

Improving News Popularity Estimation via Weak Supervision and Meta-active Learning

Mona Nashaat
University of Alberta
nashaata@ualberta.ca

James Miller
University of Alberta
jimm@ualberta.ca

Abstract

Social news has fundamentally changed the mechanisms of public perception, education, and even dis-information. Apprising the popularity of social news articles can have significant impact through a diversity of information redistribution techniques. In this article, an improved prediction algorithm is proposed to predict the long-time popularity of social news articles without the need for ground-truth observations. The proposed framework applies a novel active learning selection policy to obtain the optimal volume of observations and achieve superior predictive performance. To assess the proposed framework, a large set of experiments are undertaken; these indicate that the new solution can improve prediction performance by 28% (precision) while reducing the volume of required ground truth by 32%.

1. Introduction

Social news portals have become an essential source of information. News is increasingly consumed on the away from traditional settings such as the home. The continuous news cycle is an ideal vehicle for mobile presentation and consumption. Since they permit simple access to the latest news alongside easy integration of social media platforms, the amount at which new content is published has reached extraordinary rates [1]. According to Pew research center¹, almost two-thirds of U.S. adults are reported as Facebook users, and more than 40% of this percentage adults rely on Facebook to get news and recent updates. However, the popularity of news articles tends to show an unbalanced distribution. Previous studies [2] show that 73% of people usually skim blog articles while the rest admit to check them thoroughly. As a result, only a small percentage of the published articles gain high popularity inferred with an increased number of votes [2], comments [3], or shares

on social media [2]. Hence, in a fundamental way, the value of the mobile-consumer interface is defined as the popularity and reach of content.

Thus, precise assessment of content acceptance and predicting its popularity dynamics can have valuable implications in many areas such as social marketing and online content generation. For instance, a predictive system that estimates news popularity can recommend how news articles should be arranged in social portals to enhance the user browsing experience.

Consequently, several studies [2]–[9] proposed techniques to predict content popularity. Some studies [3], [6], [7] proposed different approaches for evaluating popularity, like examining the popularity of offline content [6] or evolution patterns [7]. Also, some research [8], [9] experimented with different models to recommend a generic model for popularity predictions. However, the existing research shows several gaps and challenges. One challenge is to decide on which metrics should be applied to express popularity [4], such as the number of user comments, the rating values, or the number of shares through social media. In many real-world applications, these metrics can be combined or even used interchangeably. Moreover, linking popularity metrics with the correct set of predictive features is an essential part of feature engineering [10]. Adopting different features according to each metric can be both expensive and time-consuming. Furthermore, several popularity factors, such as the quality of the written content or the importance of article topics to end-users, are difficult to quantify, which could further complicate the process of feature engineering.

However, the advent of new techniques of deep neural learning can alleviate most of the challenges associated with feature engineering by learning the task-specific representation of data. Nevertheless, this comes with another major cost as these techniques need massive training examples to achieve top

¹ <https://www.hubspot.com/marketing-statistics>

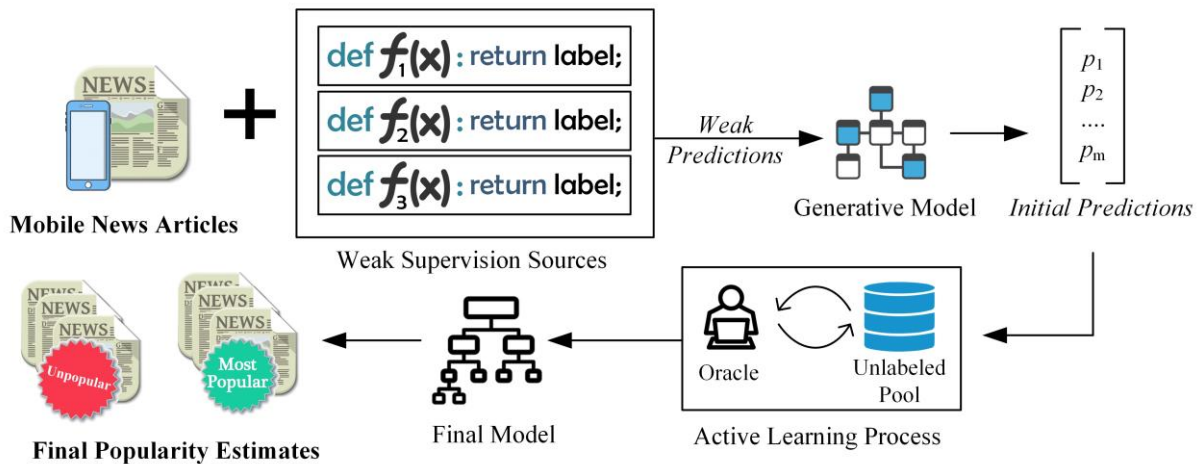


Figure 1. Overview of the proposed method

performance. Obtaining hand-labeled datasets is considered as another expensive task in the machine learning pipeline. Moreover, developing predictive systems for social content popularity depends on many varying factors, such as the structure of the news portal or the type of datasets. Therefore, different models may be required for each situation.

Moreover, changing the settings of any of these factors may result in rebuilding the model [3]. Additionally, most of the existing models are developed using publicly available datasets, which may not always be accurate or even complete. Therefore, acquiring labeled datasets for such diverse settings had turned out to be an expensive yet indispensable task in the task of predicting the popularity of news articles.

Therefore, motivated by the shortcomings of these approaches, in this article, we present ArtAI an improved prediction scheme to predict the long-time popularity of news articles without the need for ground-truth observations. The scheme extends our previous work [11], which is a labeling framework that combines Weak Supervision with Active Learning to create large-scale, high-quality training data. However, ArtAI applies a novel selection policy to engage the end-users in the process. Therefore, instead of applying traditional sampling techniques of active learning, the approach frames the active learning process as a regression problem to design the selection policy based on the underlying data distribution. To extend weakly generated labels [10], the proposed selection policy rectifies the inaccurate data points. Figure 1 illustrates an overview of the proposed model; the approach takes a collection of news articles as an input. Then, the proposed approach employs a set of weak sources to produce initial popularity estimates for the input collection. ArtAI can work with any weak supervision sources, including crowdsourced labels and knowledge bases. However,

the experiments focus on user-defined heuristics in the form of labeling functions [10], [12] since they are the most common mechanism to define weak labels [10]. After that, ArtAI applies a meta-active learning process to query the user to provide labels for the most useful observations. The output of ArtAI is a trained model for popularity prediction, along with the final predictions generated by the learned selection process.

An extensive set of experiments are performed to evaluate the proposed scheme and compare it with three state-of-the-art techniques. The comparing approaches include an ensemble model [8], a vector space model [13], a gradient boosting learning approach [14] along with traditional active learning strategies to predict social content popularity [15]. The experimental evaluation aims to estimate the proposed model's effectiveness in popularity predictions with different classification models and a set of datasets with various sizes and dimensionality.

Moreover, to assess the impact of the experimental parameters, sensitivity analysis is conducted in which the labeling budget of ArtAI is adjusted according to the number of annotations consumed by traditional active learning.

The article is structured as follows: Section 2 discusses the related background. Section 3 presents the proposed method. The experimental results are discussed in Section 4; while Section 5 concludes the article.

2. Related work

The proposed approach combines weak popularity estimates along with meta-active learning [15] to predict popularity for social news. Hence, the related work spans across various research topics, such as active learning, employing machine learning for popularity prediction, and handling weakly supervised data.

Active learning [15] is a special kind of semi-supervised learning in which a learner algorithm gets to choose which examples are added to the training set. This paradigm is proven to generate highly accurate models with minimum labeling effort. Active learning performs efficiently in situations where a large portion of the data is unlabeled, which is usually the case with social content. Most training data of social content are crawled from news portals that do not provide labels along with the data. Hence, active learning can be significantly useful in these settings. Active learning engages the users into the loop by asking them to label information to enhance the training performance of the underlying classifier. In pool-based active learning, the process is initialized with a small number of labeled instances (the seed) and a pool of unlabeled observations X_{train} . Then the learning algorithms iteratively ask the user to provide true labels for specific points from the pool. These points are then moved to the labeled set and used to retrain the classification model. The model is then evaluated using a held-out test set D_{test} , and the process is repeated. The iterative process terminates when either a performance threshold is reached, or a predefined annotation budget is exceeded. In active learning, the algorithm that decides which data instances the users should provide true labels is called the query strategy. There are many traditional query strategies [15], such as uncertainty sampling that queries the user to provide labels for the samples about which the learner is most uncertain. Another selection policy is query-by-committee, which also queries the most uncertain samples. However, it measures the uncertainty differently, as it uses a committee of classifiers and queries the instance about which the committee members disagree.

Previous studies [19]–[21] have applied active learning to different applications. For example, authors in [19] presented a human-machine collaborative model to detect misleading information in social content. The system applies active learning to cope with the problem of limited annotated samples. The system combines neural networks with active learning to reduce the labeling cost while attaining an acceptable performance. Another study [20] utilized active learning to identify malicious content in social media. The proposed model [20] initially creates a view-dependent classifier from a small labeled data and then applies active learning to enhance the model performance with additional annotated examples. Moreover, another system is presented in [21] to classify fake news by randomly selecting different sets of features to create a huge number of unbiased models; then, these models are ranked to define the best outcomes. However, although active learning has

been applied to a wide range of applications, none of these approaches has tried to examine the problem of predicting the popularity of social news. Although, since most of the publicly available datasets are known to be inaccurate, active learning can provide suboptimal solutions due to the high level of noise in input data [15].

Alternatively, previous studies [16]–[18] have focused on feature engineering as one of the challenges that face popularity estimation. For example, authors in [16] applied vocabulary clustering to social content to detect similar patterns of popular topics. Then, the model is used to estimate long-term popularity. Another research [17] presents a preliminary analysis of content popularity before developing a regression model that employs the analysis results to predict popular trends in the future. Moreover, Bao et al. [18] proposed a method that observes mobile social content to decide on the most significant attributes to build the final feature-driven model. However, most of these approaches are content-specific. For example, they focus on certain types of content, such as videos [16] and tweets [18]. Therefore, the final models are restricted to analyze content history within a single observed domain. Moreover, unlike ArtAI, none of these techniques have tried to include any domain experience in the learning process.

Finally, weakly supervised datasets [22] have been gaining popularity in machine learning tasks. Since obtaining hand-labeled large datasets has turned to be an impractical in many applications [22], inexpensive weakly supervised labels can be utilized to create accurate predictive models. In weak supervision, subject-matter experts provide some form of higher-level, low-quality supervision sources like user-defined labeling function and knowledge bases [22] to create training labels which are expected to be noisy. Since weakly supervised datasets are mostly applied to applications where obtaining accurately labeled datasets can be expensive, previous research [23], [24] has focused on text understanding, document categorization, and intent classification. For example, Meng et al. [23] have proposed a weakly-supervised method for text classifications. The model first generates a pseudo-document to pre-train the model and then fine-tune it using real unlabeled data. ArtAI applies different types of weak supervision to obtain enough training data for deep learning models. Alternatively, another recent study [24] utilizes weak supervision sources from social media to detect fake news articles with limited labeled data. The research [24] proposes a framework in which data is first collected from multiple weak sources to train a model. Then, the model runs an inference module to use the

```

def LF_popularity_estimate(x):
    if x.words_count > 3000:
        return 1;
    else:
        return 0;

```

Figure 2. Example of a user-defined labeling function the predicts popularity based on the count of words in an article

learned feature representation to predict labels for unseen data.

However, a closer look at these efforts reveals a number of shortcomings. First, applying weak sources usually results in imperfect data with conflicting and noisy data points, which affects the performance of the final model. Although most of these approaches [23], [24] have tried to automatically de-noise the data, the complex structure of these models makes it challenging for users to trust their outcomes. Secondly, none of these approaches [23], [24] have tried to engage the users in the process of training the model or assessing its performance to increase user trust. Therefore, the effectiveness of engaging the user to debug these weakly supervised sources in the domain of predicting news popularity is yet to be tested, which is what this research tries to accomplish.

3. ArtAI: The proposed method

The input to the proposed system is a collection of news articles D_N characterized as $\{\mathbf{x}_i, y_i\}_{i=1}^N$ where \mathbf{x}_i is a set of features depicting the i^{th} article in the dataset, and y_i denotes the unknown popularity flag associated with this point. As for the input $\mathbf{x}_i \in R^F$ is described as a set of A attributes to represent each article. For example, the attributes for a given article can include the number of links and images the article contains and its title subjectivity [3]. Since these attributes are a set of quantifiable features of the observed article, the set of attributes describing the i^{th} article can be represented by a feature vector \mathbf{x}_i . ArtAI also requires a small labeled set of articles of size M as $D_M = \{\mathbf{x}_i^*, y_i^*\}_{i=1}^M$ with known popularity y_i^* where $M \ll N$. As for the output, the final model predicts popularity flags for the articles in D_N as a boolean label where $y_i^* \in \{-1, 1\}$.

As Figure 1 shows, the proposed model starts by letting the users provide a group of F labeling functions of size L described as $\{f_j\}_{j=1}^L$, where $f_j: \mathbf{X} \rightarrow \{-1, 0, 1\}$. In other words, each labeling function outputs a weak prediction for each article in D_N to denote its anticipated popularity based on some user-defined heuristics. An example of a labeling function in Figure 2. As the figure shows, the function can

either output a weak prediction $\{-1, 1\}$, or abstain $\{0\}$. Consequently, the result of applying all the labeling functions F to X is a sparse matrix S where:

$$S_{i,j} = f_j(\mathbf{x}_i) \text{ where } 1 \leq i \leq N \text{ and } 1 \leq j \leq L \quad (1)$$

Afterward, ArtAI applies a generative model M_G [25] to model the accuracies of these labeling functions. The generative model models S a factor graph by encoding three factors: labeling propensity, labeling accuracy, and the function correlation for each pair of functions. These factors can be formally defined respectively as:

$$\emptyset_{i,j}^{\text{lab}}(F, Y) = \mathbf{1}\{f_j(\mathbf{x}_i) \neq 0\} \quad (2)$$

$$\emptyset_{i,j}^{\text{Acc}}(F, Y) = \mathbf{1}\{f_j(\mathbf{x}_i) = y_i\} \quad (3)$$

$$\emptyset_{i,j,k}^{\text{Corr}}(F, Y) = \mathbf{1}\{f_j(\mathbf{x}_i) = f_k(\mathbf{x}_i)\} \quad (4)$$

where $f_j, f_k \in P$ and P is the set of all functions pairs in L [26]. As mentioned earlier, these labeling functions depend on imperfect user-defined heuristics. Therefore, their outputs conflict and disagree on certain points or even abstain, which results in incomplete data. Hence, the proposed method formally describes the pairwise disagreements as:

$$\emptyset_{i,j,k}^{\text{dis}}(F, Y) = \mathbf{1}\{f_j(\mathbf{x}_i) \neq f_k(\mathbf{x}_i)\} \text{ where } j, k \in P, 1 \leq i \leq N \quad (5)$$

Furthermore, the method denotes the abstaining conditions, as

$$\emptyset_{i,j}^{\text{abstain}}(F, Y) = \mathbf{1}\{f_j(\mathbf{x}_i) = 0\} \quad (6)$$

At this point, ArtAI tries to enhance the accuracy of the labeling function by applying a meta-active learning process. The process designs the selection policy by framing the problem as a regression problem. The active learning stage aims at training a selection policy so that, when applied to a dataset, it selects the data points that would result in the maximum reduction to the generalization error. The process consists of two main steps, namely, *designing the selection policy* and *applying the policy as a meta-active learning process*.

First, as for designing the selection policy, the process is formulated as a regression problem. To initialize the regression process, the method first collects a set of labeled observation $D_S = \{y_i, \nabla_i\}_{i=1}^Q$ to train the selection policy where y_i describes a set of attributes for the i^{th} example in D_S . To only include the attributes that are related to data distribution, the

model considers the values of the factors in Equations (2)-(6). On the other hand, ∇_i describes the prospective reduction to the generalization error after adding the i^{th} point to the labeled pool. To gather these labeled examples, the model first classifies the points in S into high-conflicting points P_H and low-conflicting points P_L . The high-conflicting dataset contains the points about which the labeling functions are disagreeing or abstaining. It can be defined as:

$$P_H \subseteq \mathbf{X}, \forall x_i \in D_N \left\{ x_i | \emptyset_{i,j,k}^{\text{dis}}(F, Y) = \mathbf{1}\{f_j(x_i) \neq f_k(x_i)\} \cup \emptyset_{i,j}^{\text{abstain}}(F, Y) = \mathbf{1}\{f_j(x_i) = 0\} \right\} \quad (7)$$

while the low-conflicting points are denoted as:

$$P_L = D_M \cup (D_N - P_H) \quad (8)$$

Then, the low-conflicting points P_L is used to train and evaluate a model M_S . The model is first trained and evaluated on a subset of P_L so the initial generalization error L_g is recorded using a testing set. The generalization error is the average value of the loss weighted by how likely those examples are in the dataset. Then, ArtAI iteratively adds a new point x from P_L to the training dataset. After that, the model is evaluated again to record the generalization error related to this point L_x . Finally, the reduction in the classification loss is computed and recorded as $\nabla_x = L_g - L_x$. Consequently, the result of this process is the new training dataset D_s that is used later to train the regressor.

Second, D_s is then used to train a random forest regressor g [9] as the final selection policy that is built while considering the distribution of the underline space matrix S . The selection policy is then applied to P_H to greedily choose the points with the highest potential error reduction by taking the maximum of the value predicted by the regressor g as:

$$x^* = \arg \max_{x \in D_{\text{Test}}} g(y_x) \quad (9)$$

The model then applies the regressor function g to rank the data points in P_H . The time complexity of the ranking step is highly decreased as the size of P_H is much smaller the number of articles in D_N . Therefore, in each iteration of the active learning process, the regressor function ranks the points in P_H using (9). Then, the points denoting the articles with the highest reduction in the generalization error are selected. Next, the user is queried to provide true labels these points, which are then added to the set of final predictions. Finally, this set of predictions is used to retrain a classifier f for news popularity. As the iterations of active learning progress, ArtAI gradually builds a set of predictions D_{AL} which represents the data points that received true labels from the user during this stage. The process also outputs a predictive model f which is trained using D_{AL} and can estimate popularity for unseen articles. The framework is implemented in

Python, and a complete algorithm of the proposed method is shown in Algorithm 1.

Algorithm 1 The Proposed Method

Input: a collection of news articles D_N , small labeled dataset D_M , a set of labeling functions F , predefined labeling cost.

Output: Final classifier f for popularity predictions.

- 1: Apply F to D_N to generate a label matrix S .
 - 2: Compute disagreements factor \emptyset^{dis} (Equation 5).
 - 3: Compute abstaining factor $\emptyset^{\text{abstain}}$ (Equation 6).
 - 4: Classify S into P_H and P_L (Equation 7-8).
 - 5: Split P_L into training and testing sets.
 - 6: $D_S = \emptyset$.
 - 7: Train a classification model M_S using a subset of P_L .
 - 8: Calculate the test loss L_g .
 - 9: **Loop** for each point in training set:
 - 10: Add a point x to the training set.
 - 11: Calculate the new test loss L_x .
 - 12: Compute the reduction in classification loss ∇_x .
 - 13: Collect the data point parameters γ_x .
 - 14: $D_S = D_S \cup \gamma_x, \nabla_x$.
 - 15: **End**
 - 16: Train a random forest regressor g using D_S .
 - 17: $D_{AL} = \emptyset$.
 - 18: Initialize the unlabeled pool as P_H .
 - 19: **Loop** until labeling cost is exceeded:
 - 20: Apply g to select a point x_i from P_H (Equation 8).
 - 21: Ask the user to provide a true label for x_i .
 - 22: $D_{AL} = D_{AL} \cup x$.
 - 23: Train classifier f using D_{AL} .
 - 24: **End**
 - 25: return f .
-

4. Experimental evaluation

To estimate the effectiveness of ArtAI, the experimental evaluation considers different metrics of classification performance along with the amount of training examples needed to train each of the methods engaged in the evaluation.

4.1. Description of datasets

The experiments incorporate a set of datasets with different sizes and dimensionality. A description of datasets is summarized in Table 1. The table shows, for each dataset, the size of the data (Size), the number of attributes (Dim.), the popularity measure that is used in the experiments (Popularity Measure), and the ratio of the positive class (popular articles) to the dataset size (+/Size). The first dataset used in experiments is the **Mobile social News Popularity (mobile social News)**, which is a real-world dataset

Table 1. Overview of the datasets.

Dataset	Size	Dim.	Popularity Measure	+/- Size
News	39,797	61	# shares	49.34
Reddit Engagement	89,314	12	# comments (Reddit)	13.12
Webhose News	170,882	84	# comments (Facebook)	33.19

offered by the University of California at Irvine (UCI) Machine Learning Repository. It contains more than 39k articles with 61 attributes. The popularity term is measured by the number the article URL is shared on twitter. Also, the experiments include **Reddit Community Engagement Dataset (Reddit Engagement)**, which is a dataset of Reddit news articles crawled for three months from June to August 2017. The dataset contains 89,314 news posts with 12 attributes. Finally, **Webhose's Popular News Article (Webhose News)** is another real-world dataset that is provided by Webhose. The dataset has more than 170,000 news articles with 84 attributes.

4.2. Experiments settings

The experiments compare ArtAI with three baseline strategies:

- **Gradient boosting learning approach (GBM)** presented in [14]. The technique extends gradient boosting models to predict the number of shares using an ensemble of learning algorithms.
- **Vector space model (VSM)** proposed in [13], which applies a two-stage selection approach to predict news popularity. The method selects global features related to column information and then chooses local features related to news popularity.
- **Ensemble models (Ensemble)** presented in [8], which utilizes a group of predictive models to attain superior performance. The approach convenes decision trees along with boosting and bagging to achieve higher classification accuracy.

As for user-defined heuristics, the experiments consider threshold-based labeling function. In this type, the function assigns a popularity prediction to a given article based on certain attributes (e.g., number

of images in the article). Following the best practice found in the literature [10], [12], [27], the experiments rely on pattern matching methods to create the labeling function used in the experiments. Furthermore, to develop high accuracy labeling functions, the experiments used the set of labeled articles D_M to develop and evaluate the empirical accuracy of the generated functions. The proposed method only accommodates the labeling functions with accuracy more than a predefined threshold of 60% [10]. The experimental settings for the proposed method are presented in Table 2.

Also, the experiments must set a stopping condition for the iterative active learning process. To select the stopping condition, another set of experiments are conducted with different sampling techniques of active learning. The experiments applied uncertainty sampling (UNC), query-by committee (QBC), and random sampling (RAND) [15] with each dataset and examined the learning curves in each situation. The experiments are averaged over ten runs and stopped the active learning process when the learning curve shows no enhancements with additional points [28]. Then, to maintain fairness throughout the experiments, the same number of iterations is adopted for the proposed method. The experimental settings for active learning are also depicted in Table 2. For each dataset, the table shows the seed, the initial size of X_{train} , and the size of the test set D_{test} used to evaluate the classifier.

4.3. Experiments results

The following subsections present the results of comparing ArtAI with other methods.

4.3.1. Comparison with baseline methods. Table 3 shows the experimental results of comparing the proposed method with a set of predictive models for mobile social popularity. To avoid measurement bias, the evaluation reports several performance metrics, which include Precision (P) and Recall (R), and F1 measure (F1). Additionally, to report prediction accuracy, the experiments consider the Matthews correlation coefficient (MCC) [29] to describe the

Table 2. Experimental settings.

Dataset	# Labeling Functions	Labeling Functions Performance				Active Learning Settings		
		Accuracy	Precision	Recall	F1	Seed	X_{train}	D_{test}
Online News	6	0.74	0.82	0.78	0.80	1,989	24,675	13,133
Reddit Engagement	7	0.83	0.68	0.72	0.70	4,287	58,054	26,973
Webhose News	9	0.66	0.71	0.77	0.74	8,544	111,073	51,265

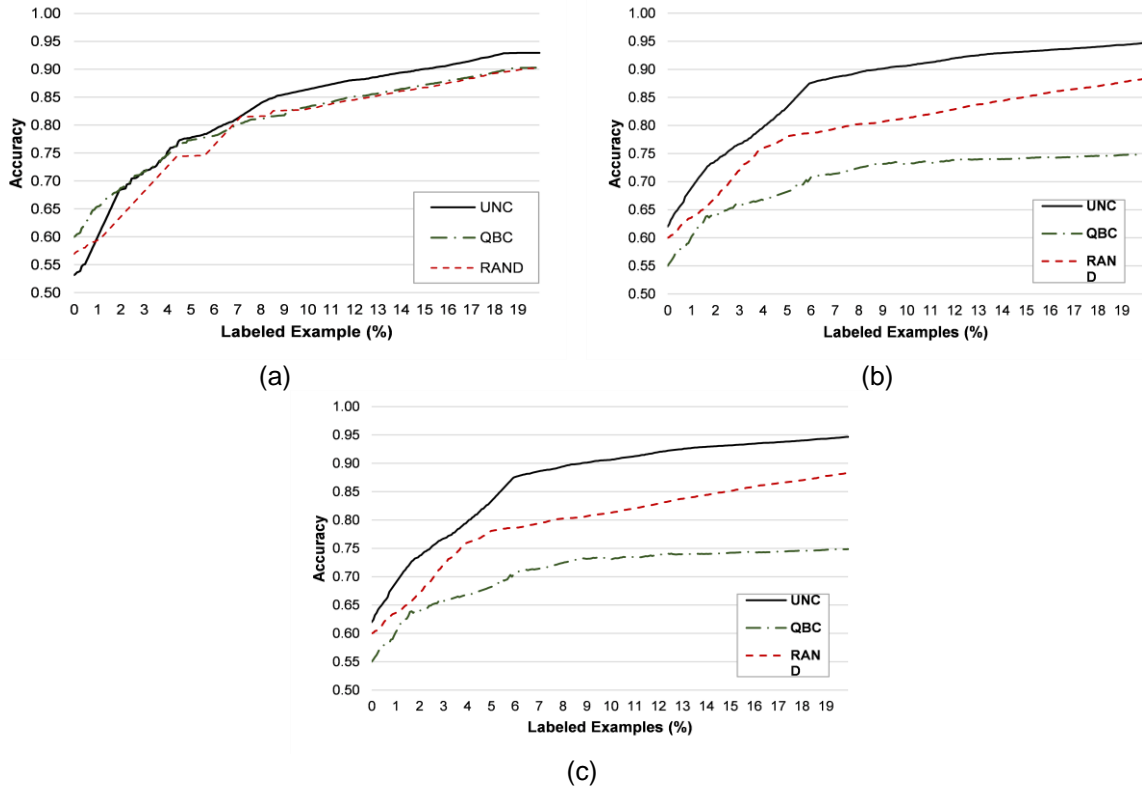


Figure 3. Learning curves of active learning for (a) Online news dataset (b) Reddit Engagement dataset (d) Webhose News dataset

confusion matrix instead of accuracy since classification accuracy can be misleading with imbalanced datasets.

The experimental results show that ArtAI attains higher precision and recall in all the problems. It managed to achieve better results than the comparing tools as it applies a meta-active learning process to enhance the accuracy of the generated popularity estimates. On average, the proposed method improves precision by up to 28.17% compared to the other methods in the Webhose News dataset.

As for the prediction accuracy, ArtAI also achieves higher MCC values in all the tasks. When compared to GBM, VSM, and ensemble models, the proposed scheme could improve the accuracy of the generated

predictions by 3.75%, 5.79%, and 3.90%, on average, respectively.

Takeaway: The proposed method maintains a comparative performance for mobile news popularity when compared to state-of-the-art techniques.

4.3.2. Comparison with active learning. To compare the labeling budget of ArtAI to traditional active learning process, we applied three query strategies to the three datasets, namely UNC, QBC, and RAND. The learning curves of the three query strategies are shown in Figure 3. The learning curves demonstrate the relationship between accuracy and the number of labeled articles consumed to achieve the corresponding accuracy value.

Table 3. Experimental results of comparison with baseline techniques.

Model	Online News				Reddit Engagement				Webhose News			
	P	R	MCC	F1	P	R	MCC	F1	P	R	MCC	F1
Proposed Method	0.88	0.97	0.96	0.92	0.93	0.88	0.92	0.90	0.91	0.95	0.85	0.93
GBM	0.81	0.84	0.83	0.82	0.83	0.82	0.84	0.82	0.89	0.83	0.81	0.86
VSM	0.86	0.89	0.91	0.87	0.91	0.80	0.88	0.85	0.71	0.91	0.72	0.80
Ensemble	0.74	0.92	0.84	0.82	0.83	0.79	0.85	0.81	0.86	0.82	0.71	0.84

Table 4. Experimental results of comparison with active learning (uncertainty sampling).

Dataset	Proposed Method					Active Learning (UNC)				
	P	R	MCC	Accuracy	# queried examples	P	R	MCC	Accuracy	# queried examples
Online News	0.93	0.95	0.85	0.92	5,374	0.89	0.9	0.8	0.9	7,764
Reddit Engagement	0.95	0.91	0.91	0.93	13,613	0.91	0.89	0.81	0.93	21,638
Webhose News	0.81	0.92	0.86	0.95	34,381	0.79	0.74	0.62	0.94	48,298

Since the curves show that UNC attained the highest efficiency for the three datasets, the experiments report the performance metrics achieved by ArtAI and UNC in Table 4. The table also shows the number of labeled articles needed to obtain the reported accuracy values. The results illustrate that the proposed method achieved better MCC values than UNC in the three datasets with an overage improvement of 19.10%. The table also shows that the proposed method maintains less labeling budget than traditional active learning, which proves that the learned selection policy in the proposed method managed to reduce the cost of manual labeling. Since the active learning process in the proposed approach starts with a much smaller unlabeled pool, the budget for manual labeling is highly reduced. For example, in the mobile social news dataset, traditional active learning needed to label 31.47% of the training pool, while the size of the unlabeled pool in the proposed method only represents 21.78% of the training set size, which results in 30.78% decrease in labeling cost. On average, ARTAI, could reduce the labeling budget in the three datasets by 32.2% when compared to UNC.

Moreover, the proposed method achieved better precision and recall values than traditional active learning in the three datasets. For Webhose News, the proposed method surpassed the recall values of active learning by 24.32%. Similarly, it improved the precision value in the same dataset by 2.53%.

Takeaway: The results empirically demonstrate that the models generated by the proposed method achieve remarkable results in real-world situations in popularity predictions for social news.

4.3.3. Sensitivity analysis of the experimental parameters. As mentioned earlier, the experiments terminated the traditional active learning process when the improvements of classification accuracy do not exceed a threshold $\lambda=0.0001$ for a successive number of iterations [28]. Then, the experiments utilized the same number of labeled articles consumed by traditional active learning as the labeling cost for the proposed method. Thus, to test the sensitivity of the stopping criterion, the experiments are repeated with different values of λ . First, the experiments with traditional active learning are repeated with values of $\lambda = 0.001, 0.0001, 0.00001$. Figure 4.a shows the accuracy values achieved by the underline classifiers with UNC using the three datasets.

Likewise, the labeling budget of the proposed method is customized according to the number of annotations consumed by UNC in each dataset. Table 5 shows the size of the unlabeled pool (X_{train}), the annotation budget used by UNC as a percentage of the total size of the unlabeled pool (Labeling cost%), and the size of the unlabeled pool in the proposed method P_H as a percentage of X_{train} . As the table shows, the size

Table 5. Experimental results with different values of λ .

Dataset	λ	Active Learning		WeSAL	
		Size of X_{train}	AL Cost %	P_H %	$B_{Labeling}$
News	0.001	24,675	12.20%	18.22%	3,010
	0.0001		31.47%		7,764
	0.00001		37.60%		9,278
Reddit Engagement	0.001	58,054	6.20%	17.19%	3,599
	0.0001		14.00%		8,128
	0.00001		16.60%		9,637
Webhose News	0.001	111,073	8.01%	21,81%	8,886
	0.0001		19.11%		20,660
	0.00001		24.91%		26,658

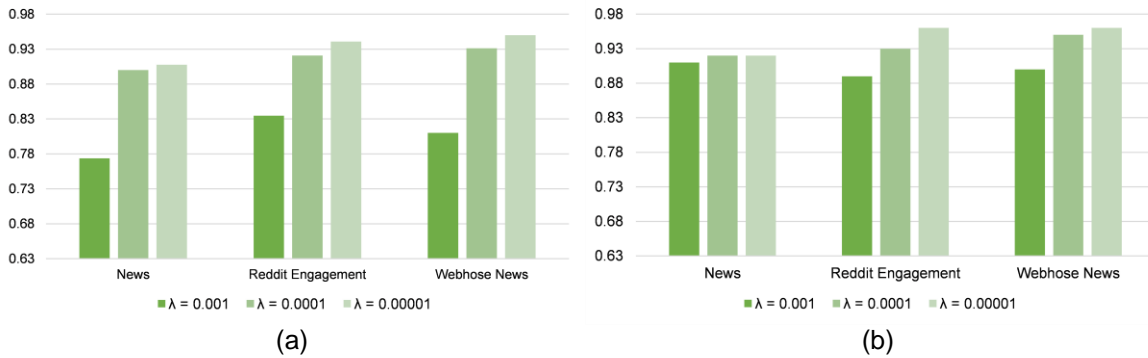


Figure 4. Accuracy values for (a) active learning (UNC) (b) the proposed method values of $\lambda = 0.001$, 0.0001 , 0.00001

of P_H is much smaller than the X_{train} in all the datasets. Also, the accuracy values achieved by the proposed method are reported in Figure 4.b.

As figure 4 shows, when the value of λ increases, this can terminate active learning too early, which results in missing useful generalizations [28]. For example, setting $\lambda = 0.001$ decreased the accuracy of UNC in mobile social News dataset by 14.06 % when compared to the accuracy with $\lambda = 0.0001$ (Figure 4.a).

Moreover, the results also attest that the labeling budget tends to increase when λ is set to a small value ($\lambda=0.00001$). However, the additional cost of manual labeling does not result in a significant enhancement in classification performance. For example, with $\lambda=0.00001$, UNC increased its labeling budget in the Webhose News by 29.03%, but with only 2.03% enactment achieved in accuracy values when compared to the performance achieved with $\lambda = 0.0001$.

Additionally, the results show that the proposed method maintained better results than active learning with different values of λ . Since the size of the unlabeled pool P_H is much less than the size of X_{train} , in some cases, the total size of P_H is less than the number of annotations consumed by active learning. Therefore, changing the value of λ did not affect the performance of the proposed method.

Takeaway: ARTAI achieves better performance than AL with all variations of λ in all the datasets.

5. Conclusions

In this article, a new scheme is presented for popularity estimates of news content. Given the abundant nature of modern mobile technology, mobile social news is now consumed essentially everywhere. The proposed approach does not entail labeled training examples to produce popularity predictions. As an

alternative, it utilizes weakly supervised labels from user-defined heuristics to generate initial estimates. Then, it applies a novel selection policy to repair these weak predictions. Since the proposed scheme engages end-users in the rectification process, this human – mobile interaction is predicated on getting the correct news article in front of the consumer at the right time. The experimental evaluation includes three real-world datasets and shows that the proposed method can outperform state-of-the-art techniques by up to 19.72% in classification performance. Also, the results empirically confirm that the proposed approach could attain better results than traditional active learning while reducing the annotation budget by up to 37.09%.

As future directions, we plan to investigate the effect of different popularity metrics on the proposed scheme's performance. For example, the absolute number of comments may not be an accurate measure for popularity since some of these comments may contain skepticism and critique of the topic. Instead, a preliminary phase of sentiment analysis may help detect the real number of positive comments.

6. References

- [1] Q. Ye, Y. Luo, G. Chen, X. Guo, Q. Wei, and S. Tan, "Users Intention for Continuous Usage of Mobile News Apps: the Roles of Quality, Switching Costs, and Personalization," *Journal of Systems Science and Systems Engineering*, vol. 28, no. 1, pp. 91–109, Feb. 2019.
- [2] J. Rezaeenour, M. Y. Eili, E. Hadavandi, and M. H. Roozbahani, "Developing a New Hybrid Intelligent Approach for Prediction Online News Popularity," *International Journal of Information Science and Management (IJISM)*, vol. 16, no. 1, Feb. 2018.
- [3] A. Tatar, J. Leguay, P. Antoniadis, A. Limbourg, M. D. de Amorim, and S. Fdida, "Predicting the Popularity of Online Articles Based on User Comments," in *Proceedings*

of the *International Conference on Web Intelligence, Mining and Semantics*, New York, NY, USA, 2011.

[4] S. Abbar, C. Castillo, and A. Sanfilippo, "To Post or Not to Post: Using Online Trends to Predict Popularity of Offline Content," in *Proceedings of the 29th on Hypertext and Social Media*, New York, NY, USA, 2018, pp. 215–219.

[5] Yu-Jen Lin, Mi-Yen Yeh, Fang-Yi Chiu, Ya-Hui Chan, and Chia-Chi Wu, "Predicting popularity of articles on bulletin board system," in *2016 International Conference on Big Data and Smart Computing*, Jan. 2016, pp. 169–176.

[6] R. Bandari, S. Asur, and B. A. Huberman, "The Pulse of News in Social Media: Forecasting Popularity," in *AAAI Conference on Weblogs and Social Media*, May 2012.

[7] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini, "A Peek into the Future: Predicting the Evolution of Popularity in User Generated Content," in *Proc. of the 6th ACM Conf. on Web Search and Data Mining*, USA, 2013, pp. 607–616.

[8] D. Deshpande, "Prediction Evaluation of Online News Popularity Using Machine Intelligence," in *International Conf. on Computing, Communication, Control and Automation*, Aug. 2017, pp. 1–6.

[9] R. Shreyas, D. M. Akshata, B. S. Mahanand, B. Shagun, and C. M. Abhishek, "Predicting popularity of online articles using Random Forest regression," in *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*, Aug. 2016, pp. 1–5.

[10] A. Ratner, S. H. Bach, H. Ehrenberg, J. Fries, S. Wu, and C. Ré, "Snorkel: rapid training data creation with weak supervision," *Proc. of VLDB*, vol. 11, pp. 269–282, 2017.

[11] M. Nashaat, A. Ghosh, J. Miller, and S. Quader, "WeSAL: Applying Active Supervision to Find High-quality Labels at Industrial Scale". *Proceedings of the 53rd Hawaii Conference on System Sciences* pp. 219-228.

[12] P. Varma, D. Iter, C. De Sa, and C. Ré, "Flipper: A Systematic Approach to Debugging Training Sets," in *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics - HILDA'17*, Chicago, IL, USA, 2017.

[13] Y. Li, Q. Peng, Z. Sun, L. Fu, and S. Khokhar, "A Two-stage Prediction Method of News Popularity only using Content Features," in *2018 13th World Congress on Intelligent Control and Automation*, Jul. 2018, pp. 767–772.

[14] M. T. Uddin, M. J. A. Patwary, T. Ahsan, and M. S. Alam, "Predicting the popularity of online news from content metadata," in *International Conference on Innovations in Sci., Eng. and Technology*, 2016, pp. 1–5.

[15] Y. Fu, X. Zhu, and B. Li, "A survey on instance selection for active learning," *Knowledge and Information Systems*, vol. 35, no. 2, pp. 249–283, 2013.

[16] R. G. Garroppo, M. Ahmed, S. Niccolini, and M. Dusi, "A Vocabulary for Growth: Topic Modeling of Content Popularity Evolution," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2683–2692, Oct. 2018.

[17] Y. Liu, T. Zhi, H. Xi, X. Duan, and H. Zhang, "A Novel Content Popularity Prediction Algorithm Based on Auto Regressive Model in Information-Centric IoT," *IEEE Access*, vol. 7, pp. 27555–27564, 2019.

[18] Z. Bao, Y. Liu, Z. Zhang, H. Liu, and J. Cheng, "Predicting popularity via a generative model with adaptive peeking window," *Physica A: Statistical Mechanics and its Applications*, vol. 522, pp. 54–68, May 2019.

[19] S. Das Bhattacharjee, A. Talukder, and B. V. Balantrapu, "Active learning based news veracity detection with feature weighting and deep-shallow fusion," in *2017 IEEE International Conference on Big Data (Big Data)*, Boston, MA, Dec. 2017, pp. 556–565.

[20] S. D. Bhattacharjee, W. J. Tolone, and V. S. Paranjape, "Identifying malicious social media contents using multi-view Context-Aware active learning," *Future Generation Computer Systems*, vol. 100, pp. 365–379, 2019.

[21] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Explainable Machine Learning for Fake News Detection," in *Proceedings of the 10th ACM Conference on Web Science*, New York, NY, USA, 2019, pp. 17–26.

[22] H. Zamani and W. B. Croft, "On the theory of weak supervision for information retrieval," in *Proceedings of the ACM SIGIR Conference on Theory of Info. Retrieval*, 2018.

[23] Y. Meng, J. Shen, C. Zhang, and J. Han, "Weakly-Supervised Neural Text Classification," in *Proc. of ACM Conf. on Info. and Knowledge Mngmt.*, 2018, pp. 983–992.

[24] K. Shu, S. Wang, D. Lee, and H. Liu, "Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements," *arXiv:2001.00623 [cs]*, Jan. 2020.

[25] A. J. Ratner, C. M. De Sa, S. Wu, D. Selsam, and C. Ré, "Data Programming: Creating Large Training Sets, Quickly," in *NIPS*, 2016, pp. 3567–3575.

[26] S. H. Bach, B. He, A. Ratner, and C. Ré, "Learning the Structure of Generative Models without Labeled Data," *arXiv:1703.00854 [cs, stat]*, Mar. 2017.

[27] P. Varma and C. Ré, "Snuba: automating weak supervision to label training data," *Proceedings of the VLDB Endowment*, vol. 12, no. 3, pp. 223–236, 2018.

[28] M. Bloodgood and K. Vijay-Shanker, "A method for stopping active learning based on stabilizing predictions and the need for user-adjustable stopping," in *Proc. of Conference on Comput. Nat. Lang. Learn.*, 2009, pp. 39–47.

[29] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2011.