Relevancy and Diversity in News Recommendations[⋆]

Shaina Raza[1,*], Chen Ding[2]

[1]*Toronto Metropolitan University, ON, Canada*
[1]*Toronto Metropolitan University, ON, Canada*

Abstract

News recommendation systems face unique challenges, including the dynamic nature of user preferences and the need for diversity in recommended news articles. To address these challenges, we propose a deep neural network architecture that learns representations for both news items and users. Our approach uses an enhanced vector for each query and news item to facilitate information interaction between these entities. To overcome selection bias in implicit user feedback, we employ negative sampling. We also promote diversity in recommended news by aligning the uneven news category representations of items in a loss function. Experimental results on a benchmark dataset demonstrate the superiority of our proposed architecture over baselines, achieving both relevancy and diversity in the news recommendations.

Keywords

News Recommender, Systems, Relevancy, Diversity, Retrieval, Deep Neural Networks

1. Introduction

Prominent news organizations such as Yahoo!, BBC, NYTimes, and CNN have introduced online news portals, which users can access from any location to peruse a wide range of news categories and stay up-to-date. However, with the abundance of information available on the internet, locating pertinent news has become a challenging and time-consuming task. News recommender systems (NRS) address the issue of information overload by presenting users with personalized and intriguing recommendations, chosen from an extensive pool of available news articles [1].

An NRS must balance between maintaining relevance to the user's interests and introducing enough diversity to keep the user engaged. If the NRS recommends news articles that are too diverse or unrelated to the user's interests, the user may lose interest and stop using the system. Conversely, if the NRS recommends only news articles that are closely related to the user's interests, it may miss out on opportunities to introduce the user to new topics and categories. Therefore, finding the optimal balance between relevance and diversity is crucial for the success of an NRS.

In general, users' preferences can be either long-term or short-term, with short-term preferences defining their current interests. For example, a user may have a long-standing interest in

✉ shaina.raza@torontomu.ca (S. Raza)

CEUR Workshop Proceedings (CEUR-WS.org)

a particular topic, such as climate change, and read numerous articles related to this topic over the course of their lifetime [2]. However, if the user has been reading only news articles related to climate change for an extended period, the NRS may consider suggesting news articles on related topics, such as sustainable living or renewable energy, to introduce some diversity into the recommendations. By doing so, the NRS can help the user explore related topics and expand their knowledge in the field, while also keeping them engaged with new and exciting news articles.

Short-term interests can also be identified by analyzing the user's recent behavior on the NRS. For instance, if a user has been reading numerous articles on a particular topic over the past few days, the NRS can infer that the user is currently interested in that topic and adjust its recommendations accordingly. By incorporating short-term interests into the recommendation process, the NRS can provide more personalized and relevant recommendations to the user.

This paper presents an NRS that prioritizes both relevancy and diversity in the recommendation process, achieving a more balanced and engaging user experience. The specific contributions of this paper are:

- The proposed deep neural network is based on a query-candidate architecture [3][4], featuring a query (user) model that retrieves relevant news items and an item (candidate news) model that ranks them based on user actions. The model introduces a similarity score between the user feedback and item representations, enabling it to recommend a more diverse set of news articles.
- The proposed model takes into account the uneven distribution of news articles across different categories, which is prevalent in news data, and ensures that recommendations include a balance of news items from all categories.
- The proposed model also offer a negative sampling approach to tackle the selection bias of implicit user feedback by bringing in news samples from the entire news corpus.

Extensive experiments on a news dataset show that our proposed approach can provide both relevant and diversified news recommendations in an NRS.

2. Proposed Approach

Next, we discuss our approach.

2.1. Problem Formulation

The objective of news recommendation is to select relevant candidate news items from a news corpus, given a set of queries. The item set is represented as $\{v_i\}_{i=1}^{N}$ and the query set as $\{v_j\}_{j=1}^{M}$. The recommendation problem, denoted as R, is learned from the query-item feedback represented by a matrix $\mathbb{R}^{N \times M}$. Each query is considered as feedback provided by the user. If query j gives positive feedback on news item i, then $R_{ij} = 1$; otherwise, it is considered as non-positive feedback.
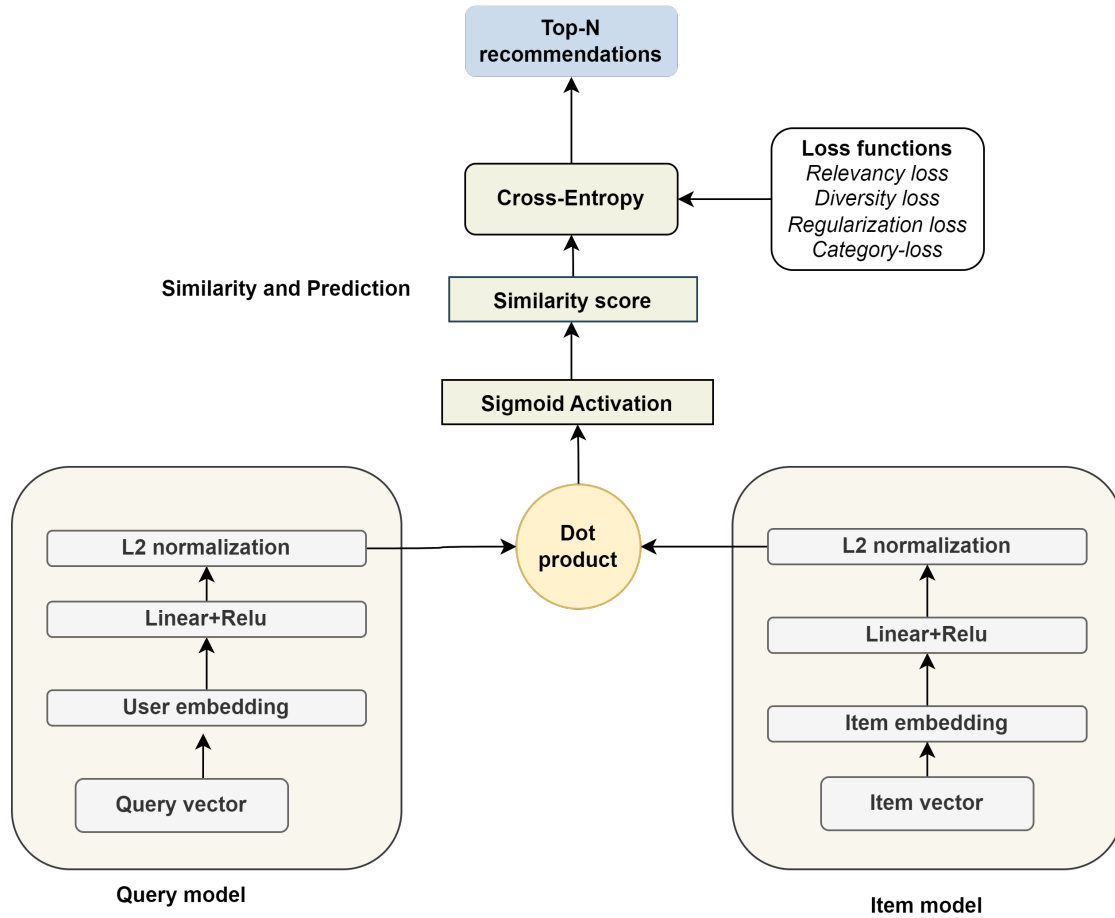
Figure 1: Proposed architecture.

2.2. Query-Candidate Model

We show our proposed architecture in Figure 1 and explain next.

The proposed model is a query-candidate model that consists of an embedding layer, an augmented layer, and two models (query and news item generation). The model takes into account different content features related to news items, such as news ID, title, body, and category, as well as contextual information.

The embedding layer is represented by an embedding matrix $E \in \mathbb{R}^{K \times D}$, which maps each piece of information (e.g., news item or user ID) in u_j and v_i to a low-dimensional dense vector $e_j \in \mathbb{R}^K$.

The augmented layer creates two augmented vectors a_u and a_v by concatenating the IDs corresponding to two input feature vectors $\mathbf{f}_u$ and $\mathbf{f}_v$, respectively. These augmented vectors are then concatenated with the original feature vectors $\mathbf{f}_u$ and $\mathbf{f}_v$ to obtain the augmented representations of query p_u and news item p_v. The fully connected layers with the ReLU activation function are applied to these concatenated vectors, and the output from the fully connected

layers goes through the L_2 normalization layer to obtain the augmented representations of query p_u and news item p_v.

The loss function of the proposed model is defined as the mean square error between the augmented vectors a_u and a_v and query/item embedding for each sample of which label equals 1. The augmented vectors a_u and a_v are used to fit all positive interactions in the model belonging to the corresponding query or item. The stop gradient strategy is applied to stop the gradient of L_u and L_v from flowing back into p_u and p_v, respectively. The output of the model is the inner product of the query embedding and news item embeddings.

To improve the generalization ability of the model and enable it to learn more transferable features across different news categories, we introduce an additional loss function during the training phase. This loss function aims to minimize the distance between the feature representations of news items from different categories. To achieve this, we first select the largest news category as the reference category. We then calculate the squared Euclidean distance between the feature representations of news items in this category and the feature representations of news items in the other categories. This loss function is added to the overall objective function, with a regularization parameter to control its relative importance.

The final loss function is calculated as the sum of the binary cross-entropy loss, the loss functions for query and news item representations, and the category-aware loss function.

2.3. Training

To train the model for news recommendation, we treat the problem as a binary classification task. We use a random negative sampling technique, where for each positive query-item pair (the label = 1), we randomly select a set of N news items from the news corpus to create negative query-item pairs (the label = 0) for that query. This process results in a dataset with a balance of positive and negative samples. We then use binary cross-entropy loss to train the model by minimizing the error between the predicted scores and the ground truth labels.

3. Experimental Settings

We use the benchmark dataset MIND-small [5] that was collected during 6 weeks (Oct. 12, 2019 to Nov. 22, 2019), with 50k users, 161,013 news, and 156,925 clicks.

Following the standard evaluation methodology s in NRS [5],[6], we conduct a time-based splitting and use the following metrics.

Relevancy metrics: We use Mean Reciprocal Rank (MRR) and F1-score (harmonic mean of precision and recall) to evaluate the relevancy of news recommendations. MRR is the average of the reciprocal rank of the first relevant item in the recommendation list for a set of queries.

To compute MRR, we first calculate the reciprocal rank for each query:

$$\text{Reciprocal Rank} = \frac{1}{\text{rank of first relevant item}} \tag{1}$$

Then, MRR is calculated as the mean of the reciprocal ranks for all queries:

$$\text{MRR} = \frac{\sum_{i=1}^{|Q|} \text{Reciprocal Rank}_i}{|Q|} \tag{2}$$

where $|Q|$ is the number of queries.

Diversity metric: In addition to the relevancy metrics of MRR and F1-score, we also use the GINI index to evaluate the diversity of news recommendations. GINI is a diversity metric that measures the inequality of item distribution across different categories or clusters [7]. In the context of news recommendation systems, GINI reflects the extent to which recommended news articles cover a variety of topics and viewpoints. To calculate GINI, we first group the recommended news articles into different categories based on their content. Then, we compute the proportion of news articles in each category and use this information to calculate the GINI index.

Tradeoff metric: In the context of news recommendation systems, trade-offs are necessary to balance multiple objectives, such as relevancy and diversity. Striking an optimal balance between these two aspects enhances user satisfaction and engagement. We use the trade-off metric, which is the product of the MRR and GINI scores, divided by their sum, multiplied by 2, to measure the balance between relevancy and diversity.

$$\text{tradeoff} = 2 * \frac{(\text{ MRR } * \text{ GINI })}{(\text{ MRR } + \text{ GINI })} \tag{3}$$

The trade-off metric ranges from 0 to 1, with higher values indicating a better balance between relevancy and diversity. By evaluating both relevancy and diversity, we can ensure that our news recommendation system provides users with personalized, engaging, and informative news articles. The trade-off metric ranges from 0 to 1, with higher values indicating a better balance between relevancy and diversity.

We use the following baseline methods for evaluation.

Wide Deep [8], a hybrid model that combines deep neural networks with linear models for recommendation. It is designed to provide both memorization and generalization capabilities.

DKN [6], a knowledge-aware news recommendation method that incorporates knowledge graphs for news representation and recommendation.

LightGCN [9], a simplified Graph Convolutional Network (GCN) for collaborative filtering, aiming to reduce complexity while maintaining performance.

SASRec [10], a self-attentive sequential recommendation model that captures long-range dependencies in user sequences for personalized recommendation.

NeuMF [11], a hybrid model that combines the strengths of Generalized Matrix Factorization (GMF) and Multi-Layer Perceptron (MLP) for better recommendation performance.

We implemented these models in TensorFlow. The embedding dimension and batch size were fixed to 32 and 256. We use the Adam optimizer for training. Other hyperparameters of all models were individually tuned to achieve optimal results to ensure a fair comparison. The dimensions of augmented vectors were both set to $d = 32$, and the tuning parameter γ_1, γ_2 were set to 0.5 and γ_3 to 1. We set top@ k to 5 and 10 as it is normally good practice to retrieve a relatively large number of candidate news items to rank.

4. Results and Discussion

4.1. Performance Evaluation

We evaluate the results for both relevancy and diversity and mainly evaluate the model performance based on the tradeoff score, as it shows a harmonic mean between relevancy and diversity. We expect a good tradeoff score to be above 50% as it is a balancing score between 2 different metrics [2][12]. The results are shown in Table 2.

Method	MRR	F1-score	GINI	Tradeoff	Precision@5	Precision@10
Wide & Deep	0.38	0.50	0.42	0.44	0.55	0.53
DKN	0.40	0.52	0.45	0.46	0.57	0.55
LightGCN	0.42	0.54	0.44	0.48	0.59	0.57
SASRec	0.41	0.53	0.47	0.47	0.58	0.56
NeuMF	0.39	0.51	0.43	0.45	0.56	0.54
Proposed Model	0.48	0.60	0.52	0.54	0.65	0.63

Table 1
Performance of different methods and the current model over the evaluation metrics and precision @5 and 10.

Overall, we observe in Table 1, our proposed approach outperforms all the other methods in terms of relevancy (MRR, F1, precision), diversity (GINI), and trade-off scores on the dataset. This indicates that the current model is capable of providing personalized news recommendations that are not only relevant to individual users but also diverse enough to expose them to a variety of content. The relevancy scores may not be optimal but we achieve balanced tradeoff scores.

The improvements in relevancy, as demonstrated by the higher MRR and F1-score, suggest that the current model can better capture users' preferences and deliver news articles that cater to their interests. This is particularly important in the news recommendation domain, where user engagement largely depends on the presentation of content that aligns with their interests and preferences.

The higher diversity, as represented by the GINI score, shows that the current model is able to recommend a more diverse set of news articles, helping users discover new and unexpected content. This is an essential aspect of a news recommender system, as it encourages users to explore different perspectives and broadens their understanding of various topics.

Moreover, the better trade-off score signifies that the current model strikes an optimal balance between relevancy and diversity, ensuring that users receive a well-rounded set of recommendations. This balance is crucial in maintaining user satisfaction and engagement, as it prevents filter bubbles and echo chambers from forming while still providing users with content that matches their interests.

These results also show that by using negative sampling, we are reducing the selection bias [13]. This is shown by the relatively higher diversity score of our model compared to other items, as all news items in the corpus get a chance to serve as negatives so that the model gets better retrievals towards diversified and long-tail items.
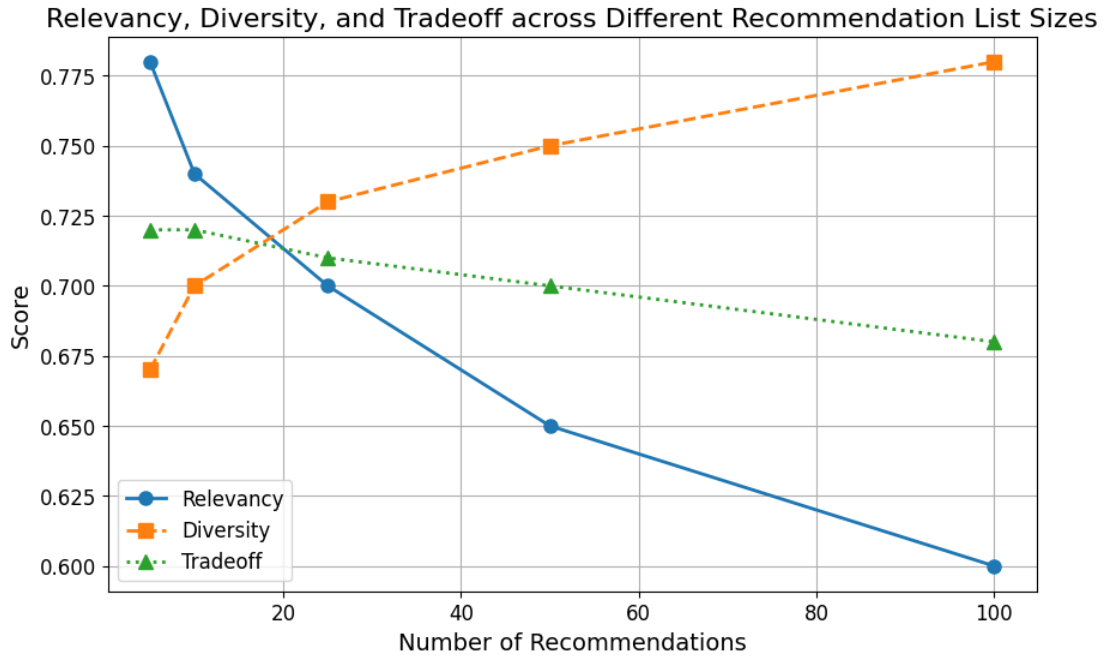
Figure 2: Relevancy and Diversity trade-off of our model

4.2. Relation Between Relevancy and Diversity

Next, we showcase the relevancy-diversity trade-off achieved by our model in Figure 2.

In Figure 2, we observe several trends and relationships between relevancy, diversity, and the tradeoff as the number of recommendations increases. The relevancy scores decrease as the number of recommendations increases. This observation suggests that it becomes more challenging for the recommender system to maintain high relevancy for all items as the recommendation list size grows. This challenge is a common trade-off in recommender systems, as providing more recommendations can increase the likelihood of including diverse items at the cost of potentially lower relevancy for some items.

Conversely, diversity scores increase with the number of recommendations. This trend indicates that the recommender system is capable of providing more diverse recommendations as the list size grows. Including a larger variety of items in the recommendation list can enhance the user experience, as it exposes users to a broader range of content that may match their interests.

The tradeoff scores, which measure the balance between relevancy and diversity, remain relatively stable across different recommendation list sizes. This stability suggests that the recommender system is maintaining a reasonable balance between providing relevant and diverse recommendations. In the given example, the tradeoff scores show a slight decrease as the number of recommendations increases, indicating a minor compromise in balancing relevancy and diversity for larger recommendation lists.

This analysis highlights the challenge of balancing relevancy and diversity in recommender

systems as the number of recommendations increases. It is crucial to find a suitable balance to ensure an optimal user experience, providing recommendations that are both relevant to the user's preferences and diverse enough to expose them to a variety of content.

4.3. Negative Sampling

To analyze the impact of negative sampling on our news recommendation system, we compared the performance of the model with and without negative sampling using precision and recall metrics at a recommendation list size of 10. As illustrated in Figure 3, employing negative sampling significantly improves both precision and recall scores. This finding highlights the importance of incorporating negative samples during the model training process, as it enables the system to provide more accurate and diverse recommendations for users.
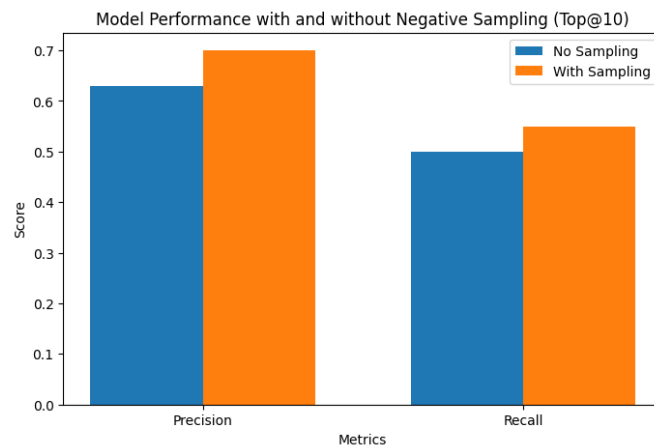

Figure 3: Model Performance with and without Negative Sampling (Top@10)

5. CONCLUSION

In this paper, we present a deep neural network-based architecture designed to model the information interaction between query and news items. Our approach incorporates a variety of features in both the news item and query representations. Additionally, we introduce a loss function that selects distinctive news items across different news categories. We briefly discuss selection bias and demonstrate how using negative sampling can mitigate this bias by including random negatives from the news corpus.

Extensive experiments on a benchmark dataset showcase the superior performance of our proposed method in achieving a balance between accuracy and diversity. For future work, we plan to conduct experiments on additional real-world news datasets and explore the potential of deeper neural networks. We also intend to incorporate more evaluation metrics to assess relevancy, diversity, and novelty in the recommendation results. Furthermore, we aim to address challenges such as mitigating biases [14] and combating fake news [15] in news recommendation systems by employing more advanced deep neural networks.

References

[1] S. Raza, C. Ding, News recommender system: a review of recent progress, challenges, and opportunities, Artificial Intelligence Review (2021) 1–52.

[2] S. Raza, C. Ding, A regularized model to trade-off between accuracy and diversity in a news recommender system, in: 2020 IEEE International Conference on Big Data (Big Data), IEEE, 2020, pp. 551–560.

[3] R. Wang, Z. Zhao, X. Yi, J. Yang, D. Z. Cheng, L. Hong, S. Tjoa, J. Kang, E. Ettinger, H. Chi, Improving relevance prediction with transfer learning in large-scale retrieval systems, in: Proceedings of the 1st Adaptive & Multitask Learning Workshop, 2019.

[4] J. Yang, X. Yi, D. Zhiyuan Cheng, L. Hong, Y. Li, S. Xiaoming Wang, T. Xu, E. H. Chi, Mixed negative sampling for learning two-tower neural networks in recommendations, in: Companion Proceedings of the Web Conference 2020, 2020, pp. 441–447.

[5] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, et al., Mind: A large-scale dataset for news recommendation, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3597–3606.

[6] H. Wang, F. Zhang, X. Xie, M. Guo, Dkn: Deep knowledge-aware network for news recommendation, in: Proceedings of the 2018 world wide web conference, 2018, pp. 1835–1844.

[7] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, P. Jiang, BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19, Association for Computing Machinery, 2019, pp. 1441–1450. doi:10.1145/3357384.3357895.

[8] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, L. Anderson, M. Pham, P. Ravichander, J. Pennington, et al., Wide & deep learning for recommender systems, in: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, 2016, pp. 7–10.

[9] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, M. Wang, Lightgcn: Simplifying and powering graph convolution network for recommendation, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2020.

[10] W.-C. Kang, J. McAuley, Self-attentive sequential recommendation, in: Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), IEEE, 2018, pp. 197–206.

[11] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2017, pp. 173–182.

[12] S. Raza, C. Ding, Deep Neural Network to Tradeoff between Accuracy and Diversity in a News Recommender System, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, 2021, pp. 5246–5256.

[13] S. Caton, C. Haas, Fairness in machine learning: A survey, arXiv preprint arXiv:2010.04053 (2020).

[14] S. Raza, J. Reji, C. Ding, Dbias: Detecting biases and ensuring Fairness in news articles, International Journal of Data Science and Analytics (2022).

[15] S. Raza, C. Ding, Fake news detection based on news content and social contexts: a

transformer-based approach, International Journal of Data Science and Analytics 13 (2022) 335–362.