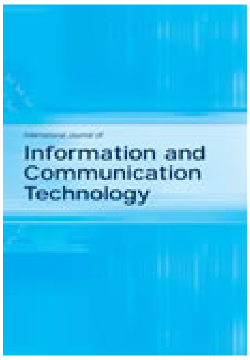


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

News selection and recommendation algorithm based on machine learning and image matching algorithms

Geng Chen

DOI: 10.1504/IJICT.2025.10074942

Article History:

Received:	19 September 2025
Last revised:	23 October 2025
Accepted:	24 October 2025
Published online:	17 December 2025

News selection and recommendation algorithm based on machine learning and image matching algorithms

Geng Chen

Department of Information Technology,
Concord University College,
Fujian Normal University,
Fuzhou, 350117, Fujian, China
Email: chengeng981@163.com

Abstract: The explosion of internet news data has led to frequent information overload, which in turn has caused the dual problems of low recommendation quality and poor user experience. The research combines machine learning with image matching algorithms to provide users with interesting news content. Firstly, the news recommendation algorithm based on multi-task learning (MTL) constructs the input sequence through the user's click history, maps the inputs of different tasks to the shared space and extracts text features. The image matching algorithm is adopted to record users' historical preferences and capture the changing trends of interests. Experimental data show that the accuracy rate of the content-based recommendation (CB) algorithm reaches 78.27%. The collaborative filtering (CF) recommendation algorithm reached 87.97%. The recommendation accuracy of the joint recommendation algorithm all exceeds 90%. Moreover, the accuracy of the joint recommendation algorithm significantly exceeds that of the two traditional recommendation methods, CB and CF.

Keywords: machine learning; image matching algorithm; collaborative filtering; news recommendations; content-based recommendation.

Reference to this paper should be made as follows: Chen, G. (2025) 'News selection and recommendation algorithm based on machine learning and image matching algorithms', *Int. J. Information and Communication Technology*, Vol. 26, No. 46, pp.141–159.

Biographical notes: Geng Chen received his MA Engineering in Software Engineering from Sichuan University in 2014. Currently, he is a Lecturer in the Concord University College Fujian Normal University. His research interests include digital media technology, image processing and computer vision.

1 Introduction

News information is an essential way for people to access social dynamics and information. How to use recommendation algorithms to accurately extract the required information from massive news data and provide support for public opinion guidance has become a current hot topic. Existing recommendation algorithms are often affected by data sparsity and lack in-depth research on user behaviour. The current multi-task learning (MTL) algorithms based on machine learning have become one of the most widely utilised recommendation algorithms due to their excellent performance, and the introduction of image matching algorithms further improves the accuracy of recommendations.

An increasing number of people read news online. However, the sheer volume of news and information posted online through different channels every day makes it hard for readers to find what they are interested in. Karimi et al. (2018) put forward a news recommendation system with the aim of providing reading suggestions with personalised approach, and specifically discussed the current academic practice of assessing and making comparisons between different algorithmic news recommendation methods on the basis of accuracy metrics. Social networks and mainstream news brands are influencing the continuous development of content personalisation. Thurman et al. explored the audience's views on news selection mechanisms and their reasons. He analysed survey data from 26 countries and explored the relationship between personal characteristics and this belief using a multi-level model (Thurman et al., 2019b). Malicious users are posting more and more erroneous information and false content, not only causing chaos in the online social media ecosystem, but also bringing indescribable pain to humanity. Therefore, there is an urgent need to discover and remove this false content from social media. Collins et al. explored various methods to combat fake news on social media. The results indicated that the detection for fake news is a problem that is challenging and complicated (Collins et al., 2021). Haim et al. performed two exploration researches to test the impact of implicit and explicit personalisation on the diversity of content and sources of Google News. Evidence suggested that concerns about algorithms in the online news context might be overstated (Haim et al., 2018). Carlson (2018) argued that recommendation algorithmic judgment ought to be distinguished from the professional judgment of journalists. It has become increasingly hard to find the information that people are interested in from the huge amount of news data, and there is also the phenomenon of low utilisation of information, resulting in the waste of information resources.

At present, recommendation algorithms are one of the important methods to address the issue of information overload. Due to the successful application and practice of machine learning in various fields, the study of combining recommendation algorithms with machine learning has also received widespread attention from scholars. Automatic recommendation has become an increasingly important issue. Shu et al. proposed a CB algorithm using a convolutional neural network and introduced a split-iteration method to solve the model. The experimental results on a public database indicated that this method has made significant improvements in quantitative evaluation compared to traditional methods, greatly improving training efficiency (Shu et al., 2018). Li et al. aimed to combine sequential recommendation with fair perception recommendation and proposed a deep end-to-end model based on MTL. This method outperformed the most single model in terms of recommendation performance (Li et al., 2022). Collaborative filtering (CF) recommendation is promising technique, but the technique suffers from cold start and sparse data. To overcome the shortcomings of CF recommendations on preference-ordered datasets, Najafabadi et al. utilised a new graph-based structure to model users' priorities and capture the associations between users and items. The experimental findings showed that compared with CF and other recommendation methods, the graph structure based method has significant improvements in recall and accuracy indicators (Najafabadi et al., 2019). Massive news data contains a lot of public opinion information. By combining machine learning recommendation algorithms, the connection between users and news data has been analysed, and users' behavioural tendencies have been judged to help users obtain relevant news information and provide effective information support for news recommendations.

The current news recommendation algorithms, such as those based on coordinated filtering, make news matching and similar content in various fields less ideal compared to the news content results that users need, and require precision and completeness in recommending the resources that users truly need (Zhang et al., 2018; Shen et al., 2020). Therefore, currently, how to provide users with rationalised and personalised resources is the main research goal of recommendation algorithms. In the article, research and analysis were conducted on these challenges, and recommendation methods based on machine learning and image matching algorithms were analysed, which effectively improved the recommendation effect. Image matching algorithms can model using the information represented by local feature point attributes, and can also measure the weights on the edges connecting feature points. For news content browsed by users, images or videos were extracted and their feature representations were matched to the similarity of news images or videos in the database. Based on the results of image matching, the correlation between each news item and the user's browsing content was analysed. In addition, local and global features can also be distinguished, and the structural features of the image can be clearly expressed, which is also an effective method for calculating similarity between images and news.

2 News recommendation algorithm

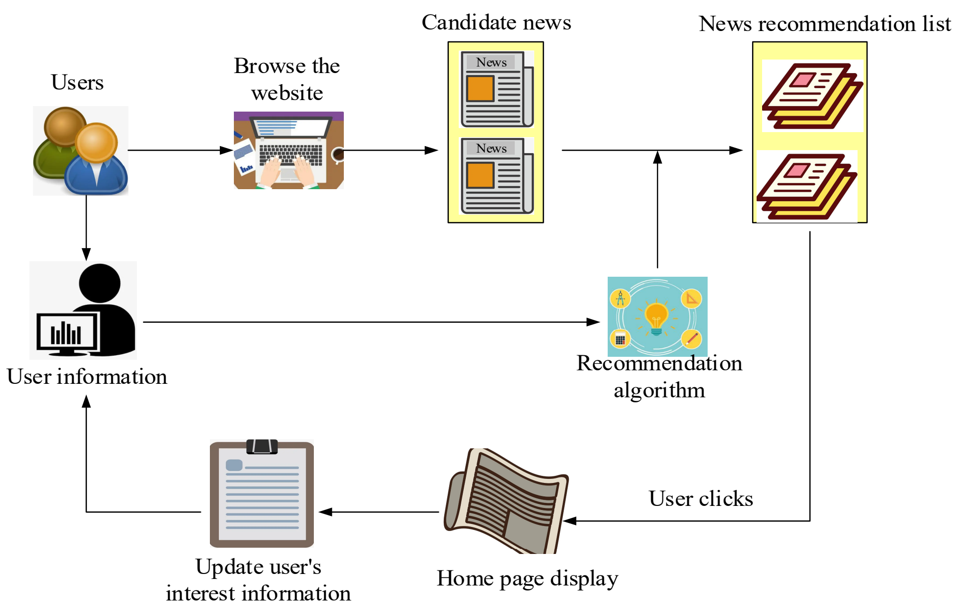
News recommendation is a means of analysing and matching information, based on data from user project interaction behaviour, to recommend users and improve webpage views and conversion rates (Zhang et al., 2019b; Wallace, 2018). News recommendations have been widely used on various news and information platforms such as NetEase News and Today's Headlines. However, due to the presence of a large amount of textual information in news, research on news recommendations has become inseparable from natural language processing techniques (Thurman et al., 019a; Javed et al., 2021). In the real world, it is often not possible to collect complete user behaviour data, so it is necessary to design specialised recommendation algorithms for the news field. The news recommendation schematic is depicted in Figure 1.

As illustrated in Figure 1: users browse the website to obtain candidate news; candidate news forms a news recommendation list; users click on the recommendation list to form a homepage display; the user's interest information is updated from it to form user information, and the recommendation algorithm then forms a news recommendation list.

When users access the news platform, the news recommendation algorithm recalls a portion of candidate news from a large-scale news pool. Then, the user's interests are modelled; the similarity between the user's interests and candidate news is calculated; the recommendation list is reordered. Finally, the top ranked news is recommended to the user (Van den Bulck and Moe, 2018; Helberger et al., 2018). The platform records users' interactive behaviour on this news, and updates and maintains user profiles to facilitate the provision of future services (Zhang and Yang, 2021).

In the field of recommendation, people often use CB and CF recommendation algorithms to filter related items, articles, and information. They only use the interaction between users and products to determine user preferences, product types, and other information for recommendation (Zhang et al., 2019a; Benkessirat et al., 2021). It has strong interpretability, simple model calculation, and low complexity. When a project is multimedia data and its content features are difficult to extract, using content based recommendation algorithms to recommend news is often not very effective (Afsar et al., 2022; Gelfert, 2018). Moreover, excessive reliance on the content characteristics of the project during the recommendation process may result in not recommending new projects to users, thus losing the diversity of recommendations.

Figure 1 Schematic diagram of news recommendation (see online version for colours)



2.1 Dataset introduction

The data in this article is sourced from Caixin.com. A total of 25,477 pieces of news data (including both browsed and unbrowsed news) were collected to conduct a preliminary analysis of the data content of Caixin.com. The data specifically includes news headlines, main text content, release time, comment texts, user ids, and news click-through rates. The data time span is from January 2024 to June 2025, with a total of 25,477 valid news entries and 840,458 user comment records collected. Invalid symbols and stop words are removed through the preprocessing step, and the keyword vectors are extracted by using the TF-IDF algorithm to construct the user interest profile. The research simulates users' browsing behaviours by designing scripts, regularly captures news items on the homepage and channel pages of the website, and ensures coverage of different topics and types of news. To ensure the legality of data sources, the research strictly adheres to the website's robots protocol and only collects publicly available information. All data is used for academic research purposes. During the data capture process, the script automatically records basic information such as the title, main text content, and release time of the news, and synchronously acquires the user's comment data on the news through the embedded comment text collection module. At the same time, to accurately track user behaviour, capture the user ID and the corresponding news click-through rate. The distribution of news types is shown in Table 1.

Table 1 Distribution of news types

News type	*Number*	*Percentage*
Politics	3263	12.81%
Economy	4211	16.53%
Society	5612	22.03%
Science and technology	3143	12.34%
Entertainment	4010	15.74%
Medical treatment	2610	10.24%
Education	2628	10.31%
Total	25477	100%

As shown in Table 1, the dataset consists of 25477 news data, with a proportion of 12.81% for political news, 16.53% for economic news, and 22.03% for social news. The proportion of social news is the highest. The browsed and not browsed situations are shown in Table 2.

Table 2 Browsed and not browsed situations

News type	*Number*	*Number of news browsed*	*Percentage of news browsed*	*Number of news not browsed*	*Percentage of news not browsed*
Politics	3263	603	18.48%	2660	81.52%
Economy	4211	797	18.93%	3414	81.07%
Society	5612	1693	30.17%	3919	69.83%
Science and technology	3143	640	20.36%	2503	79.64%
Entertainment	4010	896	22.34%	3114	77.66%
Medical treatment	2610	674	25.82%	1936	74.18%
Education	2628	551	20.97%	2077	79.03%

As shown in Table 2, the highest percentage of views is social news, which is 30.17%, and the news with the most views is only 1693. The percentage of views is much lower than the percentage of not views, and the percentage of not views is above 60%, indicating that most of the news has not been viewed.

By analysing the above data statistically, the data can be found to be somewhat sparse. Sparsity is a mathematical index used to measure the sparsity of the data, which can directly reflect the sparsity of the data. The sparsity is calculated for the user behaviour data. In order to simplify the calculation, the simplified formula for sparsity S_F in the text is shown in (1):

$$S_F = 1 - \frac{Score}{N_e} \tag{1}$$

In equation (1), *Score* represents the number of ratings (that is, the quantity of views), and N_e represents the total quantity of news, S_F represents sparsity. Based on the above data, sparsity can be calculated in (2):

$$S_F = 1 - \frac{603 + 797 + 1693 + 640 + 896 + 674 + 551}{25477} = 77.02\% \tag{2}$$

According to the analysis above, it is found that the news reading data has a high sparsity, where users view only a few news items and plenty of news are not utilised. For users, news that has not been browsed does not mean that they are not interested in the news. If recommendation algorithms are used, users can more effectively obtain more useful information.

2.2 Data filtering

This article is limited by the original database design during collection and can only uncover users' browsing behaviour towards news. News with a browsed percentage of over 20% in the dataset is selected. The filtered data is shown in Table 3.

Table 3 Filtered data

News type	*Number*	*Number of news browsed*	*Percentage of news browsed*	*Number of news not browsed*	*Percentage of news not browsed*
Society	5612	1693	30.17%	3919	69.83%
Science and technology	3143	640	20.36%	2503	79.64%
Entertainment	4010	896	22.34%	3114	77.66%
Medical treatment	2610	674	25.82%	1936	74.18%
Education	2628	551	20.97%	2077	79.03%

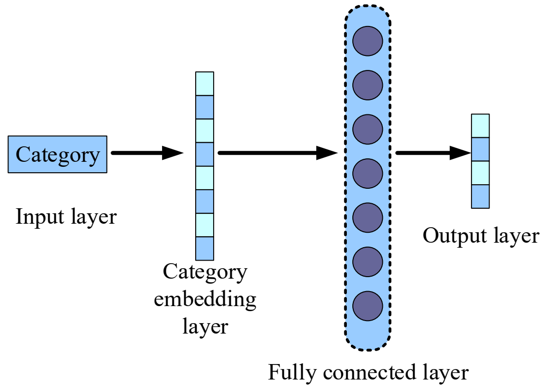
As shown in Table 3, finally, news with a browsed percentage of 20% or less are removed from the dataset, and the filtered dataset contains a total of 18,003 news articles (number of non-browsed and browsed).

2.3 MTL based news recommendation algorithm

Existing news recommendation algorithms, such as CF, usually focus only on the user's click probability of the candidate news, and the prediction method of the click rate is only good at utilising the surface features, which results in problems such as 'headline party' and affects the recommendation effect and user experience (Mohammadpour et al., 2019; Makhortykh and Bastian, 2022). To address the above issues, a more efficient recommendation algorithm for news push platforms is needed.

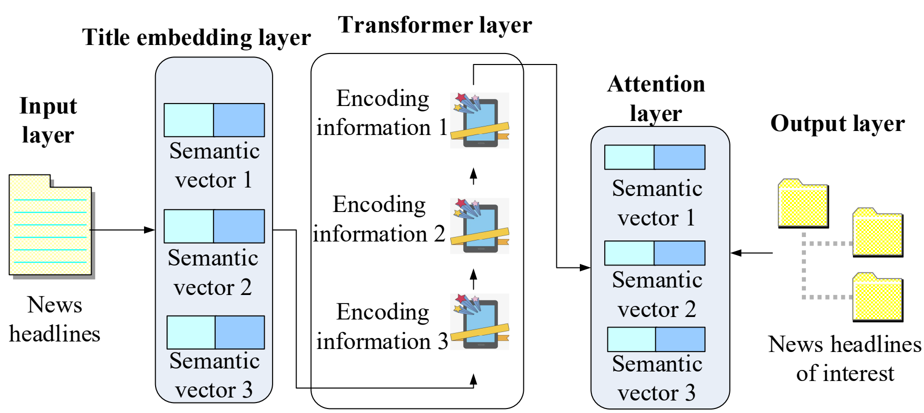
The news encoder in MTL based news recommendation consists of a category encoding layer and other layers, and the final news features can be obtained through the news encoder (Egelhofer and Lecheler, 2019; Chen et al., 2019). The category coding layer is depicted in Figure 2.

Figure 2 Schematic diagram of category encoding layer (see online version for colours)



As shown in Figure 2, the first part of the news encoder is the category encoding layer, which includes the input layer, category embedding layer, fully connected layer, and output layer.

Figure 3 Title encoder schematic (see online version for colours)



The category encoding layer is utilised to learn the news representation in categories. Currently, many online news platforms' recommendation methods mainly recommend users based on the category of the news. The numerical values corresponding to categories are represented as binary vectors, and the calculation formula for low dimensional dense representations of different categories is as (3):

$$e_c = f_{One} - Hot(x) \tag{3}$$

In equation (3), c represents the category of the previous layer, e_c is its low-dimensional representation, and r_c is the final output category, the calculation formula for r_c is (4):

$$r_c = \operatorname{Re}LU\left(W_c \times e_c + b_c\right) \tag{4}$$

In equation (4), W_c and b_c are weight parameters and bias parameters. The recommendation algorithm based on MTL can infer a user's interest in a certain type of information from their browsing records. If a user often reads news related to soccer matches, an inference can be made that this user is more interested in sports, soccer, and other types of news. Based on the MTL recommendation algorithm, news classified as sports or football can be recommended to the user (Mizgajski and Morzy, 2019).

The news encoder also includes a title encoder for learning news representations from news titles. Nowadays, for any news pushed on news portal websites, the first thing people pay attention to is the title information. The title encoder is depicted in Figure 3.

As Figure 3, the input layer inputs news titles. In addition to the input and output layers, the title encoder can be divided into a three-layer structure: title embedding layer, Transformer layer, Attention layer. The title embedding layer contains multiple semantic vectors. The Transformer layer represents the encoding information of each word in the positional embedding; the Attention layer contains the extracted semantic vector, and finally outputs the news title of interest.

The word sequence of a news title is represented as $[W_1^t, W_2^t, ..., W_M^t]$, and the length of the title is *M*. The pre trained model of *Glove* is utilised to convert the word sequence of the title into a word vector sequence with a dimension equal to *Q*. The calculation process is as (5):

$$Q = Glove(W) \tag{5}$$

In equation (5), the s_{in} function is used to calculate the positional encoding information of each word in a sentence. P_E is used to represent positional information, and positional embedding $P_E(p_{os}, 2i)$ is (6):

$$P_E\left(p_{os}, 2i\right) = s_{in} \frac{p_{os}}{1000^{\frac{2i}{d_{model}}}} \tag{6}$$

In equation (6), the position of a word in a sentence is p_{os}, and the dimension of the word vector is ordinal number *i*. Among them, the dimension embedded in the position is set to d_{model}. The position embedding obtained and the word embedding vector obtained using *Glove*(*W*) are added to obtain the true embedding vector X_{em} of the sentence, as shown in (7):

$$X_{em} = Glove(W) + \text{Positional-Encoding}(W) \tag{7}$$

The research defines the recommendation task as multiple related tasks, aiming to recommend appropriate news to users based on their historical behaviours and interests. And classify users' interests. In addition, it also includes news category prediction tasks and user behaviour pattern recognition tasks. The news category prediction task is used to accurately determine the specific category to which news belongs. And user behaviour pattern recognition tasks, which gain a deep understanding of user habits by analysing user browsing, clicking and other behaviour patterns. The research adopts a multi-task joint loss function, comprehensively considering the loss situations of each task, and fuses the losses of different tasks into an overall loss through weighted summation. For the news recommendation task, its loss is calculated based on the difference between the recommendation result and the actual click situation of users, while the loss of the user interest classification task is measured by the classification accuracy. The weights of news recommendation, user interest classification, news category prediction, and user behaviour pattern recognition were set at 0.4, 0.2, 0.2, and 0.2 respectively in the study to balance the impact of each task on model training. Some underlying feature representations are shared among different tasks, and features such as the user's basic information and browsing history can be reused in multiple tasks. Reduce the number of model parameters and improve computational efficiency through a sharing mechanism.

3 News recommendation based on image matching algorithm

Compared with traditional methods in computer vision, image matching algorithms have certain differences, with their unique characteristics and performance advantages (Liu and Bu, 2019; Jin et al., 2021). In the process of image processing, feature description is usually used to solve the problem, and the index structure is combined to calculate point matching, or a global feature model is used for image recognition and classification (Faroek et al., 2022; Zhao et al., 2018).

The detection of extreme points in scale space is carried out throughout the entire scale space and position, as accurate and stable feature points do not change with changes in scale, as shown in (8).

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \tag{8}$$

In equation (8), ⊗ indicates convolution operations in both directions x and y, as shown in (9).

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \tag{9}$$

In equation (9), σ is the scale space factor, which determines the degree of Gaussian blur. On a small scale, the details of the image are displayed, while on a large scale, the overview of the image is displayed. Therefore, small-scale represents high resolution, while large-scale represents low resolution.

In news recommendation, multiple different news genres can appear for users' historical clicks on news, and fake news can be filtered. The similarity between not browsed candidate news and these different news topics varies (Sharma et al., 2019; Molina et al., 2021). On this basis, each candidate news is assigned a different weighting matrix.

To prepare for the final news recommendation, the preference value of the news also needs to be calculated and the calculated score is utilised to determine the final recommended news (Waszak et al., 2018; Wang et al., 2021). After obtaining the feature vector $e(i)$ for the current user, it is necessary to perform a click sum operation with the candidate news word vector $e(t_j)$, as shown in (10).

$$p_{i,t_j} = G\left(e(i), e\left(t_j\right)\right) \tag{10}$$

In equation (10), p_{i,t_j} indicates the similarity score between the user and the candidate news. Due to the score being processed by the activation function, the similarity score is considered as the likelihood that users are interested in candidate news.

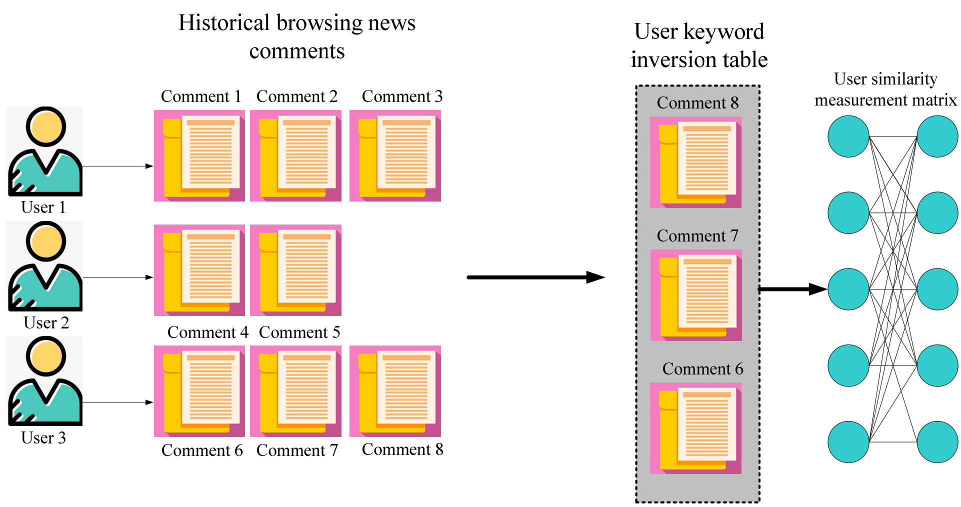
By processing news collection data, a set of news keywords in each news can be found. Data on the mapping of users to news keywords can be obtained by merging the mapping between users and their previously viewed news comments. The similarity of interests between users can be determined by using a dataset of news keywords, which is evaluated by users throughout their entire existence in the corpus. By assuming users a and b, $E(a)$ and $E(b)$ denote the set of news keywords evaluated by users a and b, respectively. The interest similarity of users a and b is calculated by (11):

$$L_{ab} = \frac{|E(a) \cap E(b)|}{|E(a) \cup E(b)|} \tag{11}$$

After obtaining the above interest similarity calculation, similarity calculation can be performed on all users to complete the division of similar user groups.

In real life, plenty of users do not have data on reading and commenting on the same news, so the computational effort can be reduced by constructing a backwards table of news keywords to users. The mapping relationship between news comments and user keyword inversion table is shown in Figure 4.

Figure 4 Mapping relationship between news comments and user keyword inversion table (see online version for colours)



As shown in Figure 4: the user keyword inverted table on the right (formed by selecting 3 keyword comments) is first gained from the mapping relationship data between users on the left and the history of browsed news comments, and then the user keyword inverted table is traversed to build a user similarity metric matrix.

When recommending news to users, the user group with similar interests is calculated. Afterwards, the news collection browsed by all users in this user group is treated as a recommended news list, and the recommended news list is traversed. Equation (12) is utilised to calculate the target user's liking level for each news item in the list of recommended news:

$$H(a, f) = \sum_{b \in S(a,k) \cap E(f)} L_{ab} g_{bf} \tag{12}$$

Among them, L_{ab} represents the similarity of interests between users a and b. After the above calculation, each recommended news data has a rating based on the user's liking level. According to this rating, the news list is sorted in descending order and used as the final recommended news data to users. This model takes the user's historical browsing behaviour and the characteristics of news content as input, and encodes the text and image information respectively through the multimodal feature extraction module. The attention mechanism is utilised to integrate the user interest vector with the news feature vector, and a matching score prediction function is constructed. The model adopts a contrastive learning strategy to optimise parameters, ensuring the accuracy of recommendations while enhancing diversity. By introducing the joint representation of image semantic features and text themes, the model can more accurately capture users' preferences for the visual style and content tendency of news. Calculate the similarity of interests among users to match their interests, and then screen out potential similar user groups. Based on the interaction behaviour of this group towards news, the system builds a candidate recommendation set and ranks the candidate news in combination with the multimodal matching score. Ultimately, the recommendation results are dynamically adjusted by integrating users' short-term interests and long-term preferences to enhance the accuracy of personalised services and the user experience. The loss function of the model adopts a weighted combination of cross-entropy loss and contrastive loss.

4 Effects of different recommendation algorithms

4.1 Time to obtain news of interest

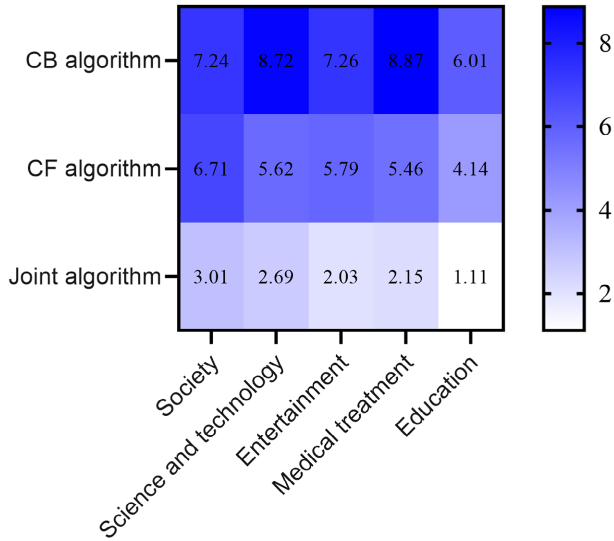
The experimental dataset has been provided earlier. This article compared and analysed the CB algorithm, CF recommendation algorithm, and joint recommendation algorithm (a combination of MTL recommendation algorithm and image matching algorithm). Due to time constraints, 30 social, technological, entertainment, medical, and educational news items were selected here, and the average time (s) of obtaining 30 different types of news items through CB algorithm, CF recommendation algorithm, and joint recommendation algorithm were calculated, as illustrated in Figure 5 (the horisontal axis of Figure 5 represents social, technological, entertainment, medical, and educational news items, while the vertical axis represents different algorithms).

As shown in Figure 5, the average time for obtaining social news through CB algorithm, CF recommendation algorithm, and joint recommendation algorithm was 7.24 seconds, 6.71 seconds, and 3.01 seconds, respectively; the average time to obtain educational news through CB algorithm, CF recommendation algorithm, and joint recommendation algorithm was 6.01 seconds, 4.14 seconds, and 1.11 seconds, respectively. It can be learned that the average time for the joint recommendation algorithm to get all kinds of news was less than 4s; that for the CB algorithm to get all kinds of news was more than 6s, and that for the CF recommendation algorithm to get all kinds of news is more than 4s.

4.2 Utilisation rate

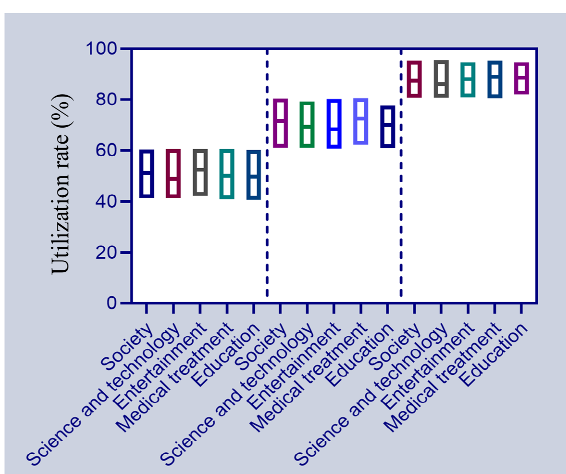
As the fast growth of the internet has led to a growing user base, all kinds of news media, public numbers, and free creators have emerged, triggering an exponential growth of data and information, and traditional news recommendation algorithms are facing many challenges.

Figure 5 Average time for news acquisition by different recommendation algorithms (see online version for colours)



The utilisation rates of CB algorithm, CF recommendation algorithm, and joint recommendation algorithm for social, technological, entertainment, medical, and educational news are shown in Figure 6 (the horizontal axis of Figure 6 shows social, technological, entertainment, medical, and educational news, and the vertical axis shows utilisation rates).

Figure 6 Utilisation rate of different algorithms for different types of news (see online version for colours)



As shown in Figure 6: part 1 of Figure 6 showed the utilisation rate of CB algorithm, which was within 70% for social, technological, entertainment, medical and educational news; part 2 of Figure 6 showed the utilisation rate of CF algorithm, which was within

80% for most of the social, technological, entertainment, medical and educational news; part 3 of Figure 6 showed the utilisation rate of the joint recommendation algorithm, which was more than 80% for the social, technological, entertainment, medical and educational news.

4.3 Recommendation accuracy

CB algorithm, CF recommendation algorithm, and joint recommendation algorithm were recommended for social, technological, medical, and educational news, and their recommendation accuracy was calculated. The study set the division ratio of the training set, validation set and test set at 6:2:2, and conducted model evaluation under the premise of ensuring balanced data distribution. During the experiment, the study utilised two recommendation algorithms to conduct recommendation tests on news related to society, science and technology, healthcare, and education. In the testing process, both were run 10 times on the same test set and the average was taken to reduce random errors. The CB algorithm recommends different news to users as shown in Table 4.

Table 4 Recommendation accuracy of CB algorithm

News type	*Number*	*Interested users*	*Not interested users*	*Recommendation accuracy rate (%)*
Society	5612	3949	1663	70.37
Science and technology	3143	2452	691	78.01
Entertainment	4010	3108	902	77.51
Medical treatment	2610	1900	710	72.80
Education	2628	2057	571	78.27

As shown in Table 4, the recommendation accuracy of CB algorithm was below 80%, with the highest recommendation accuracy for educational news being 78.27%. Recommendation accuracy of CF recommendation algorithms is depicted in Table 5.

Table 5 Recommendation accuracy of CF recommendation algorithm

News type	*Number*	*Interested users*	*Not interested users*	*Recommendation accuracy rate (%)*
Society	5612	4123	1489	73.47
Science and technology	3143	2765	378	87.97
Entertainment	4010	3269	741	81.52
Medical treatment	2610	2171	439	83.18
Education	2628	2289	339	87.10

As described in Table 5, the recommendation accuracy of the CF recommendation algorithm was mostly above 80%, with one recommendation below 80% for social news and the highest recommendation for technology news, at 87.97%. The recommendation accuracy of the joint recommendation algorithm is described in Table 6.

As shown in Table 6, the recommendation accuracy rate of the joint recommendation algorithm in various types of news exceeds 90%, which is significantly better than that of the CB and CF single algorithms. Among them, the recommendation accuracy rate of social news reached 93.30%, which was the highest among the five types of news.

Comprehensive comparison shows that the joint recommendation algorithm effectively improves the accuracy of recommendations while maintaining a high coverage rate.

Table 6 Recommendation accuracy of joint recommendation algorithms

News type	*Number*	*Interested users*	*Not interested users*	*Recommendation accuracy rate (%)*
Society	5612	5238	374	93.30
Science and technology	3143	2917	226	92.81
Entertainment	4010	3644	366	90.87
Medical treatment	2610	2368	242	90.73
Education	2628	2394	234	91.10

At present, the analysis of recommendation algorithms mainly focuses on user interest models. However, in the real world, there is a massive amount of sparse browsing data in news recommendations. Using CB algorithm and CF recommendation algorithm can easily overlook the useful information contained in it, thereby reducing the accuracy of recommendations. The less browsing data, the lower the accuracy of recommendations. The accuracy of recommendations depends on the user's rating. Improving news recommendation algorithms using machine learning methods can significantly enhance the performance of recommendation algorithms (Minaee et al., 2021).

To further verify the stability and effectiveness of the proposed joint recommendation algorithm, the study compares it with the CF and matrix factorisation methods. The comparison indicators include AUC and NDCG. All three recommendation algorithms run on the same dataset, and the parameter settings remain consistent. The results are shown in Table 7.

Table 7 performance comparison of different recommendation algorithms

Method	*AUC*	*NDGG@10*	*NDGG@50*	*NDGG@100*
Joint recommendation algorithm	0.93	0.76	0.81	0.84
Collaborative filtering	0.87	0.68	0.72	0.74
Matrix factorisation	0.89	0.70	0.75	0.78

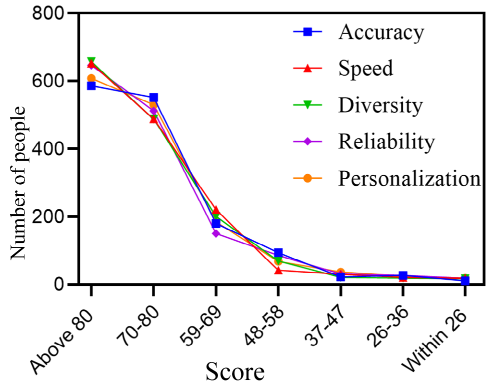
As shown in Table 7, the joint recommendation algorithm outperforms the CF and matrix factorisation methods in AUC and various NDCG indicators, especially in NDCG@100, reaching 0.84, indicating that it has a stronger ranking ability in long-tail news recommendation.

4.4 User satisfaction

The satisfaction of users with recommended content is not only related to accuracy, but also depends on the level of assistance in making decisions for users and their interests. The more recommended content aligns with user interests, the higher user satisfaction.

The 1,473 users were selected to rate the accuracy, speed, diversity, reliability, and personalisation of news recommendations using joint recommendation algorithms. The user's rating of news recommended by the joint recommendation algorithm is illustrated in Figure 7 (the horizontal coordinate denotes the score and the vertical coordinate denotes the number of people).

Figure 7 Users' ratings of news recommended by the joint recommendation algorithm (see online version for colours)



As shown in Figure 7, it can be seen that the accuracy, speed, diversity, reliability, and personalisation of news recommended by the joint recommendation algorithm were the highest among those with scores above 80, all exceeding 500, followed by those with scores between 70 and 80, ranking second, and those with scores below 26 were the lowest.

Taking into account all the interests of users, the degree of user preference was analysed. By calculating the similarity of projects, the novelty and diversity of recommendation results were measured, ensuring accuracy while improving user satisfaction.

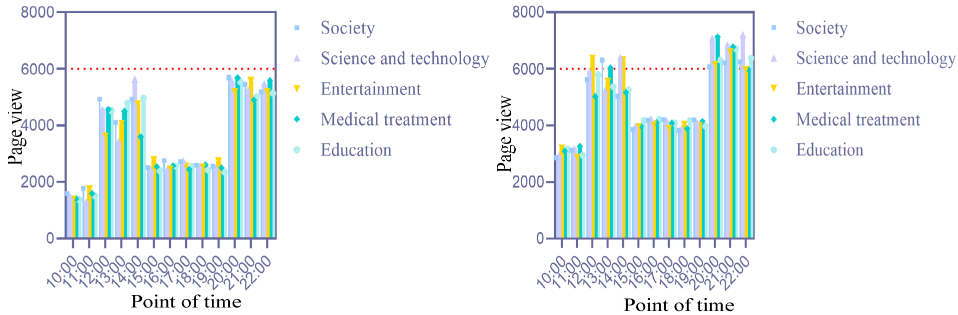
4.5 Page views

With the emergence of various news clients, reading news anytime and anywhere has become a daily behaviour of people. At the same time, the cost and time of news production have been greatly reduced, and any media and individual can produce news information data at any moment. In the information age, various types of data and information are growing at a geometric exponential rate, which has caused serious information overload problems for network users.

The number of views on a certain webpage on June 1, 2022 (where webpage S is not related to the data above) that did not use the joint recommendation algorithm and the number of views on webpage S that used the joint recommendation algorithm on June 5, 2022 were counted. The findings are illustrated in Figure 8 (the horizontal coordinate denotes the time period and the vertical coordinate denotes the number of views).

As Figure 8: Figure 8(a) showed the browsing volume of webpage S during various time periods without using the joint recommendation algorithm, and it can be seen that the highest browsing volume was below 6000; Figure 8(b) showed the browsing volume of webpage S during various time periods when using the joint recommendation algorithm, and it can be seen that the highest browsing volume was above 6000.

Figure 8 Browsing volume without and using joint recommendation algorithms, (a) browsing volume without using joint recommendation algorithm (b) browsing volume using joint recommendation algorithm (see online version for colours)



In the field of news recommendation, joint recommendation algorithms can discover the news that users love to browse, try to recommend the news that users are most likely to click to browse, in order to increase user click through rate, user stay time, and user retention rate, thus obtaining higher advertising profits.

5 Discussion

The average time for CB algorithm and CF recommendation algorithm to obtain news is longer. Users need to spend much time and energy to find the news they need, and news is a passive service. Although traditional CB algorithms and CF recommendation algorithms can help people obtain the information they want, they cannot meet all the requirements of users. Therefore, it is hard for users to describe the information they are searching for with a precise and appropriate keyword, and in many cases, relying solely on keywords for retrieval cannot meet their requirements.

The news utilisation rate recommended by the joint recommendation algorithm is higher. Among them, the joint recommendation algorithm utilises text and image matching to make recommendations simultaneously, which has been used in many news websites and clients. After adopting the joint recommendation algorithm, there is a trend of increasing user activity and actual usage time for various websites. Users feel that the recommended news content is very suitable for their personalised needs, thereby improving the quality of recommended news.

MTL can fully explore the domain information contained in each task sample, effectively improve the model's generalisation ability, and share expressions among multiple tasks, thereby effectively improving the performance of each task in the model. In addition, MTL can also be considered as the induction and transfer of knowledge, which refers to the application of existing knowledge to other related problems to improve learning effectiveness. At the same time, MTL can also achieve information sharing among multiple tasks and solve the sparsity problem of data in some tasks.

The users have high satisfaction with the joint recommendation algorithm. The primary goal of recommendation services is to provide personalised information to users. Unlike search engine technology, it applies the principle of information push to meet users' needs for personal information, that is, to provide recommendations based on users' specific requirements, or to actively provide the necessary information resources to users based on the analysis of their personality and usage habits. As internet technology evolves, more and more users are looking to utilise federated recommendation algorithms so that they can be provided with better recommendations.

6 Conclusions

The news data on the internet has shown an explosive development trend. On the one hand, these huge amounts of data provide people with more convenience, but on the other hand, people have to spend much energy and time searching for useful information for themselves. In a massive data environment, traditional recommendation algorithms can no longer realise the needs of users well. News data has a wide range of sources, complex structures, and an increasing number. In the massive amount of news data, traditional recommendation algorithms usually consume a lot of labour, material and time, which leads to information overload. Nowadays, news recommendation has been widely applied on the internet. Recommendation algorithms based on machine learning and image matching algorithms can utilise the mining of multimodal information of users, analyse their behaviour, and perform image matching on different users, so as to accurately analyse users' interests, providing them with information that interests them and meeting their personalised needs.

Declarations

The data used to support the findings of this study are all in the manuscript.

The author declares no competing interests.

References

Afsar, M.M., Crump, T. and Far, B. (2022) 'Reinforcement learning based recommender systems: a survey', *ACM Computing Surveys*, Vol. 55, No. 7, pp.1–38, https://doi.org/10.1145/3543846.

Benkessirat, S., Boustia, N. and Nachida, R. (2021) 'A new collaborative filtering approach based on game theory for recommendation systems', *Journal of Web Engineering*, Vol. 20, No. 2, pp.303–326, https://10.13052/jwe1540-9589.2024.

Carlson, M. (2018) 'Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism', *New Media & Society*, Vol. 20, No. 5, pp.1755–1772, https://doi.org/10.1177/1461444817706684.

Chen, H., Yin, C., Li, R., Rong, W., Xiong, Z. and David, B. (2019) 'Enhanced learning resource recommendation based on online learning style model', *Tsinghua Science and Technology*, Vol. 25, No. 3, pp.348–356, https://doi.org/10.26599/TST.2019.9010014.

Collins, B., Hoang, D.T., Nguyen, N.T. and Hwang, D. (2021) 'Trends in combating fake news on social media – a survey', *Journal of Information and Telecommunication*, Vol. 5, No. 2, pp.247–266, https://doi.org/10.1080/24751839.2020.1847379.

Egelhofer, J.L. and Lecheler, S. (2019) 'Fake news as a two-dimensional phenomenon: a framework and research agenda', *Annals of the International Communication Association*, Vol. 43, No. 2, pp.97–116, https://doi.org/10.1080/23808985.2019.1602782.

Faroek, D.A., Umar, R. and Riadi, I. (2022) 'Classification based on machine learning methods for identification of image matching achievements', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Vol. 6, No. 2, pp.198–206, https://doi.org/10.29207/resti.v6i2.3826.

Gelfert, A. (2018) 'Fake news: a definition', *Informal Logic*, Vol. 38, No. 1, pp.84–117, https://doi.org/10.22329/il.v38i1.5068.

Haim, M., Graefe, A. and Brosius, H.B. (2018) 'Burst of the filter bubble? Effects of personalization on the diversity of Google News', *Digital Journalism*, Vol. 6, No. 3, pp.330–343, https://doi.org/10.1080/21670811.2017.1338145.

Helberger, N., Karppinen, K. and D'acunto, L. (2018) 'Exposure diversity as a design principle for recommender systems', *Information, Communication & Society*, Vol. 21, No. 2, pp.191–207, https://doi.org/10.1080/1369118X.2016.1271900.

Javed, U., Shaukat, K., Hameed, I.A., Iqbal, F., Alam, T.M. and Luo, S. (2021) 'A review of content-based and context-based recommendation systems', *International Journal of Emerging Technologies in Learning*, Vol. 16, No. 3, pp.274–306, https://www.learntechlib.org /p/219036/.

Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K.M. and Trulls, E. (2021) 'Image matching across wide baselines: from paper to practice', *International Journal of Computer Vision*, Vol. 129, No. 2, pp.517–547, https://doi.org/10.1007/s11263-020-01385-0.

Karimi, M., Jannach, D. and Jugovac, M. (2018) 'News recommender systems – survey and roads ahead', *Information Processing & Management*, Vol. 54, No. 6, pp.1203–1227, https://doi.org/10.1016/j.ipm.2018.04.008.

Li, C.T., Hsu, C. and Zhang, Y. (2022) 'Fairsr: fairness-aware sequential recommendation through multi-task learning with preference graph embeddings', *ACM Transactions on Intelligent Systems and Technology*, Vol. 13, No. 1, pp.1–21, https://doi.org/10.1145/3495163.

Liu, J. and Bu, F. (2019) 'Improved RANSAC features image-matching method based on SURF', *The Journal of Engineering*, No. 23, pp.9118–9122, https://doi.org/10.1049/joe.2018.9198.

Makhortykh, M. and Bastian, M. (2022) 'Personalizing the war: perspectives for the adoption of news recommendation algorithms in the media coverage of the conflict in Eastern Ukraine', *Media, War & Conflict*, Vol. 15, No. 1, pp.25–45, https://doi.org/10.1177/1750635220906254.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M. and Gao, J. (2021) 'Deep learning-based text classification: a comprehensive review', *ACM Computing Surveys*, Vol. 54, No. 3, pp.1–40, https://doi.org/10.1145/3439726.

Mizgajski, J. and Morzy, M. (2019) 'Affective recommender systems in online news industry: how emotions influence reading choices', *User Modeling & User-Adapted Interaction*, Vol. 29, No. 2, pp.345–379, https://doi.org/10.1007/s11257-018-9213-x.

Mohammadpour, T., Bidgoli, A.M., Enayatifar, R. and Javadi, H.H.S. (2019) 'Efficient clustering in collaborative filtering recommender system: hybrid method based on genetic algorithm and gravitational emulation local search algorithm', *Genomics*, Vol. 111, No. 6, pp.1902–1912, https://doi.org/10.1016/j.ygeno.2019.01.001.

Molina, M.D., Sundar, S.S., Le, T. and Lee, D. (2021) ''Fake news' is not simply false information: A concept explication and taxonomy of online content', *American Behavioral Scientist*, Vol. 65, No. 2, pp.180–212, https://doi.org/10.1177/0002764219878224.

Najafabadi, M.K., Mohamed, A. and Onn, C.W. (2019) 'An impact of time and item influencer in collaborative filtering recommendations using graph-based model', *Information Processing & Management*, Vol. 56, No. 3, pp.526–540, https://doi.org/10.1016/j.ipm.2018.12.007.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M. and Liu, Y. (2019) 'Combating fake news: a survey on identification and mitigation techniques', *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 10, No. 3, pp.1–42, https://doi.org/10.1145/3305260.

Shen, J., Zhou, T. and Chen, L. (2020) 'Collaborative filtering-based recommendation system for big data', *International Journal of Computational Science and Engineering*, Vol. 21, No. 2, pp.219–225, https://doi.org/10.1504/IJCSE.2020.105727.

Shu, J., Shen, X., Liu, H., Yi, B. and Zhang, Z. (2018) 'A content-based recommendation algorithm for learning resources', *Multimedia Systems*, Vol. 24, No. 2, pp.163–173, https://doi.org/10.1007/s00530-017-0539-8.

Thurman, N., Lewis, S.C. and Kunert, J. (2019a) 'Algorithms, automation, and news', *Digital Journalism*, Vol. 7, No. 8, pp.980–992, https://doi.org/10.1080/21670811.2019.1685395.

Thurman, N., Moeller, J., Helberger, N. and Trilling, D. (2019b) 'My friends, editors, algorithms, and I: examining audience attitudes to news selection', *Digital Journalism*, Vol. 7, No. 4, pp.447–469, https://doi.org/10.1080/21670811.2018.1493936.

Van den Bulck, H. and Moe, H. (2018) 'Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas', *Media, Culture & Society*, Vol. 40, No. 6, pp.875–892, https://doi.org/10.1177/016344371773440.

Wallace, J. (2018) 'Modelling contemporary gatekeeping: the rise of individuals, algorithms and platforms in digital news dissemination', *Digital Journalism*, Vol. 6, No. 3, pp.274–293, https://doi.org/10.1080/21670811.2017.1343648.

Wang, S., Cao, L., Wang, Y., Sheng, Q.Z., Orgun, M.A. and Lian, D. (2021) 'A survey on session-based recommender systems', *ACM Computing Surveys (CSUR)*, Vol. 54, No. 7, pp.1–38, https://doi.org/10.1145/3465401.

Waszak, P.M., Kasprzycka-Waszak, W. and Kubanek, A. (2018) 'The spread of medical fake news in social media–the pilot quantitative study', *Health Policy and Technology*, Vol. 7, No. 2, pp.115–118, https://doi.org/10.1016/j.hlpt.2018.03.002.

Zhang, C., Yang, M., Lv, J. and Yang, W. (2018) 'An improved hybrid collaborative filtering algorithm based on tags and time factor', *Big Data Mining and Analytics*, Vol. 1, No. 2, pp.128–136, https://doi.org/10.26599/BDMA.2018.9020012.

Zhang, P., Zhang, Z., Tian, T. and Wang, Y. (2019a) 'Collaborative filtering recommendation algorithm integrating time windows and rating predictions', *Applied Intelligence*, Vol. 49, No. 8, pp.3146–3157, https://doi.org/10.1007/s10489-019-01443-2.

Zhang, S., Yao, L., Sun, A. and Tay, Y. (2019b) 'Deep learning based recommender system: a survey and new perspectives', *ACM Computing Surveys*, Vol. 52, No. 1, pp.1–38, https://doi.org/10.1145/3285029.

Zhang, Y. and Yang, Q. (2021) 'A survey on multi-task learning', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 12, pp.5586–5609, https://doi.org/10.1109/TKDE.2021.3070203.

Zhao, W., Yan, L. and Zhang, Y. (2018) 'Geometric-constrained multi-view image matching method based on semi-global optimization', *Geo-spatial Information Science*, Vol. 21, No. 2, pp.115–126, https://doi.org/10.1080/10095020.2018.1441754.