlbaum Associates, 2003, pp. 13-16.
- M. D. Shermis and B. Hamner, (2013). «Contrasting State-of-the-Art Automated Scoring of Essays: Analysis», in Handbook of Automated Essay Evaluation: Current Applications and New Directions, M. D. Shermis and J. Burstein, Eds. New York: Routledge, ch. 19, pp. 313-346.

- M. T. Schultz (2013) «The IntelliMetric Automated Essay Scoring Engine - A Review and an Application to Chinese Essay Scoring», in *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, M. D. Shermis and J. C. Burstein, Eds. New York: Routledge, 2013, ch. 6, pp. 89–98.
- Mayfield, E., and Rosé, C. (2010, June). An interactive tool for supporting error analysis for text mining. In *Proceedings of the NAACL HLT 2010 Demonstration Session* (pp. 25-28).
- Mimno, D., Wallach, H., Talley, E., Leenders, M., and McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 262-272).
- Misuraca, M., Scepi, G., and Spano, M. (2021). Using Opinion Mining as an educational analytic: An integrated strategy for the analysis of students' feedback. *Studies in Educational Evaluation*, 68, 100979.
- Muangkammuen, P., and Fukumoto, F. (2020). Multi-task Learning for Automated Essay Scoring with Sentiment Analysis. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop*, pp. 116-123.
- P. J. Clark and F. C. Evans, (1954). «Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations», *Ecology*, vol. 35, no. 4, pp. 445–453.
- P. W. Foltz, L. A. Streeter, K. E. Lochbaum, and T. K. Landauer (2013). «Implementation and Applications of the Intelligent Essay Assessor», in *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, M. D. Shermis and J. Burstein, Eds. New York: Routledge, ch. 5, pp. 68–88.
- P. W. Foltz (2007). «Discourse Coherence and LSA», in *Handbook of Latent Semantic Analysis*, T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, Eds. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., ch. 9, pp. 167–184.
- Peng, C., Chen, Y., Kang, Z., Chen, C., and Cheng, Q. (2020). Robust principal component analysis: A factorization-based approach with linear complexity. *Information Sciences*, 513, 581-599.
- Psalmrosi, F. H. (2019). Applying Text Mining and Machine Learning to Build Methods for Automated Grading (Master's thesis, University of Twente).
- Romero, C., and Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v10.
- Srivastava K., Dhanda N. & Shrivastava A. (2020), An Analysis of Automated Essay Grading Systems, *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-8 Issue-6, March 2020.
- T. Kakkonen, N. Myller, E. Sutinen, and J. Timonen (2008). «Comparison of Dimension Reduction Methods for Automated Essay Grading», *Educational Technology & Society*, v. 11, pp. 275–288.
- T. K. Landauer, P. W. Foltz, and D. Laham (1998). «An introduction to latent semantic analysis», *Discourse Processes*, vol. 25, pp. 259–284.
- Y. Attali, (2011), «A Differential Word Use Measure for Content Analysis in Automated Essay Scoring», *ETS Research Report Series*, vol. 36.
- Zupanc, K., and Bosnić, Z. (2017). Automated essay evaluation with semantic analysis. *Knowledge-Based Systems*, 120, 118-132.

- Zupanc, K., and Bosnic, Z. (2014). Automated essay evaluation augmented with semantic coherence measures. In 2014 IEEE International Conference on Data Mining, IEEE, pp. 1133-1138.
- Zupanc, K., and Bosnic, Z. (2016). Advances in the field of automated essay evaluation. *Informatica*, 39(4).

