systems

MDPI

*Article*

# Evaluating User Engagement in Online News: A Deep Learning Approach Based on Attractiveness and Multiple Features

**Guohui Song [1,2], Yongbin Wang [2,*], Xiaosen Chen [1], Hongbin Hu [2] and Fan Liu [3]**

[1] School of Computer and Cyber Sciences, Communication University of China, Beijing 100024, China; sgh3352952@cuc.edu.cn (G.S.); chenxiaosen_cxs@cuc.edu.cn (X.C.)
[2] State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China; hhb@cuc.edu.cn
[3] School of Journalism, Communication University of China, Beijing 100024, China; cucliufan@cuc.edu.cn
* Correspondence: ybwang@cuc.edu.cn

**Abstract:** Online news platforms have become users' primary information sources. However, they focus on attracting users to click on the news and ignore whether the news triggers a sense of engagement, which could potentially reduce users' participation in public events. Therefore, this study constructs four indicators by assessing user engagement to build an intelligent system to help platforms optimize their publishing strategies. First, this study defines user engagement evaluation as a classification task that divides user engagement into four indicators and proposes an extended LDA model based on user click–comment behavior (UCCB), using which the attractiveness of words in news headlines and content can be effectively represented. Second, this study proposes a deep user engagement evaluation (DUEE) model that integrates news attractiveness and multiple features in an attention-based deep neural network for user engagement evaluation. The DUEE model considers various elements that collectively determine the ability of the news to attract clicks and engagement. Third, the proposed model is compared with the baseline and state-of-the-art techniques, showing that it outperforms all existing methods. This study provides new research contributions and ideas for improving user engagement in online news evaluation.

**Keywords:** user engagement; text classification; deep learning; communication studies

check for updates

**Citation:** Song, G.; Wang, Y.; Chen, X.; Hu, H.; Liu, F. Evaluating User Engagement in Online News: A Deep Learning Approach Based on Attractiveness and Multiple Features. *Systems* **2024**, *12*, 274. https://doi.org/10.3390/systems12080274

Academic Editors: Bruno Oliveira, Óscar Oliveira and Telmo Matos

Received: 31 May 2024
Revised: 21 July 2024
Accepted: 27 July 2024
Published: 30 July 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

In recent years, online news platforms have become users' primary information sources [1]. A 2023 survey by the Pew Research Centre shows that most American adults (86%) regularly obtain their news from a smartphone, computer, or tablet. Online news platforms are the most popular news source [2]. In China, Toutiao and Tencent News emerged as the most popular news platforms, with a user base of around 300 million [3]. Online news platforms help those not actively seeking news information to reach news stories while also increasingly acting as intermediaries between traditional publishers and users and can significantly increase users' political engagement [4]. Therefore, news platforms attract users to read news, positively affecting the platforms' economic benefits and the public's political participation.

Platforms like Tencent New and Toutiao use personalized recommendations and other methods to attract users [5,6]. These platforms combine content such as news for viewing on mobile devices or websites and are called news aggregators. These platforms have weaker social attributes than social media platforms, and their news content is usually sourced from professional news organizations or screened as reliable sources. To increase user engagement and stickiness, these platforms often employ deep learning techniques to suggest news articles that are potentially popular or interesting [7,8]. These techniques usually attract clicks or cater to users' interests based on their historical behavior. However,

they focus only on attracting users to click, ignoring whether the news triggers user discussion. They regularly recommend that such news damages the user experience and reduces users' engagement with political issues. For example, a news platform pushing politainment news will likely attract many user clicks and lack in-depth discussion. Therefore, optimizing publishing strategies and designing more engaging content can help increase user engagement with public issues and contribute to society's development [9]. However, given the vast amount of news available, human editors cannot evaluate news's ability to attract user clicks and engagement. As a result, news platforms need an intelligent system to evaluate the news before release, enhancing platform competitiveness and user engagement by attracting users. Thus, this work will construct user engagement indicators and build an intelligent system to help news platforms optimize their publishing strategies.

Several research directions exist on building an intelligent system for evaluating news's ability to attract users. The three main research areas are popularity prediction, clickbait detection, and other methods. News popularity is often considered a market indicator of attractiveness rather than a generally accepted social standard since clickbait and fake news can also be popular. Therefore, the object of these studies tends to be platforms such as Toutiao, Tencent News, or news platforms with official backgrounds, where news content is usually sourced from professional news organizations or screened and reliable sources and tends to public issues. In these studies, popularity is an attractiveness indicator of online news articles, as it reflects the attention they receive, which can be measured by the total number of clicks or other user feedback [10–15]. However, the ability of news to generate user engagement and discussion is not emphasized. Clickbait may attract users to click news with a gimmicky headline, and clickbait detection is commonly used to identify misleading headlines [16–18]. These studies are more concerned with the credibility of the news [19,20]. Therefore, further research is needed to comprehensively evaluate the ability of news to attract user clicks and engagement.

As news platforms such as Toutiao, Tencent News, and other news aggregators often push public news in political and economic fields, they can, to some extent, reflect the public interest and value attributes of news. Therefore, this study takes these platforms as the research object and proposes a method to evaluate news, which will help news platforms attract users while promoting users' discussion at public events. To achieve such a goal, we defined four indicators in this study to define the ability of news to attract user clicks and comments.

This paper proposes a deep learning model named deep user engagement evaluation (DUEE) to build an intelligent online news evaluation system. The DUEE consists of two main parts: the representation of word attractiveness in news headlines and content and a deep learning model that incorporates this representation. This model assesses news engagement based on the number of clicks and comments on news articles. The news headline is a significant factor that triggers user clicks, while the news article's content is the main factor that triggers user comments. First, this paper proposes an extended latent Dirichlet allocation (LDA) [21] model based on user click–comment behavior (UCCB) to represent the attractiveness of news articles in DUEE and the attractiveness as a readability feature of the news. Then, the DUEE divides the information that affects user click and comment behavior into explicit and implicit parts. The explicit part consists of news headlines, source, and publish time, the main factors that trigger user click behavior. The implicit part is news content, which mainly triggers user comment behavior. The paper represents the semantics of news headlines using a bidirectional gated recurrent unit (Bi-GRU), which embeds the attractiveness of words and attention mechanisms. The news content is the main reason for triggering user comments. Since it is a long sentence, this paper uses hierarchical attention networks (HAN) [22] with embedded word attractiveness to obtain its semantic representations. Finally, the DUEE employs an attention mechanism to dynamically fuse the explicit and implicit parts of information, considering their influence on news evaluation results.

This study makes several theoretical and technical contributions. In summary, the paper's main contributions are as follows:

(1) This study proposes an extended LDA model based on users' click–comment behavior. The model can effectively represent the attractiveness of words in news headlines and content. To a certain extent, the attractiveness of words constitutes the news's readability and is an essential factor in attracting users to click and comment.
(2) This study proposes a deep learning model, DUEE, that integrates news headlines, content, meta-features, and attractiveness. More importantly, the DUEE model integrates word attraction in news headlines and content representations through attention units. The DUEE model considers various elements of news that collectively determine the ability of the news to attract clicks and engagement.
(3) This study verifies the effectiveness of the DUEE as a whole, provides empirical evidence for news attractiveness in this task, and provides a new idea for user engagement evaluation. The proposed model and indicators can better help news platforms adjust their news release strategies to promote user engagement.

The rest of this paper is organized as follows. The next section summarizes the current work related to this study. Section 3 defines four indicators to evaluate news engagement and describes ways to automate data labeling. Section 4 formally introduces the DUEE model for evaluating user engagement. Section 5 presents the experimental results and detailed analysis. Section 6 summarizes this paper.

## 2. Related Works

Since this study aims to propose methods for building intelligent systems, this section will investigate related studies from the model design perspective.

### *2.1. News Popularity Prediction Methods*

Since popularity can be quantified well and is closely related to the revenue of news platforms, there are many studies on popularity prediction. Some studies predict the popularity of news articles before their release based on their headlines, content, or other features. These studies focus on the popularity of news articles based on the number of clicks or other feedback they receive. Articles with more clicks are considered to be of high popularity. Voronov et al. [8] concluded that the news headline is the primary factor that triggers users to click. Therefore, this study used only news headlines and bi-directional long short-term memory (Bi-LSTM) neural networks to predict news popularity. Some studies have linked news popularity to named entities in headlines. Yang et al. [10] argued that news popularity is related to named entities in news headlines. They proposed a named entity topic model based on LDA to extract textual factors that can contribute to news popularity. The popularity of news articles is predicted by learning the popularity gain matrix of each named entity. In addition to news headlines, some studies have considered more comprehensive factors. The deep model proposed by Xiong et al. [11] considers not only news headlines, content, and meta-features but also the attractiveness and timeliness of news headlines. Liao et al. [12] proposed a deep learning model that fuses time series, text, and meta-features. Experimental results show that the multi-feature fusion approach can effectively improve the popularity prediction performance. The above studies are mainly based on the textual characteristics of news. Arora et al. [13] found that images in news articles also affected the popularity of news and proposed an integrated learning model using text features, meta-features, and image features. Experimental results show that this design helps improve the model's robustness. Omidvar et al. [14] proposed a deep learning model to evaluate the quality of news headlines based on some indicators. They believed that the number of clicks on the news article and the time users spend on the news content determine the quality of the headlines. Yang et al. [15] proposed a method for predicting news popularity using mining writing style.

The problem with using news popularity is that news headlines mainly trigger user click behavior, and whether news content has received user attention cannot be effectively

evaluated. Unlike popularity prediction, in this study, we define indicators that consider not only the ability of news headlines to attract users to click on them but also the ability of news content to attract users to engage in commenting.

### 2.2. Clickbait and Fake News Detection Methods

Clickbait detection is an important research direction to identify gimmicky news headlines that often do not match their content. Clickbait detection is generally considered a classification problem. There is much academic research on clickbait detection. Wang et al. [16] compiled a dataset containing up to 20,896 Chinese clickbait news articles and proposed a contextual semantic representation-based clickbait detection approach employing transfer learning. Liu et al. [17] constructed a Chinese WeChat clickbait dataset and proposed an effective deep method for clickbait detection by integrating semantic, syntactic, and auxiliary information. Pujahari et al. [18] proposed a hybrid categorization technique for separating clickbait and non-clickbait articles by combining features, sentence structure, and clustering. The main idea of these studies is to capture the consistency between news headlines and news content. A large discrepancy between the news headline and the content means the article is clickbait.

Some studies focus on detecting fake news, which is considered low-quality due to the spread of false information. Zrnec et al. [23] examined how users perceive fake news in shorter paragraphs on individual information quality dimensions. The results reveal that domain knowledge, education, and personality traits can be tools for detecting fake news. Lin et al. [24] compared six sets of expert ratings and found that they generally correlated highly. Then, they performed imputation and principal component analysis to generate a set of aggregate ratings. The results suggest that experts typically agree on the relative quality of news domains, and the aggregate ratings offer a powerful research tool for evaluating the quality of news consumed or shared and the efficacy of misinformation interventions. Mosallanezhad et al. [25] believed that some auxiliary information, such as user comments and user–news interaction, is vital in identifying fake news. They incorporated auxiliary information into a reinforcement learning-based model to address the diverse nature of news domains and expensive annotation costs.

Methods for detecting clickbait and fake news focus on identifying news headlines that are gimmicky or misleading. Although clickbait and fake news detection differ from this paper's research objectives, its multi-feature fusion approach inspires this study. In addition, as the target of this study tends to be platforms such as Toutiao, Tencent News, or news platforms with official backgrounds, the news content of these platforms usually comes from professional news organizations or screened and reliable sources and tends to be public issues. Therefore, the likelihood of clickbait and fake news is low.

### 2.3. Other News Evaluation Methods

Some studies have attempted to evaluate news from the perspective of feature extraction. In the previous study, we [26] proposed a deep learning model that integrates explicit and implicit news information to evaluate user engagement. In this study, the explicit information includes the headline, source, and time of publication of the news, which attracts users to click. The implicit information refers to the news article's content, which attracts users to comment. This work provides a new idea for user engagement evaluation. Alam et al. [27] scored the news articles by metadata, news content, and entity extraction. Wu et al. [28] developed an encoder–decoder model using a large training corpus to generate attractive news headlines by reducing the importance of sentences that are easy to classify. In this part of the study, the features that impact the news headlines are identified using various feature extraction methods. Romanou et al. [29] developed an evaluation system that automatically collects real-time contextual information about news articles and considers indicators of validity and trustworthiness. The indicators include discussions on social media, news content, and sources. Kim et al. [30] proposed a method for measuring the click value of words and analyzing how temporal trends

and linguistic attributes affect the number of clicks. Experiments show that the technique better identifies high-click-value words. Although some studies wanted to build systems from different perspectives, they did not consider user engagement. In the previous study, we [26] proposed a deep learning model integrating explicit and implicit news information to evaluate user engagement. However, we did not consider the role of word attractiveness on user engagement.

## 3. Data Collection and Labelling

In order to construct a model using credible data, this study constructed a dataset containing 12,308 online news articles from the famous Chinese online news platform "Toutiao" (www.toutiao.com, accessed on 31 May 2024). These news articles were published on the homepage of Toutiao between 5 November 2020, and 15 June 2022. The dataset contains attributes such as news source, content, headline, publication time, number of clicks, and comments. Since these news articles come from the top of the homepage and are visible to all users, they are more suitable for evaluating the user engagement of news articles. In addition, the content of these news articles is often political, economic, diplomatic, and other public issues, and, therefore, better reflects the public interest and value.

Since this work is a typical classification task, data with user engagement labels are needed. This study uses a new method to label the data automatically [26]. The method uses the number of clicks and comments on the news to classify user engagement into four classes. The number of clicks represents the breadth of news dissemination, while the number of comments is more reflective of the publicity effect of the news. We divide the information influencing user clicking and commenting behavior into explicit and implicit parts. The explicit part consists of the news headline, source, and publish time, which are the main factors that trigger user click behavior. The implicit part is the news content, which triggers user commenting behavior. Therefore, the indicators of user engagement in this study mainly consider the ability of the news headlines, content, and other features to attract users to click and comment. Then, we scaled each news article's clicks and comments to the [0, 1] by the interquartile range normalization method. As shown in Figure 1, we define four user engagement indicators. The meaning of each indicator is described in Table 1.

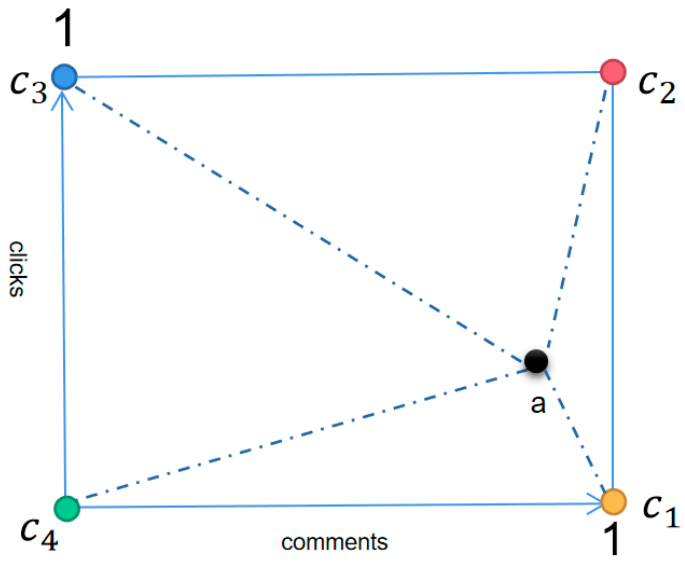



**Figure 1.** Four user engagement indicators [26].

Based on the above definition of user engagement indicators, the labeling method is shown in function (1).

$$\begin{bmatrix} P_{a,C_1} \\ P_{a,C_2} \\ P_{a,C_3} \\ P_{a,C_4} \end{bmatrix} = Softmax \begin{pmatrix} \sqrt{2} - \sqrt{x_a^2 + (1-y_a)^2} \\ \sqrt{2} - \sqrt{(1-x_a)^2 + (1-y_a)^2} \\ \sqrt{2} - \sqrt{(1-x_a)^2 + y_a^2} \\ \sqrt{2} - \sqrt{x_a^2 + y_a^2} \end{pmatrix} \quad (1)$$

**Table 1.** Definition of the user engagement indicators.

| User Engagement Indicators | Description |
|---|---|
| $C_1$ | High comments and low clicks. News articles close to this indicator represent news content that generates more attention, but the implicit information is less capable of triggering users to click. |
| $C_2$ | High comments and low clicks. News articles close to this indicator show that the implicit and explicit have enough ability to attract users to click and comment. |
| $C_3$ | High clicks and low comments. News articles close to this indicator show that explicit information attracts users, but their content does not. |
| $C_4$ | Low clicks and low comments. News articles close to this indicator show that neither explicit nor implicit information effectively attracts users. |

## 4. The Proposed Model (DUEE)

To address the limitations of existing studies, this work proposes DUEE, a deep learning approach to user engagement evaluation. In the model, we consider a variety of components that collectively determine the ability of news to attract user clicks and comments. In contrast to traditional deep learning text representation, DUEE integrates word attractiveness into the representation of news headlines and content. In addition, DUEE employs an attention mechanism to integrate both the explicit and implicit information representation. The architecture of DUEE consists of two main parts: an explicit information representation learning module that integrates four types of input, including news source, publish time, new headline, and attentiveness, and an implicit information representation learning module based on HAN, LDA, and attractiveness. An attention unit connects these two modules to learn their combinations. Then, we use the Softmax activation function to classify the user engagement, as shown in Figure 2.

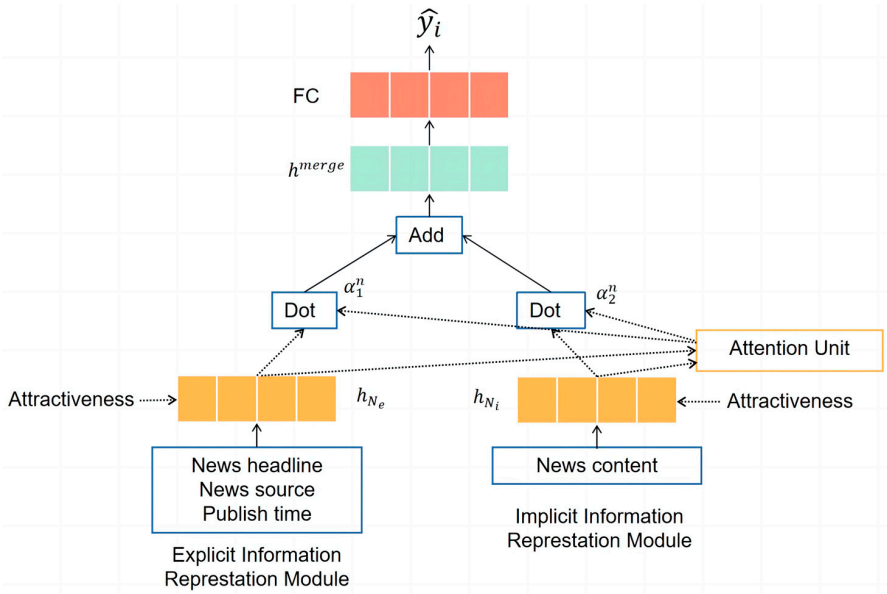



**Figure 2.** The architecture of DUEE.

### *4.1. Attractiveness Representation*

Readability is a crucial element of effective communication in journalism. It pertains to the simplicity with which a text can be read, comprehended, and processed by its intended user [31]. The choice of words and sentence composition in news is crucial to readability and can influence user engagement [32]. The attractiveness of a news article is an essential factor in triggering users to click and comment. The collection of the attractiveness of individual words in the news headline and content can determine it. In addition, the attractiveness of words depends on context. Users' click and comment behavior is an effective indicator of news attractiveness. Therefore, this study proposes a topic model based on users' clicking and commenting behavior (UCCB) to evaluate the attractiveness of words in news headlines and content, respectively. The attractiveness as a readability feature of news will be part of the input to the DUEE model. As shown in Figure 3, this study employs a probabilistic model to determine the attractiveness of words in a news article, extending the generation process of traditional LDA topic models by integrating user behavior. The topic model introduces an observable variable $c$ based on users' click–comment behavior and a latent variable $\eta$ indicating the attractiveness of each word in a given topic. The latent variable determines the likelihood that a user will click and comment on a news article containing the word. When a user sees a news article $d$, it may trigger a click action $r$. Depending on whether the user generates a comment behavior after the click, the click–comment result can be represented as $c_d^r = 1$ or $c_d^r = 1$.

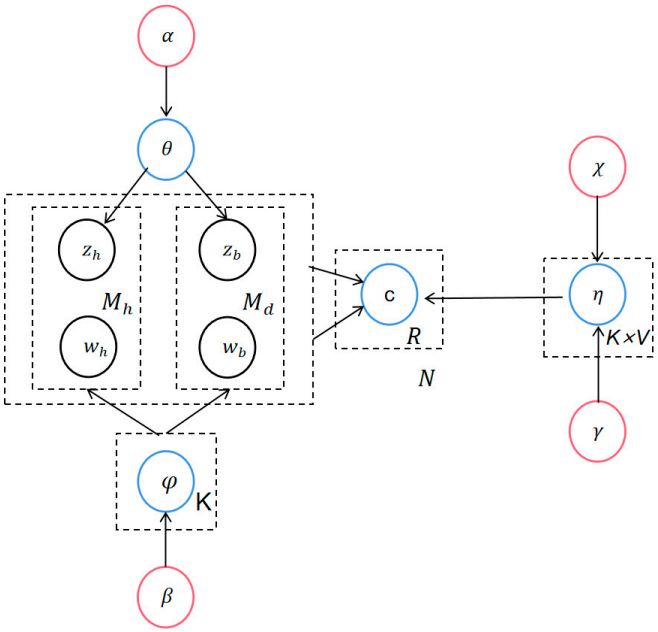



**Figure 3.** The architecture of the UCCB topic model.

Thus, the user's clicking–commenting behavior can be considered an independently distributed Bernoulli trial. In this case, the user's clicking and commenting behavior on a news article follows a binomial distribution. Since the conjugate distribution of the binomial distribution is the *Beta* distribution, to exploit the conjugacy of the binomial distribution in posterior inference, we model the prior distribution of $\eta$ as a *Beta* distribution. Thus, the probability of commenting behavior occurring after each word in a news headline and content is clicked by a user is modeled by a variable $c$ representing a Bernoulli test, and the *Beta* distribution gives the prior distribution of the probability of success of the test.

In contrast to the LDA topic generation model, our topic model also applies the Dirichlet polynomial conjugate [21] to generate the distribution of topics for each news article and the distribution of words in each topic, as well as including the process of generating the user commenting behavior $c$. The topic distributions generated by the model

can be used as supplementary information for news content representation. In the topic model, the latent variable $z$ is divided into two parts, $z_h$ representing news headlines and $z_b$ representing news content. The meaning of each variable is described in Table 2.

**Table 2.** Definition of variables in the topic model.

| Variable Name | Description |
|---|---|
| $K$ | The number of topics. |
| $V$ | Number of words in the dataset. |
| $N$ | Number of news articles in the dataset. |
| $\theta$ | Topic distribution of news articles. |
| $\varphi$ | Topic distribution of words. |
| $\eta$ | The probability that a word is attractive to a particular topic. |
| $z_h$ | Topic distribution of words in the news headlines. |
| $w_h$ | Words in news headlines. |
| $z_b$ | Topic distribution of words in the news content. |
| $w_b$ | Words in news content. |
| $c$ | Probability click–comment behavior. |
| $M_h$ | Number of words in news headlines. |
| $M_b$ | Number of words in news content. |
| $R_d$ | The number of clicks on a news article $d$. |

The process of generating this topic model is as follows:

1. For each topic $k \in K$, a word distribution $\varphi_k \sim Dir(\beta)$ is generated where each topic $k$ has a word distribution $\varphi_k = \{\varphi_{k1}, \varphi_{k2}, \ldots, \varphi_{kV}\}$.
2. Each topic–word pair $(z, w) \in K \times V$ generates a distribution of values $\eta_{z,w} \sim Beta(\chi, \gamma)$ of user comments.
3. The following calculations were made for each news article:
   (1) Generate a topic distribution $\theta_d \sim Dir(\alpha)$ for each news article, where each news article has a topic distribution $\{\theta_{d1}, \theta_{d2}, \ldots, \theta_{dK}\}$.
   (2) For each word in the news content, a topic distribution $z_b^{id} \sim Mult(\theta_d)$ is generated, and a word distribution $w_b^{id} \sim Mult\left(\varphi_{z_b}^{id}\right)$ is generated.
   (3) For each word in a news headline, a topic distribution $z_h^{jd} \sim Mult(\theta_d)$ and a word distribution $w_h^{jd} \sim Mult\left(\varphi_{z_h}^{jd}\right)$ are generated..
   (4) For a news article, click behavior $r \in R_d$ generates a distribution of comment behavior $c_d^r \sim Bin\left(1, \eta_{z_d^r, w_d^r}\right)$.

This topic model extends the classical LDA model by introducing word attractiveness distributions. LDA models are generally solved using Gibbs sampling or variational inference EM algorithms. Since Gibbs sampling can be parallelized, it is well-suited for distributed training. Therefore, this study utilizes Gibbs sampling to construct a Markov chain to converge to the target distribution. Transitions between neighboring states in a Markov chain follow a simple rule of sequential sampling from the distribution of the current values of all other variables to reach the next state. In order to realize a posteriori inference, this study uses the collapsed Gibbs sampling method to obtain the attractiveness distribution of words [33] with the following joint probability formula:

$$\begin{aligned} p(w_h, w_b, z, c, \theta, \varphi, \eta) = p(\varphi|\beta)\, p(\eta|\gamma, \chi) \prod_{d=1}^{N} p(\theta_d|\alpha) \Big(\prod_{i=1}^{M_b} p\left(w_b^{id} \middle| z_b^{id}\right) p\left(z_b^{id} \middle| \theta_d\right)\Big) \\ \Big(\prod_{j=1}^{M_h} p\left(w_h^{jd} \middle| z_h^{jd}\right) p\left(z_h^{jd} \middle| \theta_d\right)\Big) \Big(\prod_{r=1}^{R} p\left(c_d^r \middle| \eta_{z_d^r, w_d^r}\right) p(w_d^r|\eta)\Big) \end{aligned} \tag{2}$$

In this work, we use Dirichlet polynomial covariance to compute the conditional distributions of $w_h$ and $w_b$ as follows:

$$p\left(w_b^{id}\middle|w_b^{-id},\ w_h, z_b^{id}=z\right)=\frac{n_{-i,z}^{(w_{id})}+\beta}{\sum_{v=1}^{V}\left(n_{-i,z}^{(v)}+\beta\right)} \tag{3}$$

$$p\left(w_h^{jd}\middle|w_h^{-jd},\ w_b, z_h^{jd}=z\right)=\frac{n_{-j,z}^{(w_{jd})}+\beta}{\sum_{v=1}^{V}\left(n_{-j,z}^{(v)}+\beta\right)} \tag{4}$$

where $z_b^{id}$ represents the topic of the word $w_{id}$ in the news content; $z_h^{jd}$ represents the topic of the word $w_{jd}$ in the news headline; and $n_{-i,z}^{(.)}$ and $n_{-j,z}^{(.)}$ represent the number of words in the news content and headline that do not belong to topics $z_i$ and $z_j$, respectively.

Given words $w_d^{id}$ and $\eta$ in the news content, the conditional probability of the topic $z_b^{id}$ is proportional to the number of times the topic appears in news $d$. Combined with the click–comment data, the conditional distribution of $z_b^{id}$ is calculated as shown below:

$$p\left(z_b^{id}=z\middle|rest\right)\propto\left(n_{zd}^{-id}+\alpha\right)\times p\left(w_b^{id}\middle|w_b^{-id},\ w_h, z_b^{id}=z\right)\prod_r p\left(c_d^r\middle|\ \eta_{z,w_b^{id}}\right) \tag{5}$$

where $n_{zd}^{-id}$ denotes the number of times topic $z$ is assigned in news d without regard to $z_b^{id}$ and $\prod_r p\left(c_d^r\middle|\ \eta_{z,w_b^{id}}\right)$ represents the joint conditional probability of each comment associated with the word $w_b^{id}$.

Given the words $w_h^{jd}$ and $\eta$ in a news headline, the conditional distribution of $z_h^{jd}$ is proportional to the number of occurrences of that topic in the news $d$ multiplied by the conditional probability of that topic $w_h^{jd}$. The conditional distribution of $z_h^{jd}$ is calculated as shown below:

$$p\left(z_h^{jd}=z\middle|rest\right)\propto\left(n_{zd}^{-jd}+\alpha\right)\times p\left(w_h^{jd}\middle|w_h^{-jd},\ w_b, z_h^{id}=z\right)\prod_r p\left(c_d^r|\ \eta_{z,w_h^{jd}}\right) \tag{6}$$

where $n_{zd}^{-jd}$ denotes the number of times topic $z$ has been assigned in the news article $d$ without taking $z_h^{jd}$ into account.

Sampling of $z_b^{id}$ and $z_h^{jd}$ involves attractiveness distributions that are continuously updated with topic probabilities in each sampling iteration. The conditional distribution of $\eta$ is represented by the conjugate prior property of the *Beta* and binomial distributions, and the posterior distribution of $\eta$ follows the *Beta* distribution, which is represented as follows:

$$\eta_{z,w}|z,w,c\ \sim\ Beta\left(n_{z,w}^1+\gamma, n_{z,w}^0+\chi\right) \tag{7}$$

where $\chi$ and $\gamma$ are the initial parameters of the prior distribution $\eta_{z,w}$ and $n_{zw}^{(.)}$ is the number of observed comments.

The attractiveness distribution of each word in a given topic can be calculated based on the above calculation process. The final generated attractiveness scores for the $i$-th word and the $j$-th word in the news headline and news content of the $d$-th news article are calculated as shown in Equations (8) and (9):

$$g_{b,id}=\sum_{z=1}^{K}\theta_{dz_b}\varphi_{z_b w_i}\eta_{z_b w_i} \tag{8}$$

$$g_{h,jd} = \sum_{z=1}^{K} \theta_{dz_h} \varphi_{z_h w_j} \eta_{z_h w_j} \tag{9}$$

where $g_{h,jd}$ and $g_{b,id}$ are the attractiveness scores of the words in the news headlines and the news contents, the larger the value of $g_{h,jd}$ and $g_{b,id}$, the higher the attractiveness of the word. The study does not determine attractiveness by the cumulative number of news comments. However, by the ratio of comments to the number of clicks, the attractiveness scores of words remain relatively stable over time.

### *4.2. Explicit Information Representation*

Figure 4 illustrates that the headline, source, and publication time of the news article are inputs to the explicit information representation module. The left part represents the semantics of news headlines. The explicit information representation module considers various features of the news that are important in triggering a user's clicking behavior.

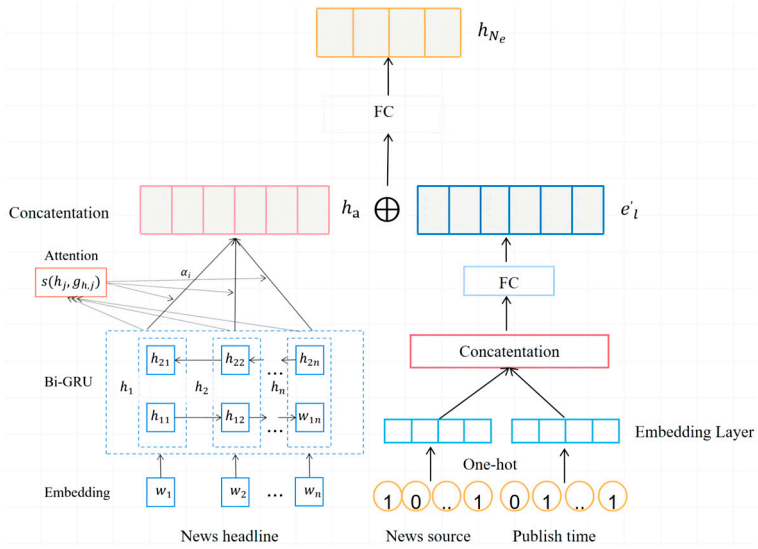



**Figure 4.** The architecture of explicit information representation.

First, the words in the news headlines are represented by a pre-trained Word2Vec model [34]. Then, Bi-GRU was used to obtain the semantic representation of the $j$-th word in the news headline as $h_j = \left[\overrightarrow{h_j}, \overleftarrow{h_j}\right]$, where $\overrightarrow{h_j}$ and $\overleftarrow{h_j}$ represent the forward and backward hidden vectors in Bi-GRU, respectively. Finally, the attractiveness features of the words need to be embedded in the news headline semantics. The potential semantic attractiveness score of the $j$-th word in the news headline is denoted as $h_j$, $g_{h,j}$. The importance function of the $j$-th word in the news headline is defined as follows:

$$s\left(h_j, g_{h,j}\right) = u^T tanh\left(w_h h_j + w_g g_{h,j} + b_h\right) \tag{10}$$

where $w_h$, $w_g$,$u$, and $b_h$ are learnable parameters.

The semantic representation $h_a$ of a news headline is derived from a weighted summation of the potential semantics of the words. The weight $\alpha_j$ of each word is derived using the Softmax function, and the final news headline representation $h_a$ is calculated as shown in Equation (11):

$$\begin{cases} \alpha_j = \frac{exp\left(s\left(h_j, g_{h,j}\right)\right)}{\sum_j exp\left(s\left(h_j, g_{h,j}\right)\right)} \\ h_a = \sum_j \alpha_j h_j \end{cases} \tag{11}$$

In addition, since source credibility is a crucial factor in attracting users [35], we have considered news sources in this module. Meanwhile, publish time measures different

time effects on users' click behavior. Therefore, we consider these factors in the explicit information representation module. These features are firstly embedded in homologous dense vectors. We then concatenate the embedding vectors as $e_l$ and apply fully connected (FC) layers to combine all the features into $e'_l$:

$$e'_l = tanh(W_l \times e_l + b_l) \tag{12}$$

where $W_l$ and $b_l$ are learnable parameters.

To obtain a final representation of the explicit information, we concatenate $h_a$ and $e'_l$ and feed the result into a FC layer. The FC layer uses the rectified linear unit (ReLu) as the activation function. The calculation process is shown below:

$$h_{N_e} = FC([e'_l; h_a]) \tag{13}$$

### *4.3. Implicit Information Representation*

Since news content is the main factor that triggers user comments and engagement, this work uses news content as input to the implicit information representation module. Since news content is usually long text, to extract the semantic information of news content fully, this chapter adopts the HAN structure to obtain the semantic representation of news content. As shown in Figure 5, the news content consists of sentences, and the $i$-th sentence obtains the semantic representation $s_i$ through the computational process shown in Figure 5. It is important to note that the implicit information module considers the news topic to represent its relevance characteristics.

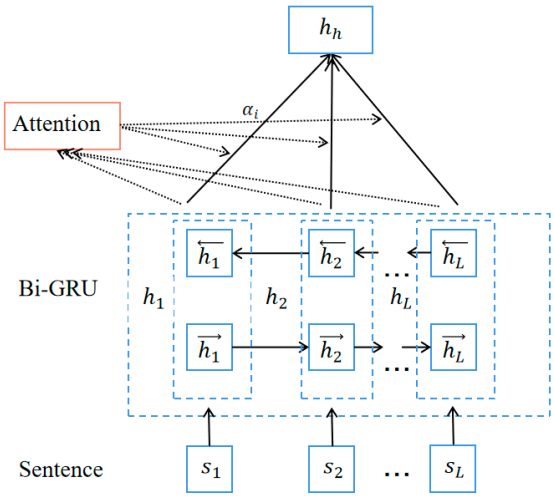



**Figure 5.** News content representation based on HAN structure.

The difference with HAN networks is that word attractiveness $g_{b,i}$ is embedded in the semantic representation of each sentence. The semantic representation of each sentence is calculated according to Equations (10) and (11) in Section 4.2. To obtain the semantic representation of the news content, we input the semantics $s_i$ of each sentence in the news content into the Bi-GRU network to obtain the potential semantic representation of the $i$-th sentence in the news content as $h_i = \left[\overrightarrow{h_i}, \overleftarrow{h_i}\right]$, where $\overrightarrow{h_i}$ and $\overleftarrow{h_i}$ represent the forward and backward hidden vectors in Bi-GRU, respectively. Then, we use the attention mechanism to compute the semantic representation of the news content; the computation process is shown in Equation (14).

$$\begin{cases} s(h_i) = v^T tanh(w_b h_i + b_b) \\ \alpha_i = \frac{exp(s(h_i))}{\sum_i s(h_i)} \\ h_h = \sum_i \alpha_i h_i \end{cases} \tag{14}$$

where $w_b$ , $v$, $b_b$ are the parameters to be trained in the model; $\alpha_i$ is the attention weight of the $i$-th sentence; and $h_h$ is the news content representation embedded with attractiveness.

Relevance is an essential criterion for evaluating news content, ensuring that the topics provided are relevant to the user's attention [36]. Therefore, considering the impact of different news topics on users' commenting behavior, we exploit the news topic distribution trained in Section 4.1, which can complement the semantics of news content. The final semantic representation of the news content is shown below:

$$h_{N_i} = FC\left(\left[h_h; h_p\right]\right) \tag{15}$$

where $h_p$ is the topic distribution of a news article; the FC layer employs ReLu as an activation function.

### *4.4. Attention Unit and Model Output*

Explicit and implicit information impacts users' clicking and commenting behaviors differently. Therefore, we employ the attention mechanism to learn the weights of the two parts automatically, and the calculation process is as follows:

$$\begin{cases} \alpha_1^n = q_n^T \times tanh(V_n \times h_{N_e} + b_n) \\ \alpha_2^n = q_n^T \times tanh\left(V_n \times h_{N_i} + b_n\right) \\ \overset{\sim}{a}_i^n = \frac{exp\left(\alpha_i^n\right)}{\sum_i exp\left(\alpha_i^n\right)}, (i = 1, 2) \end{cases} \tag{16}$$

$$h^{merge} = \overset{\sim}{a}_1^n h_{N_e} + \overset{\sim}{a}_2^n h_{N_i} \tag{17}$$

where $\overset{\sim}{a}_i^n$ is the attention weight assigned to the two parts. $q_n$, $V_n$, $b_n$ are learnable parameters. $h^{merge}$ is the result of the semantic representation of news article.

In the output layer, we input $h^{merge}$ to the FC layer, and the probability distribution of user engagement $P_C = \left\{P_{C_1}, P_{C_2}, P_{C_3}, P_{C_4}\right\}$ is output through the Softmax function. Finally, the argmax function is used to output the class to which the news belongs. The calculation process is as follows:

$$\begin{cases} P_C = Softmax(W_h \times h^{merge} + b_h) \\ \hat{y} = argmax(P_C) \end{cases} \tag{18}$$

where $W_h$ and $b_h$ are learnable parameters; $\hat{y}$ is the user engagement class of the model output.

In order to measure the classification effect of the mode, we use the cross-entropy loss function in the training phase. Since there is an imbalance in each class in the dataset, models trained using imbalanced datasets perform poorly in the weak class. Therefore, we leverage the re-weighting method to increase the weight loss of weak classes and alleviate the problem of uneven data distribution [37]. The cross-entropy loss function considering the class weights is defined as:

$$\begin{cases} E_{n_{y_i}} = \frac{1-\beta^{n_{y_i}}}{1-\beta} \\ L = -\sum_{i=1}^{M} \frac{1}{E_{n_{y_i}}} y_i log(P_C) \end{cases} \tag{19}$$

where $\mathcal{L}$ is cross-entropy loss, $E_{n_{y_i}}$ is the actual number of samples in the class to which sample $y_i$ belongs; $\beta$ is a hyperparameter whose value is usually close to 1 (for example, 0.99, 0.999, and so on).

## 5. Experiment and Result Analysis

This section compares the performance of DUEE with some classical and state-of-the-art models based on the real-world dataset constructed in Section 3.

### *5.1. Data Pre-Processing*

This work uses the news headline and content to learn the semantic representation of the text. We leverage jieba to segment news headlines and content and remove stop words and words with less than five occurrences. In addition, since 98.2% of the news headlines are less than 20 words after removing the stop words, the length of the headlines is set to 20, and the length of the news content is set to 500. In the experiment, the method of filling the short sequences and truncating the long sequences is adopted according to the set length of the headlines and the content. The number of training sets, validation sets, and test sets is divided into 60%, 20%, and 20%, and the distribution of each class is shown in Figure 6.

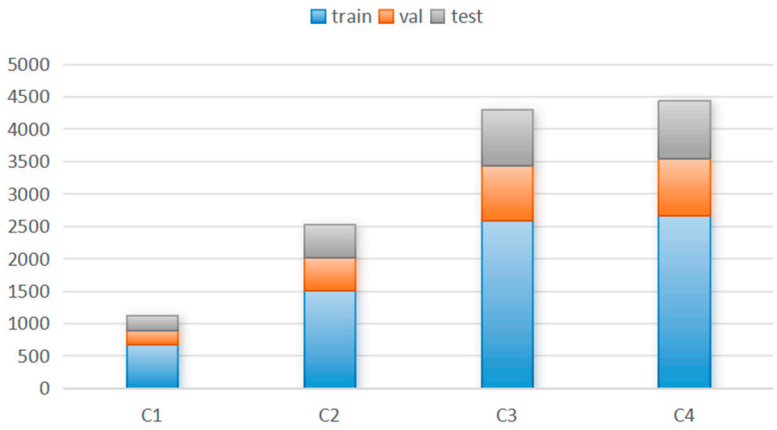



**Figure 6.** Dataset partitioning.

### *5.2. Experiment Setup*

For the hyperparameter settings of the topic model, we set $\alpha = \beta = 0.1$ and $\gamma = \chi = 0.01$, which have little effect on the model as pseudo-counts. Meanwhile, the number of iterations for Gibbs sampling is set to 1000. For the number of topics, we set $K = 30$. The word attractiveness in news headlines and content was pre-calculated according to Equations (8) and (9).

For explicit and implicit representation learning, we adopt a 100-dimensional word embedding pre-trained by Word2Vec [34] in the embedding layer. All of the hidden sizes of FC layers are empirically set as 512. The cell size of GRU is empirically set as 50.

For the training phase, the value of $\beta$ is set as 0.999 in the cross-entropy loss function. At last, we leverage the Adam optimizer for parameter learning. The intensity of dropout applied to FC layers is 0.2. This model is trained for 20 epochs with a batch size of 32. Training is based on Pytorch 1.11.0 framework, and one NVIDIA A10 is used for calculation.

### *5.3. Complexity Analysis*

The DUEE model requires the user to pre-train the topic model to obtain the word attractiveness distribution. In order to analyze the computational efficiency of the word attractiveness distribution of the UCCB proposed in this work, we compare it with the basic LDA model in terms of time and space complexity. The average length of the news and the number of iterations in Gibbs sampling are denoted by $\overline{N}$ and $N_{iter}$, respectively. The time and space complexity of the Gibbs sampling process for both models are shown in Table 3.

**Table 3.** Comparison of time and space complexity.

| Model | Time Complexity | Space Complexity |
|---|---|---|
| LDA | $O(N_{iter}N\overline{N}K)$ | $2NK + 2VK + 2N\overline{N}$ |
| UCCB | $O(2N_{iter}K(N + V + N\overline{N}))$ | $2NK + 6VK + 2N\overline{N}$ |

Time complexity: The classical LDA model generates a topic for each word in the news corpus. Its overall time complexity is $O(N_{iter}N\overline{N}K)$. For UCCB, it assigns a topic

to each word, which is the same as LDA, and the time complexity is $O(\ N_{iter}N\overline{N}K)$. The difference between LDA and UCCB lies mainly in the following steps. The first step is to calculate the attractiveness distribution for each topic $k$ over $V$ words, which has a time complexity of $O(\ N_{iter}KV)$. The second step is to update the distribution of topics for each document, which has a time complexity of $O(\ N_{iter}KN)$. Since the UCCB model requires separate sampling in news headlines and content, the overall time complexity of VCBT is about $O\big(\ 2N_{iter}K\big(N+V+N\overline{N}\big)\big)$.

Space complexity: In the LDA model, the count vector and the topic vector need to be kept in memory, and, in addition, $\theta_d$ and $\varphi_k$ must be computed after the model has converged. Therefore, the overall memory capacity required for the LDA model is $2NK+2VK+2N\overline{N}$. Meanwhile, the UCCB model needs to store additional information. The $c_{kv}$ as a count vector represents the number of comments on the $v$-th word in the $k$-th topic, which requires a memory capacity of VK. The distribution of attractiveness of each word in each topic requires a memory capacity of $VK$. Since the UCCB model needs to compute the attractiveness distribution in news headlines and content separately, the UCCB model requires an additional memory capacity of $4VK$. When the number of topics and words is large, it needs additional memory space. As can be seen from the above analysis, UCCB usually requires more computation time and memory consumption than the basic LDA model.

### *5.4. Comparison Models*

To test the effectiveness of the DUEE, we compared its performance with some classic deep learning classification models, such as Bi-GRU, CNN-Text [38], HAN [22], and RCNN [39]. These models were chosen because they are commonly used baseline models for text classification. In addition, we compare our model with state-of-the-art classification models similar to our task, such as EINQ [26] and DQNH [14], by applying them to our dataset. We use these models for comparison as their model structure can also be applied to our dataset. The final performance of each algorithm was reported based on their performance with testing data. The structure of the comparison model is described below:

Bi-GRU embeds the words into 100-dimensional vectors using Word2vec [34] for news headlines and contents, and then inputs the headlines and contents into two Bi-GRU modules. In the output stage, the last hidden states of the two Bi-GRU modules are spliced and fed into the FC layer, and the final classification result is obtained.

CNN-Text is a deep neural network based on the classical convolutional neural network (CNN) structure for text classification. The news headlines and content are inputted to the CNN network separately, then the representations are concatenated, and the classification results are obtained through the FC layer.

HAN stacks Bi-GRU network structure and attention module to better capture essential words in sentences. In the comparison experiment, the news headlines are input into Bi-GRU, the word sequences of news content are input into HAN, and the output of the two models are connected to output the classification results through the FC layer.

RCNN adds recurrent neural network (RNN) layers to process the CNN intermediate output. The model input was similar to CNN-Text.

DQNH constructs a similarity, semantic, and topic module for news headlines and content. The connection results of the three modules are passed through the FC layer to obtain the classification results.

EINQ is a model for evaluating user engagement that we previously proposed. The model uses CNN-text to obtain news headline semantics and HAN to obtain news content semantics.

### *5.5. Evaluation Metrics*

The models' performances are evaluated using accuracy and macro averaged F1 score, commonly used metrics. Accuracy is the ratio of correctly classified samples to the total number of samples. F1 score is the mean of precision (P) and recall (R) for each class. The

macro-averaged F1 calculates the mean of F1 for each class, providing a comprehensive evaluation of the model's overall performance. These metrics are defined in Equations (20) and (21).

$$Accuracy = \frac{True\ positives + True\ negatives}{True\ positives + False\ positives + True\ negatives + False\ negatives} \quad (20)$$

$$Macro - F1 = \frac{1}{C}\sum_{c=1}^{C}\frac{2P_cR_c}{P_c+R_c} \quad (21)$$

where $P_c$ is the percentage of all news articles identified as positive samples that are actually positive in class c; $R_c$ is the percentage of positive predictions correctly classified in class c.

### 5.6. Experiment Results

Table 4 shows that the proposed DUEE model produces significantly better results than the comparison models regarding macro averaged F1 score, F1 score, and accuracy. Figures 7 and 8 show the graphical representation of the experiment results.

**Table 4.** Comparison of model performance.

| Model | User Engagement Indicators | | | | Accuracy (%) | Macro Averaged F1 Score (%) | Model Size (M) | Inference Time (ms) |
|---|---|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | | | | |
| | F1 Score (%) | | | | | | | |
| Bi-GRU | 60.76 | 58.45 | 62.91 | 64.56 | 67.45 | 61.67 | 3.2 | 0.71 |
| CNN-Text | 58.79 | 59.47 | 61.24 | 62.03 | 63.65 | 60.38 | 4.8 | 0.84 |
| HAN | 65.42 | 66.89 | 73.25 | 71.18 | 71.76 | 69.19 | 5.6 | 1.07 |
| RCNN | 65.78 | 63.23 | 70.89 | 71.83 | 69.43 | 67.93 | 5.1 | 1.12 |
| DQNH | 75.61 | 77.82 | 79.44 | 76.82 | 79.29 | 77.42 | 9.1 | 1.78 |
| EINQ | 78.62 | 80.12 | 81.16 | 82.12 | 82.31 | 80.51 | 7.9 | 1.47 |
| DUEE | 82.42 | 80.41 | 86.17 | 84.32 | 85.46 | 83.33 | 8.5 | 1.67 |

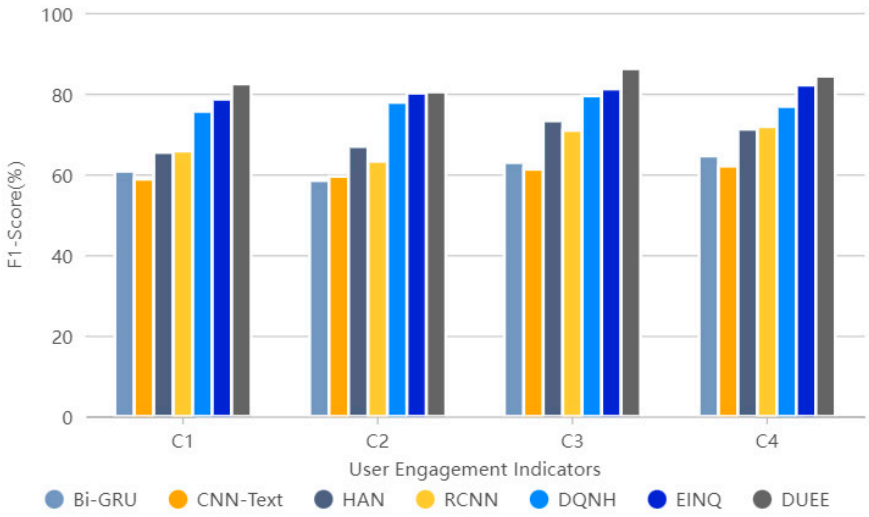



**Figure 7.** The graphical representation of the F1 scores yielded by the comparison models.

As can be seen in Table 4, the CNN structure is the least effective. Although CNN can effectively capture the semantic information of short texts, it is ineffective when facing long texts like news content. Bi-GRU utilizes two independent GRU networks to process text from two directions to capture more comprehensive contextual information and is more effective in processing news content. Therefore, it is more effective than the CNN structure. Compared with CNN, the bidirectional recurrent structure of RCNN can acquire the contextual information of news content more efficiently. It can preserve word order on a large scale when learning text representation. Therefore, it has a better classification performance

than CNN. Since HAN exploits the hierarchical structure feature of news content, the importance of words and sentences can be identified using the attention mechanism from two levels: words and sentences that affect the final classification decision. Experimental results show that HAN works better than RCNN and BI-GRU. DQNH significantly improves the model's classification performance by obtaining complete headline and content semantics through multiple semantic representations compared to other models. The EINQ model comprehensively considers the factors affecting users' browsing and commenting and assigns attention weights to explicit and implicit information through the attention unit, making integrating the two types of information more flexible. The model design is more in line with the characteristics of users' behaviors. DUEE obtains the best results on all evaluation metrics due to embedding word attractiveness in news headlines and content representation processes. Experiments show that adding word attractiveness can enhance news semantic representation, significantly improving prediction performance.

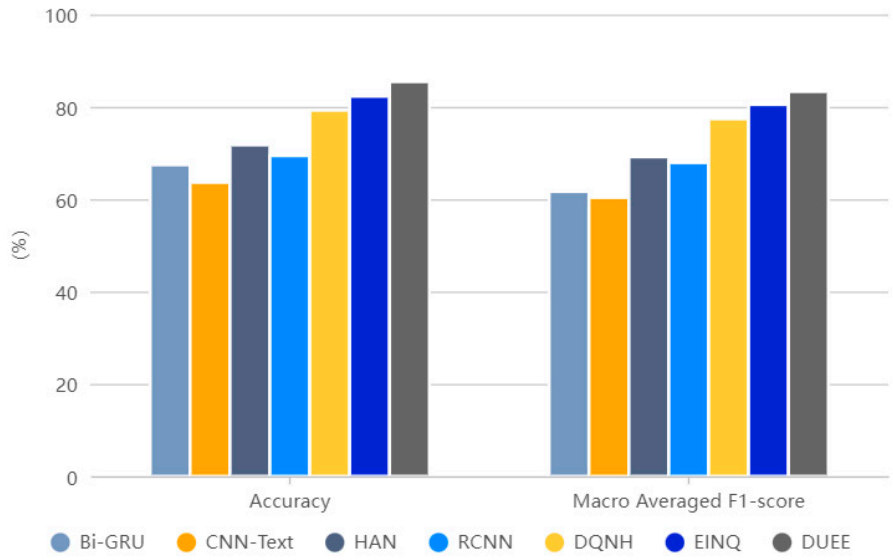



**Figure 8.** The comparison of macro averaged F1 scores and Accuracy.

In addition, we analyze the number of parameters and inference time for each model. For deep learning models, the number of parameters is the storage space occupied by the model parameters. The number of parameters of the DUEE model proposed in this paper is 8.5 M, and the inference time is 1.67 ms for a batch size of 32. Compared with the comparison models, the number of parameters and the inference time of the DUEE model increases to some extent. Compared to the EINQ model, the DUEE model significantly improves performance, although it slightly increases the number of parameters and inference time.

### *5.7. Ablation Experiments*

DUEE is an integrated model that includes text and meta-feature representation learning. Therefore, an ablation study is carried out in this section to remove some of the modules while keeping the rest of the network structure intact to analyze each module's effectiveness and impact on the performance of the DUEE model. As a result, the DUEE and the ablation models with different inputs are shown as follows:

DUEE-H*: In the explicit information representation process, only the news headlines are used, and no meta-features of the news are included to analyze the effect of news source and publication time on the results of user engagement classification.

DUEE-LDA$\hat{}$: In the implicit information representation process, the topic distribution representation is removed to analyze the effect of news topics on the results of user engagement classification.

DUEE-HAT: The attractiveness of words is considered in the news headline representation, and the traditional HAN structure is used in the news content representation.

DUEE-CAT*: The attractiveness of words is considered in the news content representation, and the traditional Bi-GRU structure is used in the news headline representation.

DUEE-AM^: All modules are utilized without assigning attention weights for explicit and implicit information. This design analyses the rationality of assigning attention weights to explicit and implicit information.

DUEE includes all components and uses all available features.

Finally, we compared the performance of DUEE with that of the other five ablation models described above. The comparison results are shown in Table 5.

**Table 5.** Comparison of ablation models performance.

| Ablation Models | User Engagement Indicators | | | | Accuracy (%) | Macro Averaged F1-Score (%) |
|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | | |
| | F1-Score (%) | | | | | |
| DUEE-H* | 77.43 | 79.14 | 81.18 | 79.16 | 81.32 | 79.23 |
| DUEE-LDA^ | 79.49 | 80.24 | 83.54 | 82.12 | 82.11 | 81.35 |
| DUEE-HAT* | 81.28 | 79.95 | 83.51 | 83.43 | 83.24 | 82.04 |
| DUEE-CAT* | 79.94 | 80.12 | 83.14 | 83.85 | 82.17 | 81.76 |
| DUEE-AM^ | 81.68 | 79.32 | 82.47 | 83.23 | 81.47 | 81.68 |
| DUEE | 82.42 | 80.41 | 86.17 | 84.32 | 85.46 | 83.33 |

* is for keeping the module, and ^ is for removing it.

As seen from Table 5, the model's performance using only news headlines (DUEE-H*) in the explicit information representation stage is poor. This suggests that the use of news headlines alone is not a comprehensive representation of the features that attract users' clicking behavior and that the source of the news and the time of publication are equally important factors in triggering users' clicking behavior. In the process of implicit information representation, there is an impact on the model performance if the news topics are not taken into account in the model (DUEE-LDA^). This suggests that news topics are also crucial in attracting user comments. Adding word attractiveness in news headlines has better results from the DUEE-HAT* and DUEE-CAT* model performance. The performance of DUEE-AM^ is not high because it does not consider the attention weights of explicit and implicit information in the fusion stage, indicating that attention mechanisms have an indispensable role in the fusion of explicit and implicit information. The results of the ablation experiments show that each module in DUEE affects the model performance. In particular, the news source, topic relevance, and attention mechanism significantly impact model performance. The differences in the performance of each ablation model are shown in Figures 9 and 10.

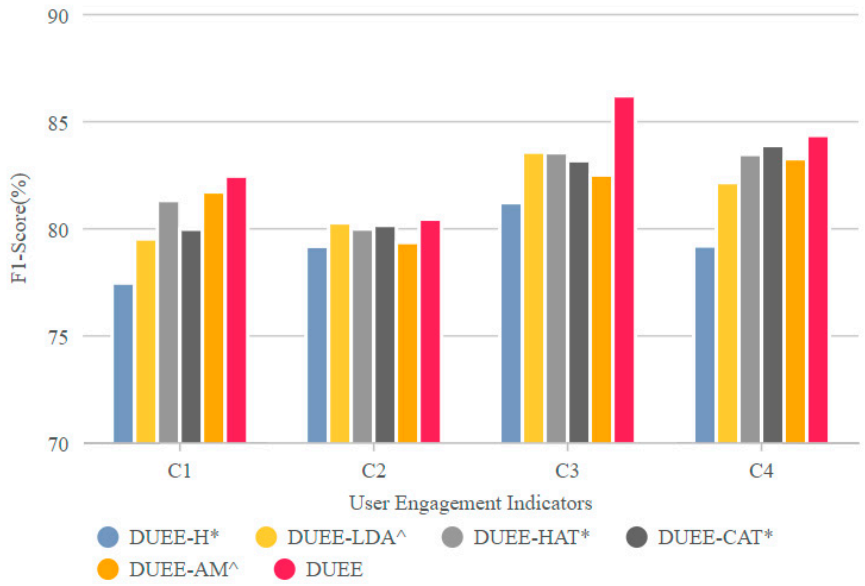



**Figure 9.** The comparison of F1 scores for ablation models.

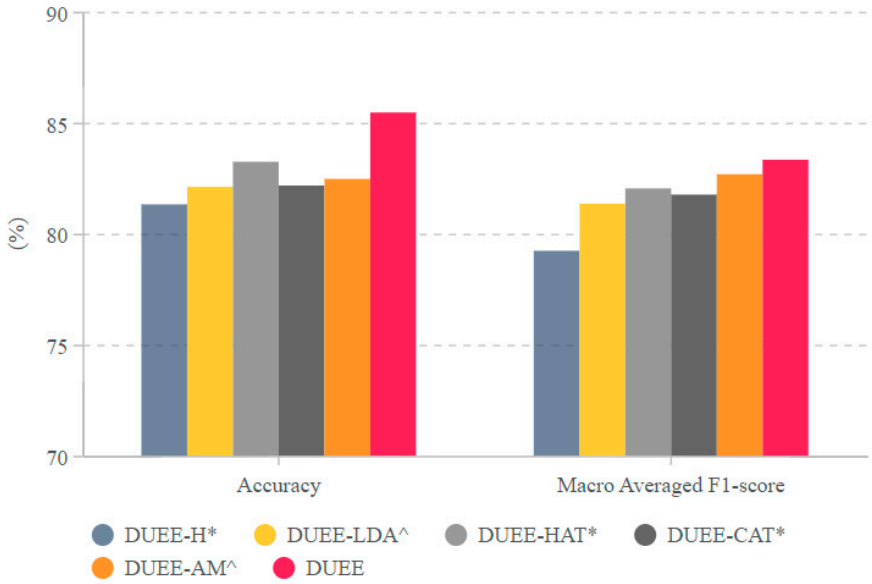



**Figure 10.** The comparison of accuracy and macro averaged F1 score for ablation models.

### *5.8. Effect of Attractiveness on the Proposed Model*

In order to verify the effect of adding attractiveness in the model, the experiment randomly selected a news article and generated a heat map on the news headline (Figure 11), which reflects the effect of attractiveness on the attention weight through the heat map. The greater the attention weight of the word in the news headline, the darker the color.

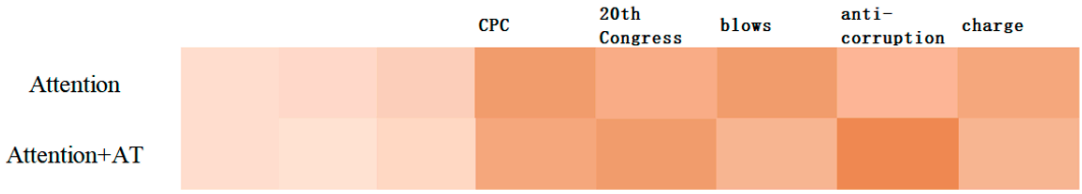



**Figure 11.** The comparison of attention weights for terms.

In Figure 11, "Attention" represents the use of traditional attention mechanisms during the representation of news headlines in Section 4.2, which represents the replacement of Equation (10) with Equation (22).

$$s(h_j) = u^T tanh(w_h h_j + b_h) \tag{22}$$

As can be seen in Figure 11, the traditional attention mechanism focuses on assigning more weights to nouns and verbs, such as "CPC", " blow", and "charge" in this example. In "Attention + AT", due to the embedded word attraction, attention weights can be increased for eye-catching words such as "anti-corruption", which means that this word can attract more users to click.

### *5.9. Effect of Dropout on the Proposed Model*

This section presents an analysis of the impact of dropout on the model's performance. The dropout is set as 0.2. Table 6 shows the results of the DUEE with and without using dropout. The results demonstrate that including a dropout layer markedly enhances all evaluation metrics. This proves the effectiveness of employing a dropout layer to reduce over-fitting.

**Table 6.** Comparison of ablation models performance.

| Dropout Layer | User Engagement Indicators | | | | Accuracy (%) | Macro Averaged F1 Score (%) |
|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | | |
| | F1 Score (%) | | | | | |
| No | 78.23 | 78.68 | 83.18 | 82.54 | 83.53 | 80.66 |
| Yes | 82.42 | 80.41 | 86.17 | 84.32 | 85.46 | 83.33 |

## 6. Conclusions

Evaluating user engagement can significantly improve news publication efficiency and identify news that attracts users to click and discuss. This study uses online news platforms such as Toutiao and Tencent News as the research object. It proposes a method to evaluate user engagement, which will help news platforms attract users while promoting users' discussion at public events. This study defines user engagement indicator as the ability to attract users' clicks and trigger users' discussions. This work proposes a deep learning model incorporating news attractiveness and multiple features for evaluating user engagement. An empirical evaluation involving a real news dataset shows that the proposed news attractiveness and multi-features significantly impact the classification results. This work makes several new contributions. First, we propose an extended LDA model based on users' click–comment behavior, using which the attractiveness of words in news headlines and content can be effectively represented.

Second, we propose a novel deep learning model named DUEE that exploits text and meta-features, significantly improving the classification result. In particular, DUEE combines news attractiveness as an auxiliary feature with textual semantic features to obtain a more comprehensive semantic representation. The DUEE model considers news sources and relevance, which collectively determine the ability of the news to attract clicks and engagement.

Third, the proposed model DUEE has been compared with some baseline models, i.e., Bi-GRU, HAN, CNN-TEXT, and RCNN. Results show that the DUEE outperforms the baseline models. The DUEE has also been compared with the state-of-the-art DQNH and EINQ models. It is shown that the proposed model performs better than the state-of-the-art models. In order to analyze the effect of multiple features on the classification performance, we performed ablation experiments. Results show that meta-features such as news sources and publication time significantly impact the model performance. The attention mechanism based on explicit and implicit information also provides a great help to the model performance. In addition, the experiments demonstrate that applying a dropout layer enhances the efficacy of the proposed model, as it effectively prevents over-fitting.

This study is subject to certain limitations, which present potential avenues for future research. First, the dataset was collected from the Chinese online news platform. In some other news platforms, such as social networks, which have a more comprehensive range of information sources and diverse user interaction behaviors, it is more challenging to evaluate user engagement. In future research, validating the proposed method using news articles in different languages and platforms would be helpful. Second, this work uses news text and meta-features for user engagement assessment. However, other news elements, such as timeliness and objectivity, are not considered. In future research, we will explore the role of additional factors in evaluating user engagement so that media organizations and readers can better assess and understand the content they consume and produce. Third, this work uses the number of clicks and comments as an indicator for evaluating user engagement. However, the number of clicks and comments may not be sufficient. The content of users' comments can further reflect users' views on news and related topics. Therefore, future research will further consider user comments to enrich the model's application scenarios and prediction capabilities.

**Author Contributions:** Conceptualization, G.S. and F.L.; methodology, G.S. and Y.W.; software, H.H.; validation, G.S., F.L., and X.C.; formal analysis, G.S. and H.H.; investigation, F.L.; resources, G.S. and X.C.; data curation, G.S.; writing—original draft preparation, G.S. and X.C.; writing—review and editing, G.S.; visualization, H.H.; supervision, Y.W.; project administration, Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Fundamental Research Funds for the Central Universities (Grant NO. CUC230B008, Grant NO. CUC24SG018) and Beijing Social Science Foundation of China under grant (Grant NO. 23JCB002).

**Data Availability Statement:** Some or all data that support the findings of this study are available from the first author upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ren, J.; Dong, H.; Popovic, A.; Sabnis, G.; Nickerson, J. Digital platforms in the news industry: How social media platforms impact traditional media news viewership. *Eur. J. Inform. Syst.* **2024**, *33*, 1–18. [CrossRef]
2. Pewreaearch. Available online: https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/ (accessed on 10 July 2024).
3. Statista. Available online: https://www.statista.com/statistics/910787/china-monthly-active-users-on-leading-news-apps/ (accessed on 10 July 2024).
4. Fletcher, R.; Nielsen, R.K. Are people incidentally exposed to news on social media? A comparative analysis. *New Media Soc.* **2018**, *20*, 2450–2468. [CrossRef]
5. Evans, R.; Jackson, D.; Murphy, J. Google News and machine gatekeepers: Algorithmic personalization and news diversity in online news search. *Digit. J.* **2023**, *11*, 1682–1700. [CrossRef]
6. Kuai, J.; Lin, B.; Karlsson, M.; Lewis, S.C. From wild east to forbidden city: Mapping algorithmic news distribution in China through a case study of Jinri Toutiao. *Digit. J.* **2023**, *11*, 1521–1541.
7. Qiu, Z.; Hu, Y.; Wu, X. Graph neural news recommendation with user existing and potential interest modeling. *ACM Trans. Knowl. Discov. Data* **2022**, *16*, 1–17. [CrossRef]
8. Voronov, A.; Shen, Y.; Mondal, P.K. Forecasting popularity of news article by title analyzing with BN-LSTM network. In Proceedings of the 2019 International Conference on Data Mining and Machine Learning, New York, NY, USA, 13–18 July 2019; pp. 19–27.
9. Cervi, L.; Tejedor, S.; Blesa, F.G. TikTok and political communication: The latest frontier of politainment? A case study. *Media Commun.* **2023**, *11*, 203–217. [CrossRef]
10. Yang, Y.; Liu, Y.; Lu, X.; Xu, J.; Wang, F. A named entity topic model for news popularity prediction. *Knowl.-Based Syst.* **2020**, *208*, 106430. [CrossRef]
11. Xiong, J.; Yu, L.; Zhang, D.S.; Leng, Y.F. DNCP: An attention-based deep learning approach enhanced with attractiveness and timeliness of news for online news click prediction. *Inform. Manag.-Amster.* **2021**, *58*, 103428. [CrossRef]
12. Liao, D.; Xu, J.; Li, G.; Huang, W.; Liu, W.; Li, J. Popularity prediction on online articles with deep fusion of temporal process and content features. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019.
13. Arora, A.; Hassija, V.; Bansal, S.; Yadav, S.; Chaloma, V.; Hussain, A. A novel multimodal online news popularity prediction model based on ensemble learning. *Expert. Syst.* **2023**, *40*, e13336. [CrossRef]
14. Omidvar, A.; Pourmodheji, H.; An, A.; Edall, G. A novel approach to determining the quality of news headlines. In Proceedings of the NLPinAI 2020, Valletta, Malta, 22–24 February 2020.
15. Yang, Y.; Cao, J.; Lu, M.; Li, J.; Lin, C. How to Write High-quality News on Social Network? Predicting News Quality by Mining Writing Style. *arXiv* **2019**, arXiv:1902.00750.
16. Wang, H.C.; Maslim, M.; Liu, H.Y. CA-CD: Context-aware clickbait detection using new Chinese clickbait dataset with transfer learning method. *Data. Technol. Appl.* **2023**, *58*, 243–266. [CrossRef]
17. Liu, T.; Yu, K.; Wang, L.; Zhang, X.; Zhou, H.; Wu, X. Clickbait detection on WeChat: A deep model integrating semantic and syntactic information. *Knowl-Based. Syst.* **2022**, *245*, 108605. [CrossRef]
18. Pujahari, A.; Sisodia, D.S. Clickbait detection using multiple categorization techniques. *J. Inf. Sci.* **2021**, *47*, 118–128. [CrossRef]
19. Kaushal, V.; Vemuri, K. Clickbait—Trust and credibility of digital news. *IEEE Trans. Technol. Soc.* **2021**, *2*, 146–154. [CrossRef]
20. Molyneux, L.; Coddington, M. Aggregation, clickbait and their effect on perceptions of journalistic credibility and quality. *J. Pract.* **2020**, *14*, 429–446. [CrossRef]
21. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
22. Yang, Z.C.; Yang, D.Y.; Dyer, C.; He, X.D.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016.

23. Zrnec, A.; Poženel, M.; Lavbič, D. Users' ability to perceive misinformation: An information quality assessment approach. *Inform. Process. Manag.* **2022**, *59*, 102739. [CrossRef]
24. Lin, H.; Lasser, J.; Lewandowsky, S.; Cole, R.; Gully, A.; Rand, D.G.; Pennycoo, G. High level of correspondence across different news domain quality rating sets. *PNAS Nexus* **2023**, *2*, pgad286. [CrossRef]
25. Mosallanezhad, A.; Karami, M.; Shu, K.; Mancenido, M.V.; Liu, H. Domain Adaptive Fake News Detection via Reinforcement Learning. In Proceedings of the ACM Web Conference 2022, Lyon, France, 25–29 April 2022.
26. Song, G.H.; Wang, Y.B.; Li, J.F.; Hu, H.B. Deep learning model for news quality evaluation based on explicit and implicit information. *Intell. Autom. Soft Comput.* **2023**, *38*, 275–295. [CrossRef]
27. Alam, S.M.; Asevska, E.; Roche, M.; Teisseire, M. A Data-Driven Score Model to Assess Online News Articles in Event-Based Surveillance System. In Proceedings of the Annual International Conference on Information Management and Big Data, Virtual Event, 1–3 December 2021.
28. Wu, Q.Y.; Li, L.; Zhou, H.; Zeng, Y.; Yu, Z. Importance-aware learning for neural headline editing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020.
29. Romanou, A.; Smeros, P.; Castillo, C.; Aberer, K. Scilens news platform: A system for real-time evaluation of news articles. *Proc. Vldb. Endow.* **2020**, *13*, 2969–2972. [CrossRef]
30. Kim, J.H.; Mantrach, A.; Jaimes, A.; Oh, A. How to compete online for news audience: Modeling words that attract clicks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
31. DuBay, W.H. The principles of readability. *Online Inf.* **2004**, *1*, 12–13.
32. Shulman, H.C.; Markowitz, D.M.; Rogers, T. Reading dies in complexity: Online news consumers prefer simple writing. *Sci. Adv.* **2024**, *10*, eadn2555. [CrossRef] [PubMed]
33. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. [CrossRef] [PubMed]
34. Li, S.; Zhao, Z.; Hu, R.F.; Li, W.S.; Liu, T.; Du, X.Y. Analogical Reasoning on Chinese Morphological and Semantic Relations. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 2: Short Papers), Melbourne, Australia, 15–20 July 2018.
35. Metzger, M.J.; Flanagin, A.J. Credibility and trust of information in online environments: The use of cognitive heuristics. *J. PRAGMATICS* **2005**, *59*, 210–220. [CrossRef]
36. Harcup, T.; O'Neill, D. What is news? Galtung and Ruge revisited. *J. Stud.* **2001**, *2*, 261–280. [CrossRef]
37. Cui, Y.; Jia, M.L.; Lin, T.Y.; Song, Y.; Belongie, S. Class-Balanced Loss Based on Effective Number of Samples. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
38. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014.
39. Lai, S.; Xu, L.H.; Liu, K.; Zhao, J. Recurrent convolutional neural networks for text classification. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–29 January 2015.


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.