CBIE 2023
CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO
Uma escola para o futuro: Tecnologia e conectividade a serviço da educação

Essays' Coherence Analysis Via Entity Grid Approach

Jessica Oliveira Brito[1], Elias de Oliveira[1]

[1]Programa de Pós-Graduação em Informática – Universidade Federal do Espírito Santo (UFES) – 29.075-910 – Vitória – ES – Brasil

{jessica.brito,elias}@lcad.inf.ufes.br

Abstract. *Coherence analysis is a challenging task, especially when applied to many domains. This paper proposes a strategy that combines Machine Learning and Linguistics to analyze text coherence by understanding entity behavior. It introduces an algorithm that automatically annotates documents based on the Entity Grid Discourse Representation. We defined two datasets, one of academic papers' abstracts and the other of students' essays. The annotation technique identifies the influence of grammatical structures on coherence levels and offers a cost-effective solution for coherence analysis. To assess coherence levels Machine Learning methods were used, and the experimental results demonstrate an accuracy of* 88% *when assessing coherence in abstracts and* 74% *in essays.*

1. Introduction

Automatically evaluating the coherence of a document is a challenging task. Understanding the connection between sentences, the role of the topic throughout a document and the recall of the main topic in the discourse are essential to the quality of a text. More than utterances randomly grouped, a coherent text is well-written, easy to read, understandable, and presents meaning.

Writing an essay is a part of the National High School Exam - Exame Nacional do Ensino Médio (henceforth, ENEM). In the ENEM essay exam, five areas of competence are evaluated: C1) Domain of the standard norm of the Portuguese language; C2) Understanding the essay proposal; C3) Organization of information and analysis of text coherence; C4) Demonstration of knowledge of the language necessary for argumentation; C5) Elaboration of a proposed solution to the problems addressed, respecting human rights and considering the social-cultural diversities[1].

Considering that the essay is a vital part of the exam, and the exam is a mechanism to access higher education, the assessment of the essay, and consequently of its coherence is an integral part of the student learning process. Therefore, assessing essays is one of the most demanding activities for teachers. Thus, a strategy to automatically grade the level of coherence of a text can be beneficial. For this reason, we evaluate students' essays written following the specifications of the ENEM exam. Focusing mainly on the C3 competence, which concerns coherence [Oliveira et al. 2021].

Several tasks in Natural Language Processing (NLP) rely on coherence models, such as coherence scoring, text generation, summarization, machine translation, and evaluation of a human-produced text, especially, in educational applications.

[1]https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem

Many coherence models aim to define local coherence by analyzing sentence-to-sentence transitions [Barzilay and Lapata 2008], [Lapata et al. 2005], [Althaus et al. 2004]. Other models focus on global coherence, which investigates relations between bigger parts of the document, or a combination of both [Elsner et al. 2007], [Li and Hovy 2014].

The coherence model described in this paper was presented by [Barzilay and Lapata 2008], [Lapata et al. 2005] and is a syntactic framework that uses entity grids to assess the coherence of documents. This discourse representation is well suited for investigating entity transition patterns of coherent and non-coherent texts.

We show the application of this method to predict the coherence of texts written in English and Brazilian Portuguese. We carried out experiments to predict the coherence behavior of texts from two datasets with different domains. One dataset is composed of abstracts and artificially created non-coherent abstracts. The other is a collection of perfectly scored essays from ENEM and other publicly available students' essays.

We employ the automatic annotation model described here and use machine learning methods to assess the coherence level of a document. We obtained better results when assessing the abstracts dataset, an accuracy of 88%, most likely because of the text structure. For the essay dataset, we achieved an accuracy of 74%, which we believe can be improved with a refinement of the pre-processing step for the Brazilian Portuguese annotation.

The next sections are organized as follows. In Section 2, some related works are briefly reviewed. In Section 3, we described the datasets. In Section 4, the automatic annotation model used to evaluate the coherence in essays is presented and discussed. In Section 5, we relate how the experiments were performed and the yielded results. The conclusions are presented in Section 6.

2. Related Works

Understanding and modeling coherence have been goals of many researchers in the fields of Computational Linguistics and NLP. In this section, we review some works that investigate coherence, from conceptual frameworks to computational approaches, in order to estimate the coherence levels of a text.

In their work, Centering Theory [Grosz et al. 1995], the authors suggest that certain entities are more central than others and that the coherence of discourse is influenced by the words used to refer to the same entity. Their work defines a set of constraints that rule the behavior of pronouns and noun phrases throughout the text. In this way, adjacent sentences will potentially focus on the same entities in a locally coherent text. The concepts and insights from the Centering Theory have influenced subsequent research in local coherence modeling.

Numerous computational approaches have been developed to try and model local and/or global coherence. Barzilay and Lapata [Barzilay and Lapata 2008], in their work, Modeling Local Coherence: An Entity Grid Approach, propose a framework to represent and measure local coherence, motivated by Centering Theory [Grosz et al. 1995].

The authors propose an approach that transforms the text into a set of entity transition sequences represented through an entity grid. The work focuses on local coherence using sentence-to-sentence transitions to capture text relatedness. By using entity-based representation, the authors argue that the properties of coherent texts can be learned from a corpus without manual annotation or a predefined knowledge base.

Althaus et al. [Althaus et al. 2004], focuses on locally coherent discourses through the task of ordering clauses to maximize local coherence and equates it to an NP-Complete problem. They contribute to the enhancement of measuring coherence limited by size.

Focusing on global coherence, Barzilay and Lee [Barzilay and Lee 2004], presented an adaptation of Hiden Markov Models (HMM) to detect topic shifts in a text. In their work, they represented the topics as hidden states and sentences as observations. The authors applied their solution to information ordering and extractive summarization, confirming the hypothesis that, for specific domains, word distribution patterns strongly correlate with discourse patterns in a text.

Elsner et al. [Elsner et al. 2007] combined the entity-based and HMM-based models and demonstrated that these two models are complementary to each other in coherence assessment. They presented a unified model to assess discourse coherence through global and local approaches and validate their work by performing an ordering task, achieving comparable results with the literature.

Not choosing an approach that uses global or local coherence, Li and Hovy [Li and Hovy 2014] propose a neural network model based on distributed sentence representation for coherence assessment. They implemented recursive and recurrent neural networks and a window approach for the coherence model and, conducted experiments on the tasks of sentence ordering and readability assessment. Their proposed architecture highlighted the need for feature engineering and is able to learn sentence representations. Which can, to some extent, understand the coherent sentence structure.

Focusing on essays, for many languages, the traditional Automated Essay Scoring (AES) techniques rely mainly on feature engineering and statistical models ([Oliveira et al. 2021], [Alves and Oliveira 2019], [da Silva Junior 2021]). In recent years, approaches using Neural Networks (e.g., Recurrent Neural Networks, Convolutional Neural Networks) arose as an intriguing idea for the overall grading of an essay ([Logeswaran et al. 2018], [Shin and Gierl 2022]).

Specifically assessing the ENEM essay, the work by [Marinho et al. 2022] analyzed the performance of different approaches for AES using feature engineering, embeddings, and neural networks, applying different methods for each of the five competencies.

The work by [Oliveira et al. 2021] discusses the use of named entities to enhance the set of shallow features usually employed by many approaches to assess the coherence of an essay and to infer its possible grade.

Alves and Oliveira [Alves and Oliveira 2019] propose a statistical model to define the profiles of the teachers according to their evaluation behavior of essays written according to the ENEM's instructions. They investigate the reliability and concordance among evaluators. Since, in the ENEM exam an essay is evaluated by at least two people. They observe that competencies that denote a sharp discrepancy between the grades given by the evaluators are a sign of the need for training to realign the evaluators.

Moreover, [Brito et al. 2021] presents an architecture for essay evaluation in virtual learning environments, in which a student can submit an essay by writing it on a Learning Management System. The written essays will be sent to a teacher for evaluation and sent back to the student, with the score given by the teacher in a peer-blinded form. The model quantifies the reliability and concordance among the evaluators.

The following sections of this paper build upon these prior works and propose a technique for enhancing local coherence in natural language processing tasks. Based on the works of [Barzilay and Lapata 2008], [Lapata et al. 2005], we build an automatic annotation model that uses entity grids to help assess the coherence of documents through machine learning models. This discourse representation is well suited for investigating entity transition patterns of coherent and non-coherent texts. We applied this approach to texts written in English and Brazilian Portuguese. In special, we aim to assess the coherence level of ENEM essays, considering the impact of this task on the lives of students and teachers.

3. Datasets

We created two different datasets for the experiments as depicted in Table 1. The first dataset consists of 100 abstracts selected from thousands of academic papers abstracts from a Kaggle Dataset[2]. Since the sample contains abstracts from actual papers, they are considered coherent by design. To balance the dataset, we artificially generated 100 non-coherent abstracts. The non-coherent text is yielded through a library that employs a Markov Chain approach to generate somewhat realistic random text from actual words. Although the abstracts generated by this method appear to be authentic text, they are not coherent. Hence, the complete abstracts dataset has 200 documents, a hundred labeled coherent, and a hundred non-coherent.

For the second dataset, we selected essays written according to the instructions for the ENEM written in Brazilian Portuguese. We collected 100 essays from the website Brasil Escola[3], where teachers grade the essays sent by students according to a monthly prompt, and 100 compiled from students that aced this part of the exam, i.e., got a 1000 points grade[4]. In the ENEM essay exam, there are five areas of competence evaluated, we focus on "C3) Organization of information and analysis of text coherence" since we want to investigate the coherence of a document. In this dataset, there are 100 essays considered coherent, with a perfect grade, and 100 non-coherent according to the grade of C3. If the C3 grade is less than 120, the text is considered non-coherent, given that the grade ranges from 0 to 200 in steps of 40 points.

Table 1. Datasets description.

Dataset	Coherent	Non-coherent	Samples
*A*bstract	100	100	200
*E*ssays	100	100	200

[2]https://www.kaggle.com/datasets/shivanandmn/multilabel-classification-dataset
[3]https://vestibular.brasilescola.uol.com.br/
[4]https://github.com/RII-LCAD/redacoesnotamil

4. Automatic Annotation Model

The model proposed in this study introduces a strategy for document annotation using the Entity Grid Model, as initially introduced by Barzilay and Lapata [Barzilay and Lapata 2008], which effectively represents discourse. The Language Model employed is a pipeline package obtained from the spaCy library[5], trained for specific languages (here, the used languages are English and Brazilian Portuguese).

For both languages, we use the morphological identification of an entity represented by the NOUN and PROPN labels, using a POS (Part-Of-Speech) Tagging feature from spaCy, to identify and select the words that represent the entities in each sentence. POS Tagging consists of annotating a word with its corresponding grammatical categories in a sentence.

We defined a dictionary that associates the syntactic relation from the word in an arch, in a dependency tree to the specific label for the category to which the entity belongs. The possible categories are Subject - S, Object - O, Other - X, and, when the entity is not in the specific sentence, a dash (-) represents the gap. Thus, a sentence in each document can be represented by a collection of nouns labeled for the role it portrays in a sentence.

To analyze how the entity behaves in a document, we look at transitions of roles played by the entity on the entire document. Specifically, we use a length two transition, i.e., adjacent sentences. Therefore, the behavior for every entity in each document is represented as sequences of $\{$S, O, X, -$\}^n$. The tables below (Tables 2 and 3) show a text fragment and its entity grid.

We build an entity grid based on the entities and their syntactic role. Table 2 depicts 3 sentences, the words selected as entities, and their roles in each sentence. Table 3 describes the entity grid, with the numbered sentence, entity, and its role. Table 4 displays the frequency of each transition throughout the document. Table 5 exhibits the probability of each transition computed as a ratio of its frequency and the total number of transitions.

Table 2. Sample Annotation.

- $[Equality]_S$ of the usual $[definitions]_X$ of $[Brakke]_X$ flow'
- In 1978 $[Brakke]_S$ introduced the mean curvature $[flow]_O$ in the $[setting]_X$ of geometric measure $[theory]_X$
- There exist multiple $[variants]_O$ of the original $[definitions]_X$

Table 3 shows the entity grid for the sentences in the document described in Table 2, with the numbered sentence, the entity, and the entity role in the sentence. We can see in Table 3, that the red box defines the transition of the entity "Equality" from a Subject(S) to a dash(-), indicating that it is not present in the adjacent sentence.

Analyzing it vertically, we can see that this transition has a length equal to two and corresponds to the transition "S -". The possible transitions with length two are SS, SO, SX, S-, OS, OO, OX, O-, XS, XO, XX, X-, -S, -O, -X, - -. Focusing on the entity "Brakke", we can see in the green box the transition "X S", meaning that in the

[5] https://spacy.io/

Table 3. Sample Entity Grid.

	Equality	Definitions	Brakke	Flow	Setting	Theory	Variants
1	S	X	X	-	-	-	-
2	-	-	S	O	X	X	-
3	-	X	-	-	-	-	O

first sentence, “Brakke” appears as a “X (Other)” and in the second sentence as a “S (Subject)”, and “S -” in the blue box, since the entity is not present on the third sentence.

The document labeled *d1* was selected from the abstract’s dataset (see Section 3). The frequency of each transition throughout the document is computed and shown in Table 4. The first line, which represents *d1*, exhibits the frequency of each transition of the text fragment shown in Table 2 and 3. As we can see, *d1* has a frequency of two for the “S -” transition, the red and blue box from Table 3.

The documents labeled *d2*, *d3*, and *d4* represent other text fragments selected from the abstract’s dataset (see Section 3). The document labeled *d4* has only two sentences and, *d4+Cn* and *d4+Cp* represent variations of the *d4* document (more detail on *d4* and variations will be presented). Table 4, displays the frequency of the possible transitions between the entities for *d1*, *d2*, *d3*, *d4*, *d4+Cn*, and *d4+Cp*.

Table 5 demonstrates the probability of each transition computed as a ratio of its frequency and the total number of transitions. We calculate the total number of transitions as the product between the number of entities and the number of sentences minus one. Analyzing *d1*, we know that it has 7 entities and 3 sentences, as can be seen in Table 2 and Table 3. Accordingly, the total number of transitions for this document is $7*(3-1) = 14$.

Table 4. Sample Frequency Table.

Document	SS	SO	SX	S-	OS	OO	OX	O-	XS	XO	XX	X-	-S	-O	-X	- -
d1	0	0	0	2	0	0	0	1	1	0	0	3	0	2	3	2
d2	0	0	0	0	0	0	0	2	0	0	0	3	0	2	2	12
d3	0	0	0	3	0	0	0	3	0	0	0	3	4	3	2	15
d4	0	0	0	1	0	0	0	1	0	0	0	2	0	1	1	0
d4+Cn	0	0	0	1	0	0	1	0	0	0	1	3	0	1	1	4
d4+Cp	0	0	0	1	0	0	0	2	0	0	1	2	0	1	2	3

Having the probability of an entity grid, we can represent each document as a feature vector of size 16. Since there are 16 transitions, we need to reduce the dimensions to better visualize the documents on the vector space, and for that, we use the Principal Component Analysis (PCA) technique, with implementation from the Scikit-Learn library[6]. Therefore, we can represent the documents in a 2D space.

Figure 1 shows the visualization of the documents in a vector space, each represented as a feature vector based on the probability of the transitions. The document

[6]https://scikit-learn.org/

Table 5. Sample Probability Table.

Doc	SS	SO	SX	S-	OS	OO	OX	O-	XS	XO	XX	X-	-S	-O	-X	- -
d1	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.07	0.07	0.00	0.00	0.21	0.00	0.14	0.21	0.14
d2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.14	0.00	0.10	0.10	0.57
d3	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.09	0.12	0.09	0.06	0.45
d4	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.33	0.00	0.17	0.17	0.00
d4+Cn	0.00	0.00	0.00	0.08	0.00	0.00	0.08	0.00	0.00	0.00	0.08	0.25	0.00	0.08	0.08	0.33
d4+Cp	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.17	0.00	0.00	0.08	0.17	0.00	0.08	0.17	0.25

represented as *d1* (green), the fragment shown in Table 2, is the only one considered coherent. We can see that *d2* (red) and *d3* (purple) are close to each other and they both are considered non-coherent since they are samples from the artificially created abstracts (see Section 3). The highlighted areas in Figures 1, 2 and 3 indicate the coherent region.

The document labeled as *d4* (blue) behaves as an outlier most likely because the frequency of the "- - " transition is equal to zero, as can be seen in Table 5. Considering that the number of sentences in *d4* is 2 and the entity grid only shows an entity present in the document, is not possible for the "- -" transition to have a value different than zero, and that is why the "- -" transition of *d4* is skewed.

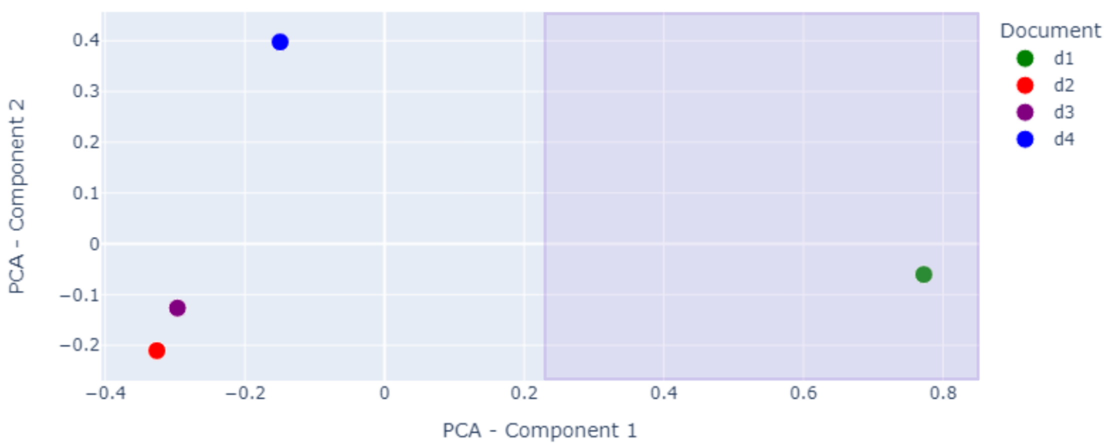

Figure 1. Visualization of the abstracts (real and artificial).

To further investigate this outlier, we altered the sample document *d4*. We created one more sentence that includes two instances of entities present throughout the document. For the document represented as *d4+Cn* (orange), a new sentence was added between the two sentences of the abstract, to connect the statements. As for *d4+Cp* (pink), the new sentence was added after the two sentences of the abstract, to complement it.

We can see in Figure 2, that both versions of *d4* are closer to *d1* (coherent) than *d4* (non-coherent). Since, by adding one more sentence we duplicate the number of transitions being evaluated at *d4*(number of entities $*$ number of sentences -1, e.g., from 6 to 12).

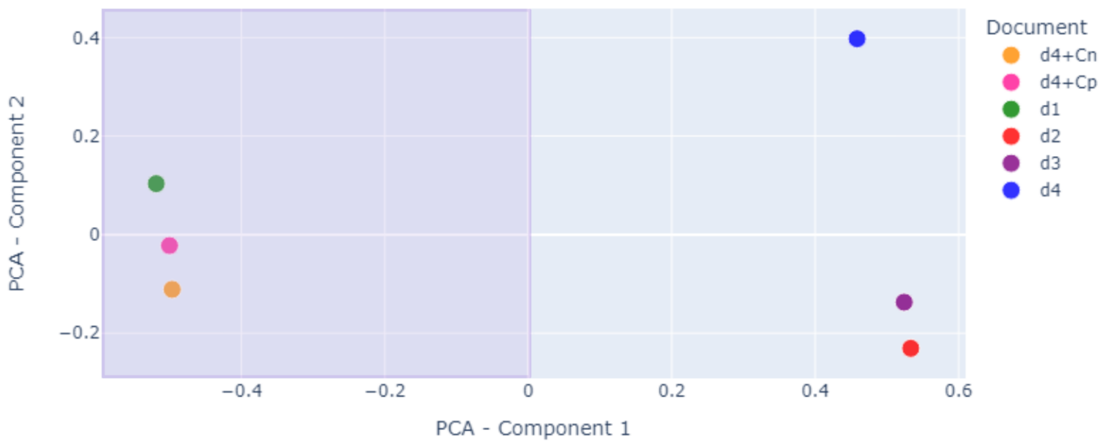

Figure 2. Visualization of the abstracts (real and artificial) and the versions of *d4*.

Based on the fact that *d1* is coherent, and assuming that the "- -" transition has a significant impact on the coherence level, we changed the value of this transition probability on the feature vector for *d4+Cp* (pink). Figure 3 shows that this simple alteration, going from 0.25 to 0.17, reduced even more the distance between *d1* and *d4+Cp*, suggesting that now *d4+Cp* is more coherent. This indicates that if an entity is absent across the sentences of the document, the coherence level will suffer substantially.

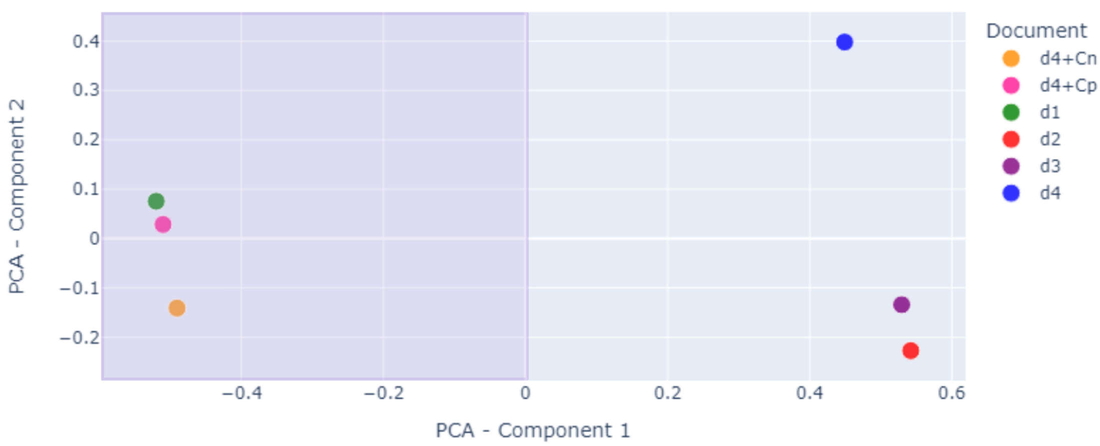

Figure 3. Visualization of the alteration made to *d4+Cp*.

The reoccurrence of an entity in a document alludes to a consistency of topic of discussion throughout the text. However, the simple fact that the entity appears in the text can steer the coherence classification of the document toward a more positive outcome.

The frequency of the transition "- -" in a document can also indicate coherence, or the lack thereof, resulting in a poor classification. Furthermore, entity coreference, i.e., when two or more words are used to refer to the same entity, can be a problem and skew the coherence levels of a document.

5. Experiments and Results

To test this approach, we designed an experiment to analyze samples of abstracts extracted from actual papers and randomly generated text and samples of real essays (see Section 3). For the abstract dataset, we trained using English, and for the essays dataset, we trained using Brazilian Portuguese.

We then applied the proposed model to annotate the documents, creating the entity grid, frequency table, and probability table. This process was carried out for the abstracts datasets and the essays datasets, separately. The coherence of a document is defined as one of two classes, labeled 0 for non-coherent and 1 for coherent. With the coherence labels defined in the dataset and the set of features created with the annotation, we trained and tested the model for each dataset.

We separated our datasets into train and test, with a proportion of 75% for training and 25% for testing, since our datasets for evaluation have a small number of samples, $n = 200$. This signifies that for the datasets with a sample size of 200, we trained with 150 documents and tested on 50 documents. The selection of samples was stratified by the classes that we were trying to predict, i.e., as our datasets were balanced so was the test split.

For the coherence evaluation, we used the set of supervised learning methods referred to as Support Vector Machines (SVMs). This set of techniques is applied to classification, regression, and outliers detection. And since they are considered effective in high dimensional spaces and memory efficient, we choose to use them for our 16 dimension vectors. To try and analyze our datasets, we employed Support Vector Regressor (SVR) and Support Vector Classification(SVC) methods from the Scikit-Learn library[7], both using a linear approach. The SVC is capable of performing class classification and was used here to analyze if a document (abstract or essay) is coherent or not. The SVR model was applied to try and comprehend the data and see how it would perform given that the prediction of coherence levels, not only if a document is coherent or not, is an essential part of grading an essay.

The goal of the experiment's first phase was to predict if a document was coherent using the SVR and SVC models and the PCA dimension reduction technique to verify if it could help improve the accuracy of the models even if we only have 16 features or give any insight on the most important features. The technique performed poorly, which means that the dimension reduction technique is making the model lose some information about the vectors.

Table 6 addresses the methods used to evaluate the abstract datasets and the essays dataset and the obtained results. We selected 100 samples, 50 coherent and 50 non-coherent from the abstracts dataset to test the PCA. The results of the experiment using the complete abstracts dataset and essay dataset, using SVR and SVC models, did not use the PCA technique (W/O). The emphasis on the table resides on the preferred method to execute the experiments onward since it shows the best accuracy.

[7]https://scikit-learn.org/

As anticipated, the SVC method presented a higher accuracy on both datasets in comparison with the SVR method. The accuracy results indicate that the automatic annotation model presented here is a suitable strategy for the analysis of coherence levels and can be applied to different domains. Particularly for essays, this annotation and prediction of coherence model can be used in order to assist the process of grading essays, accelerating it for the teachers and improving the students' performance.

Table 6. Accuracy of SVR and SVC models for the abstract dataset and the essay dataset with (W) or without (W/O) PCA technique.

Method	*D*ataset	Sample	PCA	Acc
LinearSVR	*A*bstracts - Random	100	W	66
LinearSVC			W	58
LinearSVR		200	W/O	62
LinearSVC			W/O	**88**
LinearSVR	*E*ssays		W/O	52
LinearSVC			W/O	**74**

6. Conclusion

In this paper, we presented an automatic annotation model based on the Entity Grid approach that represents the transition of entities from a text, indicating the level of coherence. Our experiments show that this annotation strategy is well suited for the prediction of coherence in a document.

The abstracts dataset and the essays dataset were annotated using this annotation model and evaluated using the SVR and SVC models, from the SVM scikit-learn library. The experimental results indicated that the presented annotation model and the SVC model produced the best overall performance with accuracy levels of, 88% and 74%, for the abstracts and essays datasets, respectively.

In section 4, we presented a study of the model using sample documents, in which the complete annotation process is described. It is noticeable that the improvements in the pre-processing phase yield better documents to be used as entries for the machine learning models. Therefore, an enhancement of this phase may lead to better results.

In future work, we intend to apply this strategy to analyze named entities produced by different Named Entity Recognition methods and compare the results with the automatic annotation model discussed here. Moreover, a complex problem that can occur during the annotation process is caused by coreference, when two or more words are used to refer to the same entity. We plan to mitigate it by using a coreference resolution tool that can act on the text written in Brazilian Portuguese. We also plan to work on the prediction of the level of coherence of a text according to the grades of the C3 competence.

Acknowledgments

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Financing Code 001

References

Althaus, E., Karamanis, N., and Koller, A. (2004). Computing locally coherent discourses. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 399–406.

Alves, J. and Oliveira, E. (2019). Avaliação de juízes: Um modelo estatístico para perfilação de avaliadores. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 199.

Barzilay, R. and Lapata, M. (2008). Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

Barzilay, R. and Lee, L. (2004). Catching the drift: Probabilistic content models, with applications to generation and summarization. *arXiv preprint cs/0405039*.

Brito, J., Alves, J., Badue, C., and Oliveira, E. (2021). An architecture for massive essays evaluations. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6. IEEE.

da Silva Junior, J. A. (2021). Um avaliador automático de redaçoes. *Master's thesis, Universidade Federal do Espírito Santo*.

Elsner, M., Austerweil, J., and Charniak, E. (2007). A unified local and global model for discourse coherence. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 436–443.

Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse.

Lapata, M., Barzilay, R., et al. (2005). Automatic evaluation of text coherence: Models and representations. In *Ijcai*, volume 5, pages 1085–1090.

Li, J. and Hovy, E. (2014). A model of coherence based on distributed sentence representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2039–2048.

Logeswaran, L., Lee, H., and Radev, D. (2018). Sentence ordering and coherence modeling using recurrent neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Marinho, J. C., Cordeiro, F., Anchiêta, R. T., and Moura, R. S. (2022). Automated essay scoring: An approach based on enem competencies. In *Anais do XIX Encontro Nacional de Inteligência Artificial e Computacional*, pages 49–60. SBC.

Oliveira, E., Alves, J., Brito, J., and Pirovani, J. (2021). The influence of ner on the essay grading. In *Intelligent Systems Design and Applications: 19th International Conference on Intelligent Systems Design and Applications (ISDA 2019) held December 3-5, 2019 19*, pages 162–171. Springer.

Shin, J. and Gierl, M. J. (2022). Evaluating coherence in writing: Comparing the capacity of automated essay scoring technologies. *Journal of Applied Testing Technology*, 23:04–20.