ASSESSING THEME ADHERENCE IN STUDENT THESIS

Tikhomirov M. M. (tikhomirov.mm@gmail.com), Loukachevitch N. V. (louk_nat@mail.ru), Dobrov B. V. (dobrov_bv@mail.ru)
Lomonosov Moscow State University, Moscow, Russia

In this paper we study approaches to assessing the quality of student theses in pedagogics. We consider a specific subtask in thesis scoring of estimating its adherence to the thesis’s theme. The special document (theme header) comprising the theme, aim, object, tasks of the thesis is formed. The theme adherence is calculated as the similarity value between the theme header and thesis segments. For evaluation we order theses in the increased value of the calculated theme adherence and compare the ordering with expert grades using the average precision measure. The best configuration for theses ranking is based on the weighted averaged sum of word embeddings (word2vec) and keywords extracted from the theme header.

Key words: Thesis assessment, embeddings, cosine similarity, ontology

ОЦЕНКА СООТВЕТСТВИЯ ТЕМЫ И ТЕКСТА В СТУДЕНЧЕСКОЙ ВЫПУСКНОЙ РАБОТЕ

Тихомиров М. М. (tikhomirov.mm@gmail.com), Лукашевич Н. В. (louk_nat@mail.ru), Добров Б. В. (dobrov_bv@mail.ru)
МГУ им. М. В. Ломоносова, Москва, Россия

В данной статье изучаются подходы к оценке качества студенческих выпускных работ по педагогике. Рассматривается подзадача определения соответствия текста теме работы. Для этого формируется специальный документ (тематический заголовок), содержащий тему, цель, объект, задачи работы. Соответствие теме рассчитывается как значение сходства между тематическим заголовком и сегментами работы. Для оценивания мы упорядочиваем работы по возрастанию значений соответствия теме и сравниваем полученный порядок с оценками экспертов, используя меру средней точности.

Ключевые слова: оценивание выпускных квалификационных работ, дистрибутивное представление слов, косинусная мера близости, онтология
1. Introduction

Currently, proper assessment of a student thesis (bachelor or magister) can be a very difficult task because of availability of various informational resources in Internet, which can be plagiarized by a student. It can seem that a thesis is well-done, but in fact the share of student own work is minimal. In order to detect borrowings in the thesis texts, so-called plagiarism detection systems (antiplagiat.ru, etxt.ru) have become widespread, which allow determining the percentage of borrowings either on the basis of their own source database or by analyzing the global search engine results (Yandex, Google) [7].

However, the quality requirements to a student thesis are not limited to plagiarism restrictions. The requirements include such important characteristics of a work as the theoretical and practical significance, elements of novelty in the work, knowledge of the modern literature on the research topic, consistency in the presentation of the material, the scientific style of presentation, and others. Checking these requirements could be automated in order to provide an expert with data on different characteristics of student works. The task of automated assessment can be compared to such a known task as automatic essay scoring [5], [6], [14], when student essays to specific topics should be assessed. But also, there are significant distinctions between thesis assessment and essay scoring tasks.

One of important characteristics of a student thesis is its relatedness to the thesis theme. The proclaimed theme is usually concretized in the following terms: aim of the study, object of the study, and the tasks of the work. It is possible to gather all these information into so-called theme header. It is usually supposed that a student should develop the theme and its details in the presented work. So, there is a subtask of thesis scoring to assess its adherence to the theme header. In the essay scoring, this subtask corresponds to the prompt relatedness subtask [6], [11].

In this paper, we study approaches to determining the relatedness between the theme header and student thesis in pedagogics. To evaluate the methods, we have the collection of 40 thousand student theses, 120 student theses among them have double expert scores. The aim of the assessment is as follows: if low relatedness is detected, then the problems should be visualized to experts and some penalties to the overall score for this work should be proposed by the system. We use several means for assessing relatedness including word embeddings and a thesaurus providing knowledge about domain term relations. As a thesaurus, we use Ontology on Natural Sciences and Technologies [4], where the pedagogics domain terms have been introduced.

We consider theme adherence as one of factors needed to be calculated for comprehensive assessment of student thesis. Also thesis fragments that found irrelevant to the thesis theme are considered as good candidates for plagiarism analysis.

2. Related Works

For assessing the quality of scientific papers, Osipov et al. [10] discuss such characteristics as the presence of the necessary sections (introduction, problem statement, list of references, etc.); scientific and non-scientific vocabularies; the presence
of logical and semantic defects in the text of a scientific publication; selecting author’s terms—new concepts defined by the authors of publications; highlighting the results presented in publications etc. Some authors study methods for the recognition of artificially generated scientific papers [2], [3], [8].

In the essay scoring, the most similar to our task is the task of prompt adherence that is assessing how the essay content corresponds to the announced essay topic [6], [11].

In [6] the Relatedness to Prompt feature is studied. The text of a essay fragment and the prompt (text of the essay question) must be related. If this relationship does not exist, this is perhaps evidence that the student has written an off-topic essay. The assessment was made for each sentence. The quality of the assessment was evaluated using double expert annotation for specific sentences. Most features proposed in this work are based on so-called Random indexing [13]. Random Indexing is a vector-based semantic representation system similar to Latent Semantic Analysis. In the current work, Random Indexing (RI) semantic space is trained on about 30 million words of newswire text. RI similarity to prompt for a sentence measures to what extent the sentence contains terms in the same semantic domain as compared to those found in the prompt. The SVM-classifier is trained on the calculated features and labeled data.

Persing and Ng [11] continue the study of the prompt relatedness in essay scoring using more diverse features. They try to predict the prompt relatedness for the whole essay, not for a single sentence. The predicted score ranges from one to four points at half-point intervals. 830 argumentative essays were annotated using a numerical score from one to four. Persing and Ng consider the task as a regression problem. Seven types of features were utilized in prompt-specific regressors based on linear SVM. Besides the random indexing features from the previous work, the authors used lemmatized unigram, bigram, and trigram similarity; thesis clarity keywords, which are the subdivision of the initial prompt to logical parts; LDA statistically generated topics.

3. Task, Data and Preprocessing

For experiments we use the collection 40 thousand theses in pedagogics from various universities defended in 2017–2018 (further FullCollection). 120 theses from this collection have double scores from two experts belonging to different institutions (further AnnotatedCollection). This collection is new and the current study is the first one based on this collection.

The theses have similar structure. They include several parts: introduction, two-three chapters, sometimes recommendations, conclusion, appendices. In the introduction, a student introduces the theme of the thesis, the aim, the object, and the tasks of the work. The first chapter presents the survey of theoretical studies related to the theme of the work. The second and third chapters often describe practical experiments carried out by the student.

To have more information about the thesis’ theme, we gather the above-mentioned structural elements (the theme, aim, object, and tasks of the thesis) into a specific document called theme header. The theme header conveys the main idea and direction of the thesis. It is clear that all parts of the thesis should correspond to the theme header in some extent. In this paper we assess how the first chapter of the
thesis, survey, is related to the theme header. We extract the theme header and chapters of the thesis using a specialized vocabulary and patterns. Figure 1 presents an example of a theme header.

---

**Figure 1:** Theme header for thesis “Safety and resilience of students of the Yaroslavl Town Planning College in the educational process”

The similarity between the theme header in a thesis and the prompt in essay writing can be seen. But the difference between two tasks: relatedness of a thesis to its theme header and an essay to the prompt is also significant:

- The theme elements are written by a student and can be poorly worded but the prompt in essay writing is formulated by professionals,
- There can be many essays for a given prompt. Therefore some methods can be specially tuned for a specific prompt [1]. The student thesis’s theme header is unique,
- The survey chapters are much longer than essays. In our data, they contain 250 sentences on average. Also, the theme headers are in most cases much longer than usual essay prompts,
- The survey chapters are never pervasive or argumentative in contrast to essays.
In the current study, we do not have any manual annotation of the theme relatedness of the first chapters. For evaluation, we use the overall score given to a thesis by two professional experts according to 2–5 scale, where 5 is the maximum grade in the scale. We suppose that low-scored theses should also have some problems in its surveys. As an example of “bad” segments for the same thesis, the theme header of which was presented, see Figure 2.

Figure 2: “Bad” segment for thesis “Safety and resilience of students of the Yaroslavl Town Planning College in the educational process”

We order all the theses according to the increase of the automatic scores of theme relatedness and evaluate methods of theme relatedness calculation using Average precision of 2-grade in the first positions of the created ranking. Table 1 shows the distribution of grades in 120 theses. Table 2 shows the deviation between grades of two experts. We calculate the Average precision measures according to the minimal grades of the theses. Thus, a thesis should have at least one 2 grade to be considered as a correct result in the beginning of the calculated rating.

Table 1: Distribution of marks in student theses

<table>
<thead>
<tr>
<th>Mark</th>
<th>Number of theses with minimal mark</th>
<th>Number of theses with maximal mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>51</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>28</td>
</tr>
</tbody>
</table>
Table 2: Deviations between thesis marks

<table>
<thead>
<tr>
<th>Points of difference</th>
<th>Number of works</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>1</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

As preprocessing, we use the procedure of segmenting the whole chapter to thematic fragments. Then we calculate similarity between specific segments and the theme header. The overall score of the theme relatedness of a chapter is based on averaging segment scores of this chapter.

4. Segmentation of Thesis Chapter to Thematic Fragments

The thematic segmentation module should break up a long sequence of the text into segments so that the sentences in one segment are thematically similar, and the boundaries of the segments signal a violation of connectivity between blocks of text. This procedure is based on the TopicTiling algorithm [12]. For splitting the text into segments, a procedure for assessing connectivity violations has been implemented, which is applied to each sentence. Based on the values of this metric, selection of separating sentences is carried out, which become the beginnings of segments. The procedure for assessing connectivity violation is as follows.

For each sentence, its “left” and “right” contexts with the length of $w$ sentences are considered (sentences having the length of less than $k$ words are ignored). For each sentence from the context, its vector representation is calculated using word2vec [9]. The weighted average of the word embeddings based on idf multiplier is calculated. Using cosine similarity, the similarity of all pairwise combinations of sentences between the “left” and “right” context is calculated. Based on these values, the coherence score is calculated by averaging all the similarities—$coherence\_score$.

$$coherence\_score = \frac{1}{w} \sum_{v_l} \sum_{v_r} similarity(v_l, v_r)$$  \hspace{1cm} (1)

After each sentence has its $coherence\_score$ calculated, in the second pass for each sentence its $depth\_score$ value is calculated using the following formula.

$$depth\_score(s_i) = 0.5 \times (top\_left + top\_right - 2 \times coherence\_score(s_i))$$  \hspace{1cm} (2)

Where $top\_left$ is the peak value of $coherence\_score$ to the left of the sentence, in non-descending order, $top\_right$ calculated by analogy.

The mean value and variance are calculated for the $depth\_score$ vector, on the basis of which the threshold values are chosen for the selection of candidate-sentences. If the $depth\_score$ of sentence is above the threshold and no separating sentences were selected in the neighborhood of several sentences, then this sentence is chosen as the separator. Sentences are considered in the order of their $depth\_score$ values. Thus, after completing the procedure described above, the source text is broken up into segments.
5. Methods of Assessing Theme Adherence

5.1. Preprocessing

The text of the thesis was pre-processed as follows:

- The whole text was lemmatised and stop words were removed. The frequency characteristics of the words were calculated and the idf values were obtained;
- In addition, some words were removed from the text of the theme header based on a part of speech: verbs, adjectives and functional parts of speech.
- The procedure of extracting the concepts was carried out using the ontology [4]. Thus for each sentence there is a list of concepts contained in it. For concepts, idf was also counted;
- The word2vec model was trained on full collection of the theses, which contains 40 thousand documents. The parameters are: CBOW, vector length is 150, window size is 10. The training was conducted using the python gensim package; 1

As an additional source of information about the domain, we use the ontology on natural sciences and technologies [4], which comprises terms of scientific fields (mathematics, physics, chemistry, geology, astronomy etc.) and terms of technological domains (oil and gas, power stations, cosmic technologies, aircrafts, etc.). Currently, about 6,500 terms (including term variants) were added to OENT to describe the pedagogics domain.

The main unit of the ontology is a concept, which has a unique name, the set of text entries, which express this concept in the text, and concept relations. For example, the concept DEAF AND HARD OF HEARING EDUCATION has the following text entries: deaf education, deaf teaching, education of the deaf, teaching of the deaf (translation from Russian).

5.2. Features of Theme Adherence Assessment

The chapter under analysis (further document) is presented in the form of $N$ segments $S$ and compared with the theme header $H$. Two baseline models were implemented to accomplish this task.

Baseline 1 (Tf-Idf). In this baseline model, the theme header and segments are represented as sparse vectors, where each element of the vector corresponds to a word from the collection's vocabulary. The values of vector elements are calculated as tf-idf. Based on the cosine similarity between the theme header and the $S_i$ segment vectors, the adherence_score with the theme is formed.

Baseline 2 (SegWord2Vec). Each segment, including the theme header, is converted into a vector using word2vec embeddings. This operation is performed by weighted averaging of the word vectors with idf as weights, as used in the segmentation procedure. The following features were implemented and combined with the baselines:

NoNorm: Disabling normalization for the theme header vector. We introduce this feature, because the normalization gives lower adherence_score's to larger and richer theme headers.

Concepts: Adding an additional vector for the theme header and segments, which is formed by analogy with the word vector, but based on the concepts of the ontology founded in the text of the segment/header. The concepts allow accounting for synonyms and multi-word expressions. In this case, the thematic adherence is formed as a weighted sum of the vectors similarities.

\[
\text{adherence_score} = \alpha \times \text{sim}_{\text{word}} + (1 - \alpha) \times \text{sim}_{\text{conc}}
\]

Keywords: For the theme header, the most significant \( k_w \) words and \( k_c \) concepts are determined according to tf-idf. The set of keywords includes words and concepts whose tf-idf weights exceed the threshold. This threshold is calculated as \( 0.2 \times \) average value to the 3 (2 for concepts) most significant (by tf-idf) words (or concepts). The weights of the keywords are multiplied by additional factors \( w_w \) and \( w_c \), respectively. Keywords for thesis “Safety and resilience of students of the Yaroslavl Town Planning College in the educational process” are as follow:

- words—resilience (1.00), safety (0.47), town-planning (0.42), Yaroslavl (0.30), college (0.21), student (0.17);
- concepts—urban planning (1.00), safety (0.63), college (0.57), student (0.20), system approach (0.17);

EmbedExp: There is an expansion of keywords for the theme header, by adding most similar words to them using word2vec embeddings. To do this, for each of the \( k_w \) keywords, the \( n_{w2v} \) closest words are calculated using the word2vec representation. Those words that are present in the whole thesis are added to the theme header vector. In order to calculate the weight of new words in the vector, the following formula is used:

\[
\text{weight}_{\text{word_ext}} = \text{sim}(\text{word}_{\text{raw}}, \text{word}_{\text{ext}}) \times \text{tf}(\text{word}_{\text{raw}}) \times \text{idf}(\text{word}_{\text{ext}})
\]

Where \( \text{word}_{\text{raw}} \)—the keyword on which the expansion is made, \( \text{word}_{\text{ext}} \)—new word. In addition, the weights of the new words are multiplied by the factor of \( w_{\text{ext}} \). The set of expanded words for the thesis “Safety and resilience of students of the Yaroslavl Town Planning College in the educational process”, includes:

- new words—hardiness (0.88), Maddy (0.76), tough (0.66), stories (0.62), coping (0.54), freshman (0.48), security (0.44), safe (0.41), highschool (0.14), scholar (0.13);

Regardless of the specific configuration, each segment and theme header are represented by a vector (or vectors) and adherence_score is calculated as cosine similarity between corresponding vectors. We calculate adherence_score for the whole chapter in two ways:

- \text{mean}: The average value of adherence_score’s of all segments;
- \text{mean\_worse}: The average value of adherence_score’s among the worst 20% of segments;
The \textit{mean\_worse} variant corresponds to the hypothesis that a thesis is characterized by its worst fragments. Further, theses are ranked in ascending order of their adherence\_score values.

6. Evaluation and Results

As mentioned earlier, we have 2 expert grades for each thesis. The minimum score (2) is chosen as the reference value, assuming that at least one expert could find serious problems in the theses. It was also previously shown that in the reference collection there are 42 works with the minimum grade of 2.

The evaluation methodology is proposed as follows: the algorithm for each thesis forms the values of adherence\_score (mean and mean\_worse), on the basis of which the reference collection is ranked so that the “worst” thesis was “above”. For evaluation we use average precision measure.

\[
\text{average\_precision}(n) = \frac{\sum_{k=1}^{n} P(k) \cdot \text{rel}(k)}{n}
\]  \hspace{1cm} (5)

Where \(P(k)\)—precision at \(k\), \(\text{rel}(k)\) is equal to 1 if \(k\)-th element of the list is relevant, otherwise 0.

We calculate average\_precision(25) measure in 25th position (20\% of the collection). The thesis is considered as relevant if it has grade 2 for at least one expert. In addition, the mean values of the average\_precision\((n)\) for positions from 1 to 25 were calculated.

The results of the evaluation of the configurations can be seen in Table 3. It also presents the result of random ordering (averaging 25000 random permutations). NoNorm did not give any significant improvements for any configurations. The results of average precision measures are also shown in the Figures 3 and 4.

Table 3: Evaluation results on 120 reference theses

<table>
<thead>
<tr>
<th></th>
<th>av_prec(25) by mean_worse</th>
<th>av_prec(25) by mean</th>
<th>mean av_prec(25) by mean_worse</th>
<th>mean av_prec(25) by mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.15</td>
<td>0.15</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Tf-Idf</td>
<td>0.15</td>
<td>0.16</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>Tf-Idf + Concepts</td>
<td>0.15</td>
<td>0.22</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>Tf-Idf + Keywords</td>
<td>0.23</td>
<td>0.21</td>
<td>0.16</td>
<td>0.29</td>
</tr>
<tr>
<td>Tf-Idf + Concepts + Keywords</td>
<td>0.20</td>
<td>0.27</td>
<td>0.26</td>
<td>0.43</td>
</tr>
<tr>
<td>Tf-Idf + Keywords + EmbedExp</td>
<td>0.21</td>
<td>0.22</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>Tf-Idf + Concepts + Keywords + EmbedExp</td>
<td>0.20</td>
<td>0.28</td>
<td>0.25</td>
<td>0.41</td>
</tr>
<tr>
<td>SegWord2Vec</td>
<td>0.28</td>
<td>0.22</td>
<td>0.37</td>
<td>0.18</td>
</tr>
<tr>
<td>SegWord2Vec + Concepts</td>
<td>0.28</td>
<td>0.27</td>
<td>0.37</td>
<td>0.40</td>
</tr>
<tr>
<td>SegWord2Vec + Keywords</td>
<td>0.30</td>
<td>0.19</td>
<td>0.47</td>
<td>0.36</td>
</tr>
<tr>
<td>SegWord2Vec + Concepts + Keywords</td>
<td>0.29</td>
<td>0.29</td>
<td>0.46</td>
<td>0.49</td>
</tr>
</tbody>
</table>
For the best configuration (accounting for both mean and mean_worse), the optimal parameters were as follows:

- **mean**: SegWord2Vec + Concepts + Keywords:
  \( \alpha = 0.25, k_w = 12, w_w = 1.25, k_c = 5, w_c = 1. \)

- **mean_worse**: SegWord2Vec + Concepts + Keywords:
  \( \alpha = 0.75, k_w = 12, w_w = 1.25, k_c = 3, w_c = 5. \)

Where \( \alpha \) — the parameter which controls the participation of concepts, 
\( k_w \) — number of word keywords, \( w_w \) — multiplier of word keywords, 
\( k_c \) — number of concept keywords, \( w_c \) — multiplier of concept keywords.

**Figure 3:** Average precisions graphs for base and best configurations for av_precision by mean_worse

**Figure 4:** Average precisions graphs for base and best configurations for av_precision by mean
Based on the results, we can conclude the following:

- The configurations based on SegWord2Vec are better than those based on Tf-Idf.
- Ranking based on mean_worse (worst segments) at least not worse then mean adherence_score and better corresponds to the task of finding the worst thesis.
- Use of Keywords always leads to better results.
- Use of Concepts is very useful for Tf-Idf and sometimes useful for SegWord2Vec.

7. Error analysis

For the analysis of the system, we consider two configurations: Tf-Idf и SegWord2Vec + Concepts + Keywords.

**Tf-Idf.** The first 5 theses have the following scores and grades:

1) 0.0 : 4;
2) 0.001 : 3;
3) 0.001 : 2;
4) 0.001 : 3;
5) 0.001 : 4.

The scores of the worst theses are very close to each other and are practically zero. This means that the algorithm did not reveal similarities between the worst segments and theme header. The first thesis with the grade 4 has score 0.0. Among the 13 worst segments the similarity to the theme header is zero. The thesis itself has the name “Differentiated approach in improving the physical fitness of students in 6th grade”, but in all worst segments author is talking about the biological characteristics of children and their development. The amount of specific biological information seems excessive. In addition, all the worst segments are plagiarized, and the text was deliberately distorted (every 3–4 words were simply removed from the text), which prevented the anti-plagiarism systems from detecting plagiarism.

The second thesis with grade 4 also has a low score of 0.001. Among the 10 worst segments, there are also no exact matches of words with the theme header and this is due to the fact that they also deviate from the main theme of the work. The theme of the thesis is “The implementation of a systematic approach to teaching biology in primary school” but in all the worst segments there is a serious bias in the philosophical direction, to questions of knowledge.

**SegWord2Vec + Concepts + Keywords.** The first 5 theses have the following scores and grades:

1) −0.099 : 2;
2) −0.098 : 3;
3) −0.076 : 2;
4) −0.056 : 2;
5) −0.037 : 2.

The spread of values here is much larger, which suggests that the model better separates different theses, among other things, it is clearly seen that the estimates are better grouped (this is also evident on the Figure 3).
We looked at the content of the worst theses and the text of the worst segments is poorly relevant to a given topic. But at the same time there were situations that large segments sometimes get low scores even if they contained some amount of relevant information. This is due to the fact that the averaging of word2vec vectors works worse on large texts. In addition, in comparison with \textit{Tf-Idf}, it became more difficult to interpret, why exactly one segment is worse than the other.

As a result, we can draw the following conclusions from the error analysis:

- \textit{Tf-Idf} badly detects links between related segments, but the text of which use different words. This leads to the fact that among the worst segments are those that do not seem to relate directly to the topic, but at the same time have some consistency with it.
- \textit{Tf-Idf} poorly separates bad theses from each other.
- \textit{SegWord2Vec + Concepts + Keywords} on the other hand, organizes theses well and, on average, highlights really bad segments, in which there is no useful information for thesis, but at the same time, the interpretability suffers a little.
- \textit{SegWord2Vec + Concepts + Keywords} also sometimes puts very low weights on large segments, but at the same time in which there is some amount of relevant information.

\section{Conclusion}

In this paper we studied approaches to assessing the quality of student theses in pedagogics. We considered a specific subtask in thesis scoring of estimating its adherence to the thesis's theme. The special document (theme header) comprising the theme, aim, object, tasks of the thesis is formed. The theme adherence is calculated as the similarity value between the theme header and thesis segments.

For evaluation we ordered theses in the increased value of the calculated theme adherence and compared the ordering with expert grades using the average precision measure. The best configuration for theses ranking is based on the weighted averaged sum of word embeddings (word2vec) and keywords extracted from the theme header.

\section{Acknowledgements}

This article contains the results of the project “Developing new methods for analyzing large text data using linguistic and ontological resources, machine learning methods and neural networks”, carried out as part of the Competence Center program of the National Technology Initiative “Center for Big Data Storage and Analysis”, supported by the Ministry of Science and Higher Education of the Russian Federation under the Contract of Moscow State University with the Fund for Support of Projects of the National Technology Initiative No. 13/1251/2018 dated December 11, 2018.
References