Aligning Learning Outcomes to Learning Resources: A Lexico-Semantic Spatial Approach

Swarnadeep Saha[1] , **Malolan Chetlur**[1] , **Tejas Indulal Dhamecha**[1] , **W M Gayathri K Wijayarathna**[2] , **Red Mendoza**[2] , **Paul Gagnon**[2] , **Nabil Zary**[2] , **Shantanu Godbole**[1]
[1]IBM Research - India
[2]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
{swarnads, mchetlur, tidhamecha, shantanugodbole}@in.ibm.com, {gwkumari, mendozaredante, pgagnon, nabil.zary}@ntu.edu.sg,

Abstract

Aligning Learning Outcomes (LO) to relevant portions of Learning Resources (LR) is necessary to help students quickly navigate within the recommended learning material. In general, the problem can be viewed as finding the relevant sections of a document (LR) that is pertinent to a broad question (LO). In this paper, we introduce the novel problem of aligning LOs (LO is usually a sentence long text) to relevant pages of LRs (LRs are in the form of slide decks). We observe that the set of relevant pages can be composed of multiple chunks (a chunk is a contiguous set of pages) and the same page of an LR might be relevant to multiple LOs. To this end, we develop a novel Lexico-Semantic Spatial approach that captures the lexical, semantic, and spatial aspects of the task, and also alleviates the limited availability of training data. Our approach first identifies the relevancy of a page to an LO by using lexical and semantic features from each page independently. The spatial model at a later stage exploits the dependencies between the sequence of pages in the LR to further improve the alignment task. We empirically establish the importance of the lexical, semantic, and spatial models within the proposed approach. We show that, on average, a student can navigate to a relevant page from the first predicted page by about four clicks within a 38 page slide deck, as compared to two clicks by human experts.

1 Introduction

Problem based learning involves students to refer to learning materials while solving the problem at hand [Wood, 2003]. In such scenarios, it is important to show only the pertinent pages relevant from a Learning Resource (LR) to engage the learner. LRs range from books with several pages of information delving on various topics to well curated and concise slide decks that are organized based on desired learning outcomes. Learning Outcomes (LO) are statements that describe the demonstrable knowledge to be acquired during a learning activity [Harden, 2002]. Each LO is a brief text consisting of bloom verb(s) [Bloom and others, 1956] and the concepts or topics to be mastered during the learning activity.

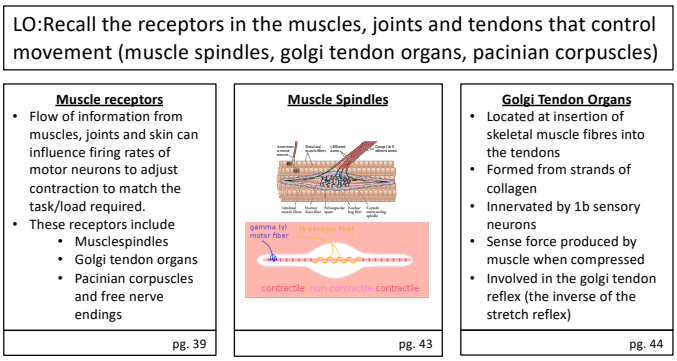

Figure 1: Figure showing an LO and three relevant pages from the corresponding LR with lexical aspect (left), semantic aspect (right) and spatial aspect (middle).

In this paper, we present the novel problem of identifying the set of relevant pages for an LO within an LR (slide deck). An LO can be considered as a broad question and this task is similar to finding pertinent pages in the document (LR) that answer the question. We consider a catalogue of LOs and their corresponding LRs spanning across five years of a medical curriculum. Figure 1 shows one such LO pertaining to muscle receptors and movement. The three pages numbered 39, 41, and 44 are identified relevant by content experts who are both doctors and faculty in the medical school. The individual pages of LRs, being slides, have prominent titles and bodies. The slides are usually designed with visual appeal and brevity in mind. Thus, they are typically composed of very few lines of text and contain images or tables or a combination of these. The brevity within a page and any arbitrary order of information pose significant challenges. Techniques to learn semantic representations of structured documents may not be sufficient in these scenarios where short sentences and keywords appear without any specific order among them.

We observe that identifying the relevancy of a page has three key aspects – (1) Lexical, (2) Semantic, and (3) Spatial. We motivate these aspects with the help of an example. In Figure 1, page 39 can be identified based on the overlap between the terms in the LO and those in the page (lexical aspect). On the other hand, although page 44 has significantly lesser overlaps, semantically it describes the receptors in the

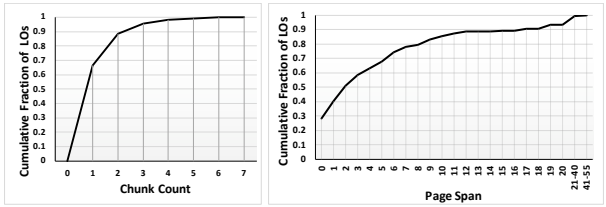

Figure 2: Distribution of chunk counts and page spans for the relevant pages of the LOs in the dataset.

Golgi Tendon Organ (semantic aspect). Finally, neither lexical nor semantic overlap will be beneficial for page 43 as it contains very little unstructured text used for labeling the images. However, the page, being in close proximity to a semantically relevant page, might refer to a continuation of the same concept and is also likely to be relevant (spatial aspect).

Obtaining annotated training data for this task is expensive as the expert has to go through the entire LR for annotating each LO. Constrained by a limited set of annotated LOs (694 in total), we observe that training a joint model that captures all the three aspects of the task might be prohibitive and almost impossible to scale to unseen LRs. Thus, we develop a pipelined approach consisting of separate models capturing each of the aspects that converts the alignment problem into a page relevancy classification problem. In summary, we make the following contributions in this paper.

1. We define the novel task of aligning Learning Outcomes (LO) to specific pages of Learning Resources (LR), where each LR is a slide deck.
2. We propose a novel two-stage *Lexico-Semantic Spatial* approach consisting of lexical, semantic and spatial models. Our approach is easily extensible and also alleviates the limited availability of training data.
3. We evaluate the effectiveness of our approach for the *page relevancy task* as well as the final *LO alignment task* using both standard metrics and a novel Click metric. Our approach not only achieves strong results against natural baselines but also learns important characteristics of the task.

2 Related Work

The task of aligning LOs to relevant pages in LRs is related to two broad bodies of work – Document Retrieval and Text Segmentation. However, our problem is unique compared to existing problems in these areas as discussed below.

2.1 Document Retrieval

Query-based document retrieval [Voorhees *et al.*, 2005; Mitra and Craswell, 2017] is a long-standing problem in Information Retrieval. For our task, an LO can be thought of as a single-sentence query and the pages of an LR can be treated as documents. In the education domain, [Contractor *et al.*, 2015] also use retrieval techniques for labeling learning content with learning standards. They build semantic representations of the documents and the standards, and finally rank them using a matching score. However, unlike such retrieval tasks, our task of LO alignment has two key differences - (1) The pages of an LR are not independent documents. In fact, the sequential order of the pages affects the relevancy output, as exploited by our spatial model. (2) The pages also do not follow a ranked order of relevancy; rather there are two distinct classes of relevance and irrelevance. Thus, we explicitly model the task as a classification problem and not a ranking problem. Our idea of using lexical and semantic features to evaluate the relevance of an independent document (page in our case) is however not new and has been explored before in multiple IR and NLP tasks. In fact, [Mitra *et al.*, 2017] show that for web search, some queries can be answered by exact matches while others require semantic matching in the embedding space. On similar lines, we also develop a lexical and semantic model for identifying a page's relevancy. Our semantic model is inspired from existing deep learning models for textual similarity and entailment tasks [Mueller and Thyagarajan, 2016; Conneau *et al.*, 2017].

2.2 Text Segmentation

Another close body of work is that of Text Segmentation, the problem of semantically dividing a document into contiguous segments. Techniques range from early unsupervised models like LDA [Riedl and Biemann, 2012] and graph-based methods [Glavaš *et al.*, 2016] to recent supervised deep models [Koshorek *et al.*, 2018]. However, these methods segment documents based on topical shifts only and are not governed by any query. A more closely related work is by [Bhartiya *et al.*, 2016] where they perform document segmentation for LO alignment using an unsupervised segmentation technique followed by ranking. To the best of our knowledge, LO alignment to slide decks has not been explored before. Unlike text books, slide decks are already segmented into a sequence of slides. Also, these slides pose multiple structural challenges including limited and unstructured texts among others.

3 Data Description and Analysis

Our dataset consists of a total of 100 LRs. The total number of LOs across these LRs is 694, with an average of 7 LOs per LR. The minimum, maximum and average number of pages for an LR are 10, 165, and 42, respectively.

We make the following observations from the expert-annotated pages for each LO. The average number of pages pertinent to an LO is 4.73, with the median at 3 pages, and the maximum and minimum being 45 and 1, respectively. Although the number of pages can be fairly large for certain LOs, around 71% of them have less than 6 relevant pages.

Our second key observation is that the set of relevant pages is not always a contiguous set. Figure 2 shows the cumulative distribution of the LOs with the number of chunks (a chunk is a contiguous span of pages). We find 34% of the LOs have more than 1 chunk. We also observe that more than 8% of the aligned pages belong to more than one LO. Thus, we treat the pages and the LOs independently in the lexical and semantic components of our model, without enforcing any constraint on the contiguity of the pages.

Our third observation is also plotted in Figure 2 where we show the distribution of the span of the pages. Page span of an

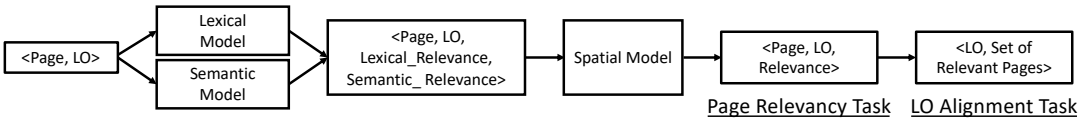

Figure 3: Block Diagram of the Lexico-Semantic Spatial approach.

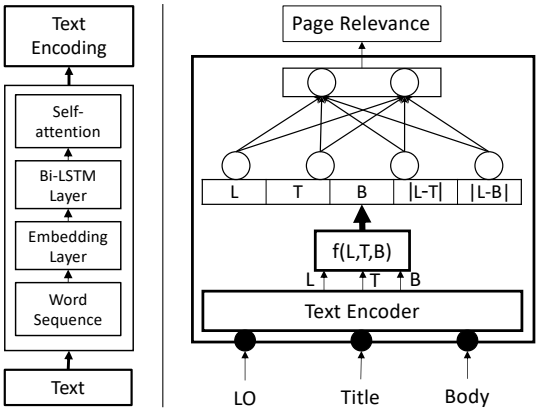

Figure 4: Architecture diagram of the Text Encoder (left) and the Semantic model (right).

LO is the difference between the maximum and the minimum page numbers from the relevant set. We observe that around 80% of the LOs have a span below 11, suggesting that the chunks tend to be in close proximity of each other. Our spatial model relaxes the independence assumption of the pages and achieves similar distributions in chunk counts and page spans.

4 Proposed Approach

In this section, we first formulate the problem and then describe the proposed lexico-semantic spatial approach.

4.1 Problem Formulation

We formulate the problem as a binary classification task where given an LO and a page of an LR , the model predicts its relevance. In this task, for each LO, all the relevant pages are positive samples and all other pages from the same LR are the negative samples. We call this task as the *page relevancy task*. During the *LO alignment task*, we collate all the predicted relevant pages by the page relevancy task as pertinent pages for the LO. Note that our choice of modeling has two key advantages - (1) It largely alleviates the data scarcity problem. From only 694 samples in the LO alignment task, we are now able to generate 26,253 samples for the page relevancy task, with 3,278 positive and 22,975 negative samples. (2) Our approach directly learns the distribution of the number of pages as opposed to a ranking formulation.

4.2 Lexico-Semantic Spatial Approach

Our approach is a combination of three models – a lexical model, a semantic model, and a spatial model. The spatial model infers the final relevancy using the relevancy scores of the other two models. Figure 3 shows the block diagram.

Lexical Model

The lexical model is composed of lexical overlap-based features between an LO and a page. For a page, we initially extract the tokens in the title and the body of the page separately. The details of the extraction algorithm are provided in the experiments section. We develop two features as follows.

LO and Title Overlap - This feature computes the proportion of overlap between the LO tokens and the title tokens of the page.

LO and Body Overlap - This feature computes the proportion of LO tokens that are overlapping with the body of the page. The body is an explanation of the title and the title tokens are often repeated multiple times. Therefore, we consider only those overlapping tokens within the body of a page that are not part of the title.

We tried experimenting with inexact overlaps of tokens using the cosine similarity of word vectors followed by thresholding. However, this deteriorated the performance because a well-curated LR mostly uses the same terms as mentioned in the LO. For example, a relevant page for the LO "Explain the mechanisms underlying Starling's Law of the Heart" would be using the exact phrase "Starling's Law of the Heart".

Semantic Model

Our semantic model, as shown in Figure 4, is a self-attention based neural architecture, aimed specifically for the samples where only lexical overlaps are insufficient to infer a page's relevancy. One of the key components of this model is a text encoder, used for encoding the LO, the title and the body of the page.

Text Encoder - The text encoder provides a dense feature representation of an input text. We use a Bidirectional Long Short-Term Memory Network [Hochreiter and Schmidhuber, 1997] (BiLSTM) with self-attention to encode the text. Each word in the text is first embedded using an embedding layer. The words are initialized with pre-trained word embeddings, which are further trained to reflect the task dependent nature of the words. The sequence of words are then passed through a BiLSTM layer. It generates a sequence of hidden representations. Not all words in the text contribute equally to its meaning. This is especially true for the body of the page which can be fairly long and a single BiLSTM layer might not be able to learn a good representation of it. Thus, we apply self-attention [Liu *et al.*, 2016; Lin *et al.*, 2017] to extract the importance of the words and aggregate their representations to form a vector. Formally,

$$u_t = \tanh(Wh_t + b)\ ,\ \alpha_t = \frac{e^{u_t \cdot u_w}}{\sum_t e^{u_t \cdot u_w}}\ ,\ v = \sum_t \alpha_t h_t$$

Each u_t is a hidden representation of the BiLSTM output h_t formed by passing through a dense layer with weights W and bias b. The importance of each word is measured by taking the similarity between u_t and the context vector u_w. α_t is the normalized importance of each word obtained by passing the similarity values through a softmax function. The final text embedding is a weighted sum of the BiLSTM outputs.

Final Architecture - Our model is a siamese network [Mueller and Thyagarajan, 2016] i.e. we use the same text encoder with the same weights to encode the LO, the title, and the body. Let L, T, and B be d-dimensional embeddings of the LO, the title and the body respectively. Our final feature representation is computed as the concatenation of L, T, B, absolute difference between L, T and absolute difference between L, B. The absolute differences between the LO embeddings with those of the title and the body capture their relatedness in the semantic space. Finally, a dense layer with a softmax function converts the features into relevancy scores.

4.3 Spatial Model

The lexical and semantic models output a relevancy score for each page of an LR. The spatial model relaxes the independence assumption of each page and models the space spanned by the relevant pages in the resource. It uses the scores from all the pages of the same LR to improve the current page's relevance. For example, a page appearing between two highly relevant pages is also likely to be relevant (Consider pg 43 in Figure 1) Specifically, we develop the following features using the relevancy scores of the lexical model.

Relevance (R) - The relevance of the page.

Context Relevance (CR) - The relevancy scores of the neighbouring pages within a window of size k on either sides. In our experiments, we choose $k = 1$.

Range of Relevance (RC) - The absolute difference between the relevance of the current page and the pages in the LR with the minimum and maximum relevance scores. Thus, even if the page's relevance is low, if it is one of the higher values in the entire LR, it is likely to be relevant.

Range of Pages (RP) - The absolute difference between the page number of the current page with those of maximum and minimum relevance scores. The features are normalized by dividing by the total number of pages in the LR. It computes how far the page is with respect to the minimum and the maximum relevance pages.

Histogram of Relevance (H) - We compute a histogram of the relevance scores for all the pages in the LR. The features are these bin counts normalized between 0 and 1. They represent an approximate distribution of the relevance scores of the entire LR. In our experiments, we choose the number of bins to be 10 and bin size to be 0.1.

We also compute the same set of features using the scores of the Semantic model. Our final feature representation for the spatial model is a 34-dimensional vector, formed by concatenating the 17 features from both the lexical and the semantic model.

5 Experiments

We conduct experiments to show the effectiveness of our approach for both the page relevancy task and the LO alignment task.

5.1 Implementation Details

We use the PdfBox[1] java library to parse the pdf LRs. From each page, we extract all the text snippets in order along with their font sizes and if they are marked in bold. The title of the page is identified as the bold text snippets with largest font in the page. Everything else is considered as part of the body.

[1]https://pdfbox.apache.org/

	Train	Test	Total
LRs	75	25	100
LOs	535	159	694
(Page, LO) relevant pairs	2,547	731	3,278
(Page, LO) irrelevant pairs	18,192	4,783	22,975
(Page, LO) all pairs	20,739	5,514	26,253

Table 1: Details of train-test splits.

	M-F1	W-F1
Lexical on Title Only	0.5316	0.8261
Lexical on Title and Body	0.5659	**0.8359**
Semantic on Title Only	0.5958	0.8123
Semantic on Title and Body	0.5644	0.8156
Semantic on Title and Body with Self-Attention	**0.6375**	0.8356

Table 2: Comparison of the Lexical and Semantic models.

We perform the train-test split at the level of LRs – we keep 75 LRs for training and 25 for testing. We intentionally do so as our model needs to align LOs for unseen LRs. The page relevancy task is performed using 20,739 training samples and 5,514 test samples. Note that the original LO alignment task is still tested on 159 LOs. Table 1 shows the details.

The lexical and spatial models are trained using a Random Forest classifier with 250 estimators. For the Semantic model, the maximum length of the LO, the title and the body is set to 20 words. All word vectors are initialized with 250-dimensional embeddings, pre-trained on the PubMed corpus [Chiu *et al.*, 2016] and further updated for our task. The size of the LSTM hidden units is set to 100. We train the model for 10 epochs using a batch size of 32, categorical cross-entropy loss and Adam optimizer with a learning rate of 0.001.

5.2 Page Relevancy Task

We evaluate the effectiveness of our approach as follows. First, we compare the lexical and semantic models. Second, we show the effectiveness of the lexico-semantic spatial approach and finally, we perform a detailed ablation of the spatial features. Although we report both macro average-F1 and weighted-F1, we compare using the former due to the imbalance in the dataset.

Comparison of Lexical and Semantic Models

Table 2 compares the performance of the various lexical and semantic models. The inclusion of the overlapping tokens from the body of the page improves the macro-F1 of the lexical model by 3 points. The semantic model, which learns only the LO and the title embeddings improves the macro-F1 by further 3 points. However, we observe a drop in performance after including the body embedding. We believe this is because the text snippets in the body are often ungrammatical and un-ordered and considering them as a single sequence of text might be detrimental. The best macro-F1 is obtained by applying self-attention and is 7 points better than the lexical model. Self-attention helps the model to focus on only the key tokens that are useful for aligning the LO to the page.

	M-F1	W-F1	Rel P	Rel R	Rel F1
Lexical	0.5659	0.8359	**0.7304**	0.1149	0.1986
Lexical+Spatial	0.6332	0.8559	0.6991	0.2161	0.3302
Semantic	0.6375	0.8356	0.3812	0.3557	0.3680
Semantic+Spatial	0.6577	**0.8487**	0.4430	0.3611	0.3892
Lexico-Semantic Spatial	**0.6609**	0.8486	0.4381	**0.3776**	**0.4056**
Human	0.8362	0.9256	0.7369	0.6936	0.7146

Table 3: Comparison of Lexico-Semantic Spatial approach with ablated models and human performance for the Page Relevancy Task.

	Lexical		Semantic	
	M-F1	W-F1	M-F1	W-F1
R	0.5671	0.8363	0.5917	0.7975
R+CR	0.5878	0.8348	0.6335	0.8306
R+CR+RC	0.6253	0.8437	0.6392	0.8314
R+CR+RC+RP	0.6273	0.8500	0.6485	0.8391
R+CR+RC+RP+H	**0.6332**	**0.8559**	**0.6577**	**0.8487**

Table 4: Ablation of Spatial features on Lexical and Semantic model for the Page Relevancy Task.

Effectiveness of Lexico-Semantic Spatial approach

We now demonstrate the effectiveness of the proposed approach by performing ablation on the individual components. Specifically, we compare the lexico-semantic spatial approach against four baselines – (1) lexical model (2) lexical + spatial model – spatial features from the lexical model only. (3) semantic model, and (4) semantic + spatial model – spatial features from the semantic model only. We also compare against human performance by asking another content expert to align the LOs in the test set and then evaluating these against the gold results. In Table 3, we report the overall macro-F1, weighted-F1 and also the precision, recall and F1 of the relevant class for all the models. Application of the spatial features improves the macro-F1 of the lexical model by 7 points and that of the semantic model by 2 points. This shows that the spatial features are generic and can be used to improve any model that captures the relevancy of each page independently. The lexico-semantic spatial approach obtains the best macro-F1 at 0.6609, a further improvement from the lexical+spatial and semantic+spatial models. We believe that the spatial model can act on any ensemble of independent page relevancy models capturing complementary information and should lead to further improvements.

We observe that the lexical model has the best relevant class precision but a significantly low recall. The semantic+spatial model has the best recall but much worse precision. Overall, the lexico-semantic spatial model achieves the best F1 for the relevant class at 0.4056, a significant 20 points improvement over the lexical model. Compared to human performance, our approach's relevant class recall is significantly lower. This observation is further discussed as part of the LO alignment task and future work.

Ablation Study of Spatial Features

We provide a detailed ablation study of the spatial features on both our lexical and semantic models. Table 4 shows the results. We first describe our observations for the lexical model. We start with a single feature, the lexical relevance (R) of the corresponding page. Inclusion of the relevance scores of the neighbouring pages (CC) improves the macro-F1 by 2 points. The range features (RC and RP) improve performance by a significant 4 points. Finally, the histogram of relevance scores (H) leads to an overall improvement of 7 points in macro-F1. We observe consistent improvements with the spatial features for the semantic model as well.

5.3 LO Alignment Task

We now evaluate the effectiveness of our approach on the LO alignment task. For each LO, we have a gold set and a predicted set of pages obtained after collating all the relevant pages from the page relevancy task. Note that this might lead to scenarios where the predicted set is empty for certain LOs. Since we know that at least one page will be relevant for such LOs, we also add a post processing step (PP) to our approach whereby we include the most relevant page (highest relevance class confidence) within the LR. We compare the models using two metrics – (1) Precision, Recall, F1 and (2) Click.

F1 - Precision for each LO is computed as the number of common pages upon the number of predicted pages, while recall is the number of common pages upon the number of gold pages. Final precision and recall are obtained by averaging over the per LO precision and recall values in the test set. F1 is the harmonic mean of the final precision and recall.

Click Metric - The main objective of the LO alignment task is to allow easy navigation of pertinent pages in an LR to a student during problem solving. In our application, for an LO, the predicted pages are first ordered based on the page numbers and then the first predicted page is presented along with links to all the other predicted pages. The student, at any time, can also navigate to an adjacent page. Therefore, the presence of a relevant page closer to first predicted page becomes critical. We develop a novel *click metric* to measure the number of clicks a student has to spend to get to a relevant page from the first predicted page. For the LOs with at least one correctly predicted page, the click metric value is the index of the first correctly predicted page. In the absence of any correctly predicted page, the expected click metric is

$$click = |predicted\_pages| + \frac{f - f_b}{2} + \frac{f_w - f}{2}$$

We assume that the student first looks at all the predicted pages, and then, with equal probability, navigates to the adjacent pages on either directions from the first predicted page (f). f_b and f_w are the page numbers of the first pages encountered on the backward and forward direction respectively that overlap with the gold set. The final click value is the average of the per LO click values.

Table 6 shows the results of the LO alignment Task. The semantic model significantly outperforms the lexical model by 17 points better F1. Similar to the page relevancy task, application of spatial features improve both these models – 15 points better F1 for the lexical model and 5 points for the semantic model. In this task too, the lexico-semantic spatial model has the best precision, recall, and F1 scores. After the post-processing step, the lexico-semantic spatial model improves further. In order to gather insights about our approach's relatively low recall, we analyzed the pages where only the human expert was able to correctly classify them as relevant. Interestingly, we find that 68.3% of these pages contain images with very less or no text at all.

LOs	Lexical	Lexical+Spatial	Semantic	Semantic+Spatial	Lexico-Semantic Spatial	Gold
LO_1	[35-37] , 38 , 39	[35-37] , [38-39]	10,33 , [35-39]	34 , [35-39]	[35-39]	[35-39]
LO_2	6 , [8-9]	[6-7] , [8-9]	7 , 8 , 9 , 15	7 , [8-9] , 15	7 , [8-9]	[8-9]
LO_3	17 , 18 , 19 , [26-28]	[17-19] , [26-28]	17 , 18 , 19 , [26-28] , 30	[14-16] , 17 , 18 , 19 , [26-28] , 30	[17-19] , [26-28]	[17-19] , [26-28]
LO_4	[5-11] []	[5-11] []	5 , [6-9] , [10-11]	[5-6] , 7 , [8-10] , 11	[5-10] , 11	[5-11]
LO_5	39 , [40-45] , 46	[39-40] , [41-43] , [44-46]	6 , [40-41] , [42-43] , [44-45]	[40-41] , [42-43] , [44-45]	[39-41] , 42 , [43-46]	[39-46]

Table 5: Comparison of the gold pages with the pages predicted by the different models for five sample LOs. Green, red, and gray represent pages correctly classified as relevant, incorrectly classified as relevant, and incorrectly classified as irrelevant respectively.

	Precision	Recall	F1	Clicks
Baseline-Seq	-	-	-	18.35
Lexical	0.2315	0.1531	0.1843	12.55
Lexical+Spatial	0.3969	0.2936	0.3375	7.81
Semantic	0.3210	0.3846	0.3499	7.67
Semantic+Spatial	0.3947	0.4107	0.4026	5.80
Lexico-Semantic Spatial	0.4013	0.4342	0.4171	5.34
Lexico-Semantic Spatial +Post Processing	**0.4327**	**0.4609**	**0.4464**	**4.96**
Human	0.7330	0.7671	0.7497	2.01

Table 6: Comparison of the different models with human performance for the LO Alignment Task. For click metric lower value is desirable, for other metrics higher values are desirable.

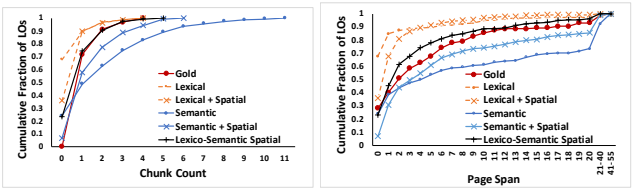

Figure 5: Comparison of the chunk counts (left) and the page spans (right) for the different models with the gold labels on the test set.

For the click metric, we additionally compute the sequential number of clicks – number of clicks spent by the students if they reach the first relevant page from the beginning of an LR, sequentially. This provides a baseline performance in the absence of the LO alignment system. Further, we compute human performance based on the annotated pages by the content expert. This provides a skyline performance in the presence of a highly accurate LO alignment system. Our experiments show that compared to the baseline click value of 18 and skyline human performance of 2 clicks for an average 38 page LR, our best approach yields click value of 4. This implies, in our application, a student can reach a relevant page by 4 clicks from the first shown page.

5.4 Deeper Analysis of the Spatial Model

We demonstrate how the spatial model helps in learning the important characteristics of the task. Figure 5 compares the chunk counts of the predicted pages by all the algorithms with those of the gold pages for the test LOs. For 70% of the LOs, the lexical model does not output any relevant page and hence the chunk count is also 0. The semantic model, on the other hand, outputs lot more relevant pages which are typically scattered, leading to larger number of chunks. The spatial model improves it by reducing the number of chunks. The lexico-semantic spatial model has the best chunk distribution and is closest to the gold.

We perform a similar analysis on the span of the predicted pages for the test LOs. Since the lexical model predicts only a small number of relevant pages, the span of the pages for majority of the LOs is also smaller. The semantic model, however, predicts pages that are scattered throughout the LR and hence the spans for many LOs are significantly higher. The graph corresponding to the lexico-semantic spatial model is again the closest to the gold one. Both these plots demonstrate that apart from predicting more correct pages, our approach also learns important characteristics of the task.

In Table 5 we provide examples showing how the spatial model reduces the chunk counts and the page spans. We observe two different phenomena with the output of the spatial model – (1) It fills intermediate pages between lexically or semantically relevant ones, and (2) It removes isolated irrelevant pages. The first phenomenon is visible in LO_3 where the spatial model brings in page 18 to complete the chunk [17-19]. The second phenomenon is observed in LO_1, where pages 10 and 33 are removed. Note that both these operations lead to a reduction in chunk counts. The span also typically decreases when irrelevant pages that are not in proximity of other chunks are removed.

6 Conclusion and Future Work

In this work, we introduced the novel problem of aligning LOs to relevant pages of LRs, the LRs being slide decks. The main contribution of this paper is in developing a novel pipelined approach capturing the lexical, semantic, and spatial aspects of the task with limited availability of annotated data. Our approach also learns important characteristics like number of chunks and page span for the set of relevant pages.

Empirical evaluation of our approach, through ablation studies, establishes the effectiveness of the lexical, semantic, and spatial models. While there is scope for improving the precision and recall metrics of the proposed approach, our experiment with the click metric is promising. We show that the learning experience in terms of a student navigating to a relevant page from the first predicted page is four clicks compared to human performance of two (for an average sized LR of 38 pages).

We observe that our approach does not exploit information contained in images leading to significantly lower recall. Therefore, as part of future work, we plan to extend our approach to include a model similar to the lexical and the semantic ones that processes the images. Finally, we see the need to improve the text encoder in the semantic model. This is mainly due to the challenges posed by the slides where the text snippets are non-sentential and without a definite order among them. These improvements should further enhance the LO alignment task. We believe that our approach is generic and as part of future work, we plan to validate this approach in the context of finding relevant pages of a document that answer a broad question.

References

[Bhartiya *et al.*, 2016] Divyanshu Bhartiya, Danish Contractor, Sovan Biswas, Bikram Sengupta, and Mukesh K Mohania. Document segmentation for labeling with academic learning objectives. In *EDM*, pages 282–287, 2016.

[Bloom and others, 1956] Benjamin S Bloom et al. Taxonomy of educational objectives. vol. 1: Cognitive domain. *New York: McKay*, pages 20–24, 1956.

[Chiu *et al.*, 2016] Billy Chiu, Gamal Crichton, Anna Korhonen, and Sampo Pyysalo. How to train good word embeddings for biomedical nlp. In *Proceedings of the 15th workshop on biomedical natural language processing*, pages 166–174, 2016.

[Conneau *et al.*, 2017] Alexis Conneau, Douwe Kiela, Holger Schwenk, Loic Barrault, and Antoine Bordes. Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*, 2017.

[Contractor *et al.*, 2015] Danish Contractor, Kashyap Popat, Shajith Ikbal, Sumit Negi, Bikram Sengupta, and Mukesh K Mohania. Labeling educational content with academic learning standards. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 136–144. SIAM, 2015.

[Glavaš *et al.*, 2016] Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto. Unsupervised text segmentation using semantic relatedness graphs. Association for Computational Linguistics, 2016.

[Harden, 2002] R.M. Harden. Learning outcomes and instructional objectives: is there a difference? *Medical Teacher*, 24(2):151–155, 2002. PMID: 12098434.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Koshorek *et al.*, 2018] Omri Koshorek, Adir Cohen, Noam Mor, Michael Rotman, and Jonathan Berant. Text segmentation as a supervised learning task. *arXiv preprint arXiv:1803.09337*, 2018.

[Lin *et al.*, 2017] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding. *arXiv preprint arXiv:1703.03130*, 2017.

[Liu *et al.*, 2016] Yang Liu, Chengjie Sun, Lei Lin, and Xiaolong Wang. Learning natural language inference using bidirectional lstm model and inner-attention. *arXiv preprint arXiv:1605.09090*, 2016.

[Mitra and Craswell, 2017] Bhaskar Mitra and Nick Craswell. Neural models for information retrieval. *arXiv preprint arXiv:1705.01509*, 2017.

[Mitra *et al.*, 2017] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. Learning to match using local and distributed representations of text for web search. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1291–1299. International World Wide Web Conferences Steering Committee, 2017.

[Mueller and Thyagarajan, 2016] Jonas Mueller and Aditya Thyagarajan. Siamese recurrent architectures for learning sentence similarity. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[Riedl and Biemann, 2012] Martin Riedl and Chris Biemann. Topictiling: a text segmentation algorithm based on lda. In *Proceedings of ACL 2012 Student Research Workshop*, pages 37–42. Association for Computational Linguistics, 2012.

[Voorhees *et al.*, 2005] Ellen M Voorhees, Donna K Harman, et al. *TREC: Experiment and evaluation in information retrieval*, volume 63. MIT press Cambridge, 2005.

[Wood, 2003] Diana F Wood. Problem based learning. *BMJ (Clinical research ed.)*, 326(7384):328–330, 02 2003.