Journal of Universal Computer Science, vol. 29, no. 1 (2023), 3-15
submitted: 3/4/2022, accepted: 24/10/2022, appeared: 28/1/2023

EntailClass: A Classification Approach to EntailSum and End-to-End Document Extraction, Identification, and Evaluation

Purvaja Balaji
(MIT CSAIL, Cambridge, Massachusetts, USA
https://orcid.org/0000-0002-5043-7695, pbalaji@mit.edu)

Helena Merker
(MIT CSAIL, Cambridge, Massachusetts, USA
https://orcid.org/0000-0001-5330-3230, hmerker@mit.edu)

Amar Gupta
(MIT CSAIL, Cambridge, Massachusetts, USA
https://orcid.org/0000-0001-9306-1256, agupta@mit.edu)

Abstract: The novelty of zero-shot text classification can address the fundamental challenge of the lack of labeled training data. With the current plethora of multidisciplinary, unstandardized text data, scalable classification models favor unsupervised methods over their supervised counterparts. Overall, the aim is to automate the labelling of each sentence in an input document consisting of section titles and section text. We propose an end-to-end pipeline that includes a document parser, a text classification model called EntailClass, and finally an evaluator to determine balanced accuracy. The suggested pipeline employs a zero-shot approach to classify text within any desired set of aspects. Moreover, text sentences are paired with their section titles and chronological order is maintained within sentences of the same aspect. The proposed automated, three-step pipeline represents a step towards solving the challenge of text classification without the need for an individual dataset for each aspect. It also offers the potential for seamless integration into existing workflows. This zero-shot, generalizable pipeline has achieved 87.2% accuracy and outperformed other state-of-the-art models when applied to supervisory documents.

Keywords: Text classification, Entailment, Zero-shot, Natural Language Processing
Categories: I, I.7, J.0, J.7
DOI: 10.3897/jucs.84647

1 Introduction

This paper proposes an unsupervised approach for aspect-based text classification to identify issues and actions within documents. The main objective of text classification is to enable users to extract information from textual resources and to organize large numbers of documents [Kadhim 2019]. In particular, text classification is critical to information management applications by allocating a specific document to one or more predefined classes [Kadhim 2019].

The purpose of aspect extraction is to identify and extract topics, aspects, and objects in the document [Maharani et al. 2015]. Aspect extraction identifies the category or categories contained in a given piece of text where the aspect is the topic while the term is a word or phrase in the text that references the topic [Sokhin et al. 2020]. Aspect identification usually involves extracting terms explicitly mentioned within the sentence, rather than implied terms [Byrne 2017].

Aspect-based text summarization provides targeted summaries, while generic summarization methods can touch on many aspects. Aspect-based summarization entails, given a document and a target aspect, generating a summary centered around that aspect [Fremann et al. 2019]. Current aspect-based summarization methods fall into two general categories. In the unsupervised category, these models are responsible for both identifying the aspect (often accounting for the polarity of words in Aspect Based Sentiment Analysis) and generating a short summary around that word. In the supervised realm, models cannot adapt to new aspects without training on a different dataset based around the new aspect. Several datasets including WikiAsp and MA-News have been created to test these methods [Frermann et al. 2019, Hayashi et al. 2021, Krishna et al. 2018].

Supervised methods result in models with inherent dependencies on the training dataset's domain, and thus require a new set of training data for each new category [Byrne 2017]. Within many frameworks, there is a tradeoff between accuracy and robustness. Achieving a higher accuracy often involves supervised learning with sentence-level datasets identifying aspect terms, while robust domain-neutral models rely on more general features and unsupervised approaches [Byrne 2017]. As a result, while supervised learning tends to be more effective, these models necessitate a well-defined example dataset for training [Mir et al. 2017]. Supervised learning algorithms require human experts to organize and correctly label a large quantity of documents [Shafiabady et al. 2016]. Preparing such a dataset requires significant manual labor. Furthermore, the knowledge to label different classes might be unavailable in some disciplines, in turn necessitating unsupervised learning schemes [Shafiabady et al. 2016]. In situations where expert knowledge on how to properly label the training documents is limited or very expensive, unsupervised learning with a zero-shot approach can be utilized to properly categorize document text.

The use of transformer-based pre-trained models such as BERT and RoBERTa in aspect-based text analysis improves model performance beyond benchmark results [Narayanaswamy 2021]. The potential for zero-shot transformers for the task of NLI was recently demonstrated by [Yin et al. 2019]. Transformers are integral to this process because they enable the creation of a generic, non-aspect aware summary.

While there has been much work performed by others on aspect-based summarization, EntailSum decomposes aspect-based summarization into a modular, two-step approach: aspect-based sentence pruning and generic summarization. Ankner et al demonstrated that this decomposition beats current state-of-the-art approaches, such as few-shot and weakly-supervised models trained on zero documents. This paper elaborates on the aspect-based sentence classification step (EntailClass) that is performed before pruning. There are three main explorations in this paper:

1. Integrating EntailClass into an end-to-end pipeline for repetitive use
2. Utilizing the pipeline to explore the results of feeding financial documents called supervisory reports from an external national banking organization to the model to demonstrate the utility of this approach in industry
3. Comparing EntailClass to other state-of-the-art zero-shot text classification models

2 Related Work

Several other approaches have explored aspect-based text classification and summarization. [Maharani et al. 2015] implemented a learning-based approach where a decision tree and rule learning were utilized to generate a pattern set, which were in turn used to extract aspects and opinions. [Bhamre et al. 2016] explored a weight-based ranking method for sentiment classification where aspects were extracted using a shallow dependency parser, an SVM classifier differentiated between positive and negative opinion, and the aspects were then ranked based on weights, the measure of the aspects' degree of importance. [Alasmari et al. 2017] discussed sentence-level sentiment classification, which involves studying each sentence individually and then determining whether that sentence includes a singular opinion and if that opinion is positive, negative, or neutral. [Kansal et al. 2014] also implemented aspect-based summarization of opinion words where an online dictionary was used to classify context, natural linguistic rules were utilized to assign polarity, and then a summary was generated after classifying each opinion word. In the domain of sentiment summarization, [Titov et al. 2008] created a statistical model without annotated data that used aspect ratings to determine the corresponding topics and then extracted fragments of text discussing these topics.

Regarding neural networks, [Krishna et al. 2018] used an attention-based RNN framework to produce multiple summaries of a singular document, each centered around a distinct topic, by adjusting the attention. To build an aspect-based summarization system, [Yauris et al. 2017] employed a modified Double Propagation (DP) that extracted pairs of aspects and corresponding sentiments, grouped them through aspect categorization, and combined them into a summary by aggregating aspect expressions of each aspect category based on sentiment orientation. [Shen et al. 2010] handled query-based multi-document summarization by proposing a topic aspect extraction method to identify aspect words and sentences, which were subsequently entered into a cross propagation framework and then an inner cluster sentence selection model to generate a summary. [Narayanaswamy 2021] employed a transformer-based architecture to extract aspect terms from sequences using dependencies between the words, and this approach is particularly effective on domain-dependent data as it employs parts of speech to identify the data's nouns and adverb.

In addition, clustering was a critical factor in several recent approaches. [Vargas et al. 2018] focused on aspect clustering methods to cluster explicit and implicit aspects. [Coavoux et al. 2019] addressed opinion summarization with an unsupervised abstractive summarization neural system that was characterized by clustering review sentences about the same aspects and then generating summary sentences focused on these aspects. For aspect-oriented summarization, [Li et al. 2011] clustered sentences into aspects, utilized an extended LexRank algorithm to rank each cluster's sentences, and then employed an integer linear programming (ILP) framework to select aspect-relevant sentences. Similarly, [Woodsend et al. 2012] explored a method of aspect summarization where individual aspects were learned separately from data but optimized jointly through the use of an ILP framework. [Shimada et al. 2011] conducted multi-aspect review summarization by using a clustering method to integrate similar opinions, in turn leading to redundancy reduction within the output summary. [Stammbach et al. 2021] employed clustering within an unsupervised text classification approach that utilized a pre-trained language model to create semantically informative vectors for each document. The approach assigns classes without a need for ground-truth labels by clustering neighbors as similar documents with proximate vectors [Stammbach et al. 2021].

Regarding the use of document structure, [Zhou et al. 2009] utilized each document's implicit structure by identifying the latent structure of specified topics, which was then used to compute the initial retrieval results. Likewise, [Fremann et al. 2019] applied latent document structure to identify aspect-relevant segments of the input document. This structure was then leveraged to produce both abstractive and extractive aspect-based summaries. [Shimada et al. 2010] also discusses aspect summarization regarding sentiment analysis by identifying aspects of each word in the target documents through structure information and rating information. The documents were then summarized using a scoring process to determine word-aspect relations.

EntailClass fills the need for a fully integrated, zero-shot method of aspect-based text classification. The key difference between aspect-based classification and aspect-based extractive summarization is that aspect-based classification does not prune sentences. The benefits of including human intervention to feed in the aspect pair is that there will be more coverage of the specific, desired topic, since EntailClass is only focused on the pair of aspects provided. Another benefit is that human-intervention eliminates noise, since the model will not classify sentences for which the human does not have express interest. There are many situations where binary classification or aspect-based summaries regarding two topics are useful, such as long text-based discussion on the pros and cons of a debated topic or email threads containing multiple documents. This paper also demonstrates that EntailClass provides another competitive approach to existing zero-shot classification techniques.

3 Proposed Approach

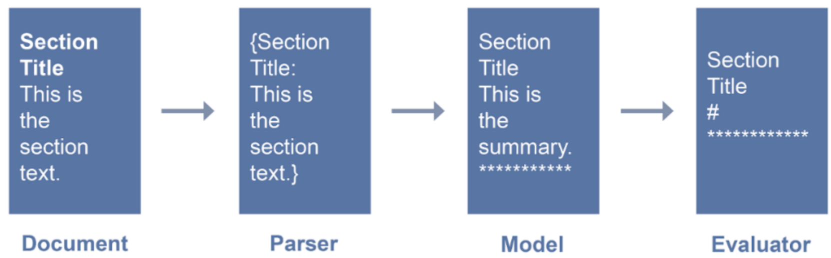

Figure 1: The proposed end-to-end approach.

This approach involves two inputs: a pair of aspects and a machine-readable document. The aspects are the two keywords that will serve as the labels for the binary classification of each sentence. Examples of aspects include "entertainment", "culture", "politics", etc. This pair of aspects is passed into to EntailClass in a list. The document that is fed into the model is composed of section titles and section text. The document parser extracts the section titles and matches the section text to the corresponding section title. The zero-shot entailment module EntailClass labels each sentence of the section text as one of two aspects. The output consists of all the sentences split between the two aspects passed in as prompts. The chronological order of the sentences is maintained. The evaluator utilizes the applicable metrics, balanced accuracy, precision, and recall, to determine the performance of the output.

3.1 Data

The input documents to EntailClass for the following experiments include supervisory reports, which are redacted versions of financial documents provided by the central bank of a country. The input document is a machine-readable pdf composed of section titles and corresponding section text. Text in bullet point format is treated as a set of full sentences. The model's architecture can be extended for multiclass classification, but the nature of supervisory documents is to address the issues and areas of enhancement for financial institutions, which naturally prompts binary classification. While the results mentioned focus on experiments run on one supervisory document, consisting of tens of section titles and section text, the application of EntailClass to ten different supervisory documents has been successful. This use of supervisory documents rather than synthetic data demonstrates the impact of EntailClass on the financial sector and its potential application to fields with expansive text-based communication.

3.2 Parser

The parser is tuned on the common document structure of supervisory reports. The parser identifies section titles and section text, and then matches each section title with its corresponding section text. Section text is processed as paragraphs (both unnumbered and numbered), bullet points, numbered lists, and various formats of nested lists and bullet points. In addition, section titles are identified as separate from section text even if they themselves are numbered or are composed of words that begin with lowercase letters. The parser also handles nested titles and titles of multiple lines. After inputting the document into the parser, the resulting output dictionary of section titles and section text is passed to the zero-shot entailment module.

3.3 EntailClass Model

The model EntailClass is adopted from Ankner et al's EntailSum model and is modified for classification. Ankner et al.'s results revealed that this zero-shot summarizer is competitive with Tan et al's weakly supervised, abstractive summarizer model, indicating that aspect-based pruning paired with generic summarization is an effective framework for topic-oriented summaries [Anker et al. 2022, Tan et al. 2020]. This entailment approach will be the basis for the following binary classification experiments to test the limits of zero-shot models.

A pair of aspects are fed into the model as prompts, and they as labels in sentence classification of each sentence. We use the transformer DaBERTa for sentence-aspect entailment. DaBERTa is pre-trained on self-supervised tasks on large bodies of texts and fine-tuned for NLI. For each aspect, the network generates an entailment, neutral, and contradiction logit. These logits are generated by passing in a premise and text, where the premise is the text from the input document and the hypothesis is "This text is about {aspect}". At this point, we have the sentence from the input document, the entailment logit, and the contradiction logit. The neutral logit is ignored for these experiments. Then, each sentence's entailment with the aspect is calculated and stored as tuples. The scores are normalized and converted to probabilities using the sigmoid function. These normalized scores can be interpreted as the probability a sentence belongs to an aspect.

3.4 Evaluator

The evaluator compares the actual results with the results from the model. Since supervisory reports are written with the intent of issues and actions in mind, the following equations will consider aspect 1 to be "Issues" and aspect 2 to be "Actions." A human can intervene to determine if sentences are issues or actions. These are considered to be the true issue (TI) and action (TA) sentences, and they are passed into the evaluator in a dictionary where the keys of the dictionary are the sentence titles. Once these true issues and actions are identified, the first step of the evaluator is to determine the values for a confusion matrix for each section. In addition to the TI and TA, the two confusion matrices calculated for each section include false issue (FI) and false action (FA). FI and FA account for the sentences that are incorrectly labelled as aspect 1 and aspect 2, respectively. To compute the accuracy, precision, and recall of the classification model, we use Equations 1-3. This binary text classification model is evaluated with balanced accuracy, the average of precision and recall, because the sections are often imbalanced in terms of the proportion of target classes. Meaning, some sections contain many sentences related to aspect 1 over aspect 2 and vice versa.

1	Precision	$\frac{TI}{TI + FI}$
2	Recall	$\frac{TI}{TI + FA}$
3	Balanced Accuracy	$\frac{TPR+TNR}{2}$
4	True Positive Rate	$\frac{TI}{TI + FA}$
5	True Negative Rate	$\frac{TA}{TA + FI}$

Table 1: Equations to quantify the performance of EntailClass

Since the aspects passed into the module may result in most sentences in a section to be classified as either issues or actions, we decided to include balanced accuracy in our evaluator, to balance the skew of the results.

Figure 2 provides an example of the of the pipeline of the execution of EntailClass. Like any other execution of EntailClass the input to the pipeline is a supervisory document and the pair of aspects. The supervisory document contains the section titled

"Inappropriate Credit Scoring Model" and its corresponding section text contained within the two bullet point as pictured in Figure 2. The pair of aspects are ["Issue", "Mandates"]. The parser extracts the section title and section text. The sentences from the bullets are placed into a list of sentences with the same hierarchy, meaning there are no embedded lists of sentences in this example. This list of sentences is passed to the model, which labels each sentence as one of the aspects passed in. Not pictured is the final step of evaluating this classification with the balanced accuracy, precision, and recall metrics.

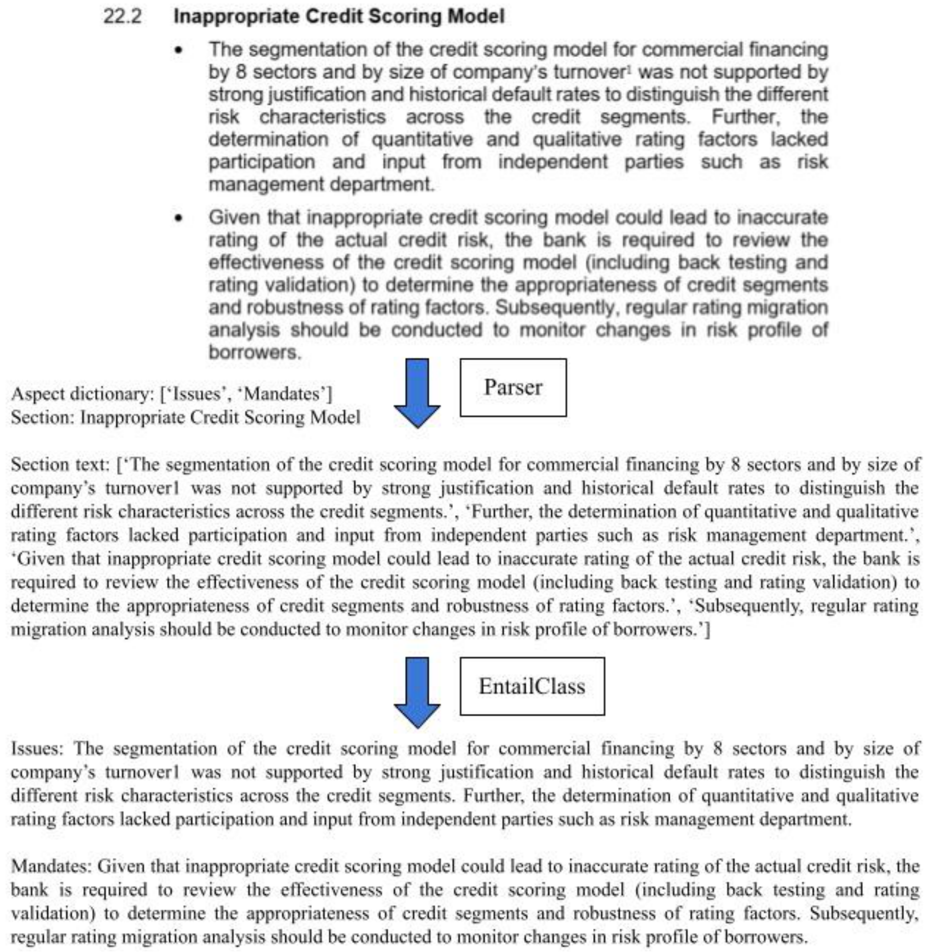

Figure 2: Example pipeline execution from input document, parser, to model before the final evaluation step

4 Experiments

The concerned central bank provided redacted supervisory reports, and we began by computing the accuracy statistics for each section of the first sample. To compute the accuracy of the classification model, we used equation 1. In addition, precision was computed using equation 2, recall was computed using equation 3, and balanced accuracy was computed using equation 4.

4.1 Setup

The following experiments were run on the individual sections of the document before a final average resulted. These experiments were used to determine the level of accuracy, precision, recall, and balanced accuracy with various aspects. The first step was to gather a list of common words of interest in finance. Then these common words were split into general issue aspects and action aspects. To understand the most relevant aspects to run the following experiments, the first experiment was running all issue aspects with one action aspect and one issue aspect with all the different action aspects. This gave us a final list of keywords and more focused results.

5 Results and Discussion

5.1 Parser Results

The parser does not rely on learnt document structure. Instead, it is able to accurately extract text from the document. Each section, including any embedded bullet points, are handled appropriately to feed into the model. Limitations to the parser include documents formatted in with different symbols (i.e. bullet points are asterisks). This was not a significant problem as the input documents, the supervisory reports, which maintain consistent formatting and the parser code is ready for change.

5.2 Aspect Pair Experimentation

Prompt	Balanced Accuracy	Precision	Recall
[“Issues”, “Mandates”]	**0.8717**	0.9445	0.8328
[“Observation”, “Mandates”]	0.8667	0.9531	0.8049
[“Shortcomings”, “Should consider”]	0.8301	0.9383	0.7961
[“Shortcomings”, “Mandates”]	0.8269	0.9284	0.7813
[“Lack”, “Intensify”]	0.8246	0.9583	0.7658

Table 2: Top 5 Keyword Experiment Results

Table 2 displays the five most accurate keywords. The highest accuracy achieved is 0.872. There were two pairs that achieved a balanced accuracy >0.85 and many pairs, beyond those listed, that achieved an accuracy >0.8. Due to the imbalanced nature of the number of sentences in each section related to aspect 1 vs aspect 2, balanced accuracy was favored as the accuracy metric.

A takeaway from these results is specificity vs generalizability. In term of specificity, the most accurate aspects are categorically related, and they match the typical topics of concern when writing a supervisory report. While all of these aspects match typical topics, the specific synonym used varies accuracy.

Related to generalizability, the variance in the most accurate aspects demonstrate the relevance of the aspect pairs passed in, but the number of accurate aspects supports the model's general versatility. These same top keywords can be applied to other supervisory reports, due to the categorical nature of these documents, and further experimentation would be required for documents related to other aspects. The most novel development 87.17% accuracy of EntailClass, supporting that zero-shot, aspect pruning is a competitive classification method.

5.3 Comparison to Generic Segmentation Methods

['Issues', 'Mandates]			
Segmentation method	Balanced Accuracy	Precision	Recall
EntailClass	0.8717	0.9445	0.8328
Top Half Aspect 1	0.6413	0.7360	0.6214
Top Half Aspect 2	0.2173	0.3409	0.2735
Random	0.3958	0.4765	0.3861

Table 3: Comparing EntailClass to Generic Segmentation Methods

To compare the model's classification performance with pruning to other generic segmentation methods, the experiment compiled in Table 3 indicates three other segmentation methods in the left column. Top Half Aspect 1 splits the sentences in half, where the top half are classified towards Aspect 1 and the bottom half as Aspect 2. Top Half Aspect 2 classifies the top half as Action 2 and the bottom half as Aspect 1 (reverse of row 2). Random classifies each sentence randomly. We compared these segmentation methods to the model's performance using the aspects ["Issues", "Mandates"].

Top Half Aspect 1, Top Half Aspect 2, and Random are three naïve ways of segmenting data. A comparison to these models serves as a baseline test as to whether EntailClass aids in classifying text faster than three ways a human could without NLP. It is reasonable that classifying the top half of the sentences as the first aspect leads to greater accuracy, since text often presents the problem before elaborating on action required to solve the aforementioned problem. Even with this reasoning, EntailClass outperforms these naïve methods that do not use entailment or machine learning.

5.4 Comparison to Zero-Shot Models for Text Classification

Balanced Accuracy Scores			
	EntailClass	facebook/bart-large-mnli	cross-encoder/nli-distilroberta-base
["Issues", "Mandates"]	**0.8717**	0.5906	0.5091
["Observation", "Mandates"]	0.8667	0.6422	0.5373
["Shortcomings", "Should consider"]	0.8301	0.7339	**0.8289**
["Shortcomings", "Mandates"]	0.8269	0.6642	0.5768
["Lack", "Intensify"]	0.8246	**0.7939**	0.8107

Table 4: Comparing EntailClass to Other Zero-Shot Models

Yin et al's method of using pre-trained NLI models to create a ready-to-use zero-shot classifier inspired significant work in the area of zero-shot text classification, including EntailClass and other state-of-the-art models. To compare EntailClass to existing zero-shot models for text classification, two models from the Hugging Face transformers library were evaluated on the same supervisory reports as the proposed model [Wolf et al. 2020]: facebook/bart-large-mnli and cross-encoder/nli-distilroberta-base. Both are also inspired by Yin et al's work and produces three scores, entailment, neutral, and contradiction for a given sentence. The first comparison model facebook/bart-large-mnli is a bart-large model trained on the MultiNLI (MNLI) dataset [Williams et al. 2018]. The second model is cross-encoder/nli-distilroberta-base and is trained using both the SNLI and MNLI datasets [Bowman et al. 2015].

The bolded numbers in Table 4 indicate the highest balanced accuracy for each of the models over the specified aspect pairs. EntailClass outperforms the other two models when any of the listed aspect pairs are passed in. Also, the highest balanced accuracy of EntailClass is significantly higher than the other two models. This demonstrates that EntailClass's method of employing DaBERTa pre-trained on the task of NLI is beats other zero-shot text classification methods in this application domain.

5.5 End-to-End Pipeline

The novel development is the end-to-end pipeline that strings together the parser, EntailClass, and the evaluator. Unlike related work, EntailClass does not rely on opinion words, rule learning, or clustering, and no training data is necessary. From now on, this workflow for all machine-readable documents with a common structure of bolded titles and paragraph text can be automated—both in the workplace and for experimentation.

6 Future Work

The parser delineates between section titles and text by formatting. Bolded text is identified as section titles and underlined to differentiate them for subtitles (bolded and not underlined) and section text (non-bold and not underlined). While this structure is common in documentation, changes to this generalizable structure, including italics, would require parser revisions to include learning. To do this, future models can include models to learn the structure of a document. Such future work can manifest by considering whitespace when processing a document or applying NLP techniques. This enhancement to the parser would allow EntailClass to have wider application to other documents and tables with text. Proposed experimentation includes fine-tuning the model for, possible, more accurate selections of issues and actions. The current model utilizes an entailment-based approach, and integration of zero-shot or few-shot models can improve results and functionality. This end-to-end model could also be applied to other related documents, such as long emails, annual reports, and medical records, which could benefit from the classification of the problems presented and the corresponding action items for a company, a company's employees, or an individual / patient. Fields of application include banking, healthcare, and insurance, all of which could save hours and the necessity for a large amount of training data by avoiding training time due to the zero-shot nature of this model.

Acknowledgements

The authors are grateful to Zachary Ankner for his collaboration on the development of EntailSum and Kenneth Hiew Chun Keat and others at the Bank of Malaysia for providing redacted supervisory documents as the dataset and for their valuable inputs on the application domain. This research is supported in part by funds from FinTech@CSAIL initiative.

References

[Alasmari et al. 2017] Alasmari, S. F., Dahab, M.: "Sentiment Detection, Recognition and Aspect Identification"; International Journal of Computer Applications, 177, 2 (2017), 31–38.

[Anker et al. 2022] Anker, Z., Balaji, P., Zhu, Y., Wang, P., Hiew, K.: "EntailSum: An Entailment Based Approach to Aspect-Based Text Summarization with Automated Aspect Adaptation"; International Journal of Pattern Recognition and AI (accepted for publication), 36, (2022).

[Bhamre et al. 2016] Bhamre, N. R., Patil, N. N.: "Aspect rating analysis based product ranking"; 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), (2016), 197-202.

[Bowman et al. 2015] Bowman, S., Angeli, G., Potts, C., Manning, C.: "A large annotated corpus for learning natural language inference"; Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, (2015), 632-642.

[Byrne 2017] Byrne, S.: "Aspect Identification and Sentiment Analysis in Text-Based Reviews"; Lehigh Preserve Institutional Repository, (2017).

[Coavoux et al. 2019] Coavoux, M., Elsahar, H., Gallé, M.: "Unsupervised Aspect-Based Multi-Document Abstractive Summarization"; Proceedings of the 2nd Workshop on New Frontiers in Summarization, (2019), 42-47.

[Frermann et al. 2019] Fremann, L., Klementiev, A.: “Inducing Document Structure for Aspect-Based Summarization”; Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, (2019), 6263-6273.

[Hayashi et al. 2021] Hayashi, H., Budania, P., Wang, P., Ackerson, C., Neervannan, R., Neubig, G.: “WikiAsp: A Dataset for Multi-domain Aspect-based Summarization”; Transactions of the Association for Computational Linguistics (TACL), (2020), 211-225.

[Kadhim 2019] Kadhim, Ammar Ismael.: “Survey on supervised machine learning techniques for automatic text classification”; Artificial Intelligence Review, 52, 1 (2019), 273–292.

[Kansal et al. 2014] Kansal, H., Toshniwal, D.: “Aspect Based Summarization of Context Dependent Opinion Words”; Procedia Computer Science, 35, (2014), 166–175.

[Krishna et al. 2018] Krishna, K., Srinivasan, B. V.: "Generating Topic-Oriented Summaries Using Neural Attention"; Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, (2018), 1697-1705.

[Li et al. 2011] Li, P., Wang, Y., Gao, W., Jiang, J.: “Generating Aspect-oriented Multi-Document Summarization with Event-aspect model”; Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, (2011), 1137-1146.

[Maharani et al. 2015] Maharani, W., Widyantoro, D. H., Khodra, M. L.: "Learning-based aspect identification in customer review products"; 2015 International Conference on Electrical Engineering and Informatics (ICEEI), (2015), 71-76.

[Mir et al. 2017] Mir, J., Mahmood, A., Khatoon, S.: “Aspect Based Classification Model for Social Reviews”; Engineering, Technology & Applied Science Research, 7, 6 (2017), 2296–2302.

[Narayanaswamy 2021] Narayanaswamy, G.R.: “Exploiting BERT and RoBERTa to Improve Performance for Aspect Based Sentiment Analysis”; Technological University Dublin, (2021).

[Shafiabady et al. 2016] Shafiabady, N., Lee, L. H., Rajkumar, R., Kallimani, V. P., Akram, N. A., Isa, D.: “Using unsupervised clustering approach to train the Support Vector Machine for text classification”; Neurocomputing, 211, (2016), 4-10.

[Shen et al. 2010] Shen, C., Wang, D., Li, T.: “Topic Aspect Analysis for Multi-Document Summarization”; Proceedings of the 19th ACM International Conference on Information and Knowledge Management, (2010), 1545-1548.

[Shimada et al. 2010] Shimada, K., Yamaumi, M., Tadano, R., Hadano, M., Endo, T.: “Interactive Aspect Summarization Using Word-Aspect Relations for Review Documents”; SCIS & ISIS, (2010), 183-188.

[Shimada et al. 2011] Shimada, K., Tadano, R., Endo, T.: “Multi-aspects review summarization with objective information”; Procedia - Social and Behavioral Sciences, 27, (2011), 140-149.

[Sokhin et al. 2020] Sokhin, T., Khodorchenko, M., Butakov, N.: “Unsupervised Neural Aspect Extraction with Related Terms”; Artificial Intelligence and Natural Language (AINL), 1292, (2020), 75-86.

[Stammbach et al. 2021] Stammbach, D., Ash, E.: “DocSCAN: Unsupervised Text Classification via Learning from Neighbors”; arXiv:2105.04024v2 [cs.CL], (2021).

[Tan et al. 2020] Tan, B., Qin, L., Xing, E., Hu, Z.: “Summarizing Text on Any Aspects: A Knowledge-Informed Weakly-Supervised Approach”; Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), (2020), 6301-6309.

[Titov et al. 2008] Titov, I., McDonald, R.: "A Joint Model of Text and Aspect Ratings for Sentiment Summarization"; Proceedings of ACL-08: HLT, (2008), 308-316.

[Vargas et al. 2018] Vargas, F. A., Pardo, T. A. S: "Aspect Clustering Methods for Sentiment Analysis"; Computational Processing of the Portuguese Language, (2018), 365-374.

[Williams et al. 2018] Williams, A., Nangia, N., Bowman, S.: "A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference"; Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (2018), 1112-1122.

[Wolf et al. 2020] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush A.: "Transformers: State-of-the-Art Natural Language Processing"; Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, (2020), 38-45.

[Woodsend et al. 2012] Woodsend, K., Lapata, M.: "Multiple Aspect Summarization Using Integer Linear Programming"; Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, (2012), 233-243.

[Yauris et al. 2017] Yauris, K., Khodra, M. L.: "Aspect-based summarization for game review using double propagation"; 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), (2017), 1-6.

[Yin et al. 2019] Yin, W., Hay, J., Roth, D.: "Benchmarking Zero-Shot Text Classification: Datasets, Evaluation and Entailment Approach"; Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), (2019), 3914-3923.

[Zhou et al. 2009] Zhou, D., Wade, V.: "Latent Document Re-Ranking"; Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, (2009), 1571-1580.