

Locally Adaptive Density Ratio for Detecting Novelty in Twitter Streams

Yun-Qian Miao
University of Waterloo
Waterloo, Ontario, Canada.
N2L 3G1
yqmiao@uwaterloo.ca

Ahmed K. Farahat
University of Waterloo
Waterloo, Ontario, Canada.
N2L 3G1
afarahat@uwaterloo.ca

Mohamed S. Kamel
University of Waterloo
Waterloo, Ontario, Canada.
N2L 3G1
mkamel@uwaterloo.ca

ABSTRACT

With the massive growth of social data, a huge attention has been given to the task of detecting key topics in the Twitter stream. In this paper, we propose the use of novelty detection techniques for identifying both emerging and evolving topics in new tweets. In specific, we propose a locally adaptive approach for density-ratio estimation in which the density ratio between new and reference data is used to capture evolving novelties, and at the same time a locally adaptive kernel is employed into the density-ratio objective function to capture emerging novelties based on the local neighborhood structure. In order to address the challenges associated with short text, we adopt an efficient approach for calculating semantic kernels with the proposed density-ratio method. A comparison to different methods shows the superiority of the proposed algorithm.

Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning

Keywords

Novelty Detection; Density Ratio Estimation; Locally Adaptive Kernel; Semantic Kernel Representation; Social Media Analysis.

1. INTRODUCTION

With the increasing popularity of social networks, a huge amount of diverse and dynamic information is continually being generated. Mining this rich information and analyzing the trend of topics in social media has the potential to be useful in many aspects, such as for helping political parties and companies to understand people's opinions, responding to customer needs, or even discovering natural or social disasters as early as possible [19]. Accurately detecting novel content from this short text in a timely fashion is an important task, which involves identifying novel instances that

include new topics or have sudden increases of intensity on existing topics in comparison to past.

A common approach to novelty detection is to first build a model from the past data, and then recognize any deviation from that model as novel. Different existing methods for novelty detection capture different aspects of novel. For instance, some methods, such as one-class support vector machine [22] and local outlier factor [3], focus on detecting novel instances that emerge in new data which are completely different from the previously-seen instances. Other methods, such as relative [23] and density-ratio-based [12] novelty detection, are effective in detecting instances which are not new by themselves, but their intensities are considerably different from the previously-seen data. These two types of novelties have been referred as emerging and evolving in [21], and the authors proposed a dynamic non-negative matrix factorization approach to identify emerging topics from evolving topics.

Due to the time continuity of social media streams, these two types of novelties are not easily distinguishable and a novel concept is usually characterized by the combination of emerging and evolving. One reason is the existence of large common vocabularies between different topics. Another reason is that there is high possibility of topics being continuously discussed in sequential batch of collections, but showing different level of intensity. While the previous approaches for novelty detection exhibit appealing successes in specific applications, methods that focus on detecting emerging novelties are quite limited in identifying evolving novelties, and vice versa.

To address the limitation, a locally adaptive density ratio approach is presented in this paper that combines the strength of both categories of methods and recognizes both emerging and evolving novelties. The basic idea behind the proposed approach is to use the density-ratio that provides a measure of how evolving the novelty is, while the structural similarity learned from local neighborhoods captures how emerging it is. Moreover, to deal with the challenges associated with the very short text of social media data, we propose the use of an efficient representation model by calculating semantic kernels with the proposed density-ratio method. The semantic kernels are constructed with the low-rank approximation of statistical term-term correlations, which greatly alleviate the dimensionality and sparsity problem while reserving the most important bases of semantic meanings.

The rest of this paper is organized as follows. Section 2 introduces the problem formulation and briefly reviews related

work. Section 3 describes the proposed locally adaptive density ratio approach. Section 4 presents experimental results. Lastly, the paper is concluded in Section 5.

2. RELATED WORK

Given samples from historical records as the reference collection \mathcal{X}_{rf} and a batch of new coming instances as the test collection \mathcal{X}_{ts} , the novelty detection is a task to identify test instances which exhibit any deviation from the reference as novel. This is an assumption that has been adopted in many well-recognized work on the novelty detection task [3, 12, 15, 22, 23].

In this setting, the model of normality $M(x)$ is learned from the given normal examples \mathcal{X}_{rf} . Then, in the test stage, previously-unseen instances \mathcal{X}_{ts} are tested against the model M , and the corresponding novelty scores are calculated. The novelty score for a test sample x , i.e. $m = M(x)$, is compared to a decision threshold τ . If $m \leq \tau$, this instance x is classified as normal. Otherwise, it is classified as novel.

A variety of approaches for novelty detection have been proposed with different inspirations. The probabilistic-based novelty detection method estimates the Probability Density Function (PDF) of reference data, and assumes that low density areas have a high probability of being novel [9]. The One-class Support Vector Machine (OSVM) method models the boundary of reference data and assumes that samples locating outside of the boundary are novel [22]. The neighborhood-based approach analyzes the distances of k -nearest neighbors, and identifies novel instances if they are relatively far from their neighbors in the reference collection [3]. In [23], Smola et al. proposed a concept of relative novelty and modified the OSVM to incorporate the reference densities as density ratios.

Following, Hido et al. [12] proposed an inlier-based outlier detection method and defined the inlier score by using density ratios between the reference data and the test data. For the regions that the inlier scores are small, it means the reference data density is low and the test data density is high. Thus, with the relative densities, the novel instance can be identified if its inlier score is below a threshold. Recently a number of methods have been proposed to estimate the Density Ratio (DR) [16, 17]. With different optimization formulation, several well-known methods have been proposed as the Kernel Mean Matching (KMM) algorithm [13], the Kullback-Leibler Importance Estimation Procedure (KLIEP) algorithm and the unconstrained Least-Squares Importance Fitting (uLSIF) algorithm [11].

In the topic detection field, there are many techniques being proposed, which include latent semantic indexing, Non-negative Matrix Factorization (NMF), and different clustering methods [1, 2, 24]. In [14], Karkali et al. proposed a new novelty score by modifying the Inverse Document Frequency (IDF) scoring function. In [10], Guille and Favre proposed the use of social links to evaluate tweets impact and enhance the abnormal social events detection. One closely related work is from Saha and Sindhwani [21], which introduced the concepts of emerging and evolving and proposed a dynamic NMF framework to detect them.

To detect novel content in twitter streams, the goal of this paper is to explore the use of novelty detection techniques in the topic identification task, which is a gap in the literature. The following section describes the details of our proposed approach.

3. PROPOSED APPROACH

In the literature of novelty detection, existing techniques for novelty detection focus on either 1) detecting novel instances that are completely different from existing instances in the reference data, or 2) identifying a group of novel instances that are not completely new in themselves, but rather appear with a density which is different from that of similar instances in the reference data. As highlighted in the introduction, we refer these two types of novelties as emerging and evolving, respectively. The objective of this work is to identify the limitations of each of these methods and introduce a method that is capable of identifying both emerging and evolving novelties.

The basic idea behind the proposed method is to start with the effective kernel mean matching method for density-ratio estimation, which has been very successful in identifying evolving novelties [12], and to augment this method to capture emerging novelties. We observed that the density ratio as novelty measure is very unreliable when encountering a few novelties in completely new areas of the feature space. In this case, the density-ratio estimation depends mainly on the absolute dissimilarity between the potentially novel instances and other reference instances, without considering information about how these reference instances are similar to each other. We propose to alleviate this problem by adaptively estimating the density ratio based on the neighborhood of both the new instance as well as the reference instances.

In the rest of this section, we first introduce the semantic representation model which is used to deal with the sparsity problem associated with the very short text of social media data. Then, we formalize the aforementioned intuition by analyzing the kernel mean matching algorithm for density ratio estimation and identify the different components of the novelty score that quantify how emerging/evolving the novelty is. Following, we analyze why the component responsible for detecting emerging novelties is inaccurate. After that, we propose the use of locally adaptive kernels that results in accurate estimations.

3.1 Semantic Representation Model

The conventional text representation is the Vector Space Model (VSM), which represents documents with a document-term matrix. The most problematic disadvantages of VSM representation is its sparsity and high dimensionality. This becomes severe when dealing with the short social media data. For example, the maximum length of twitter messages is only 140 characters, with the average length being approximately 15 words per tweet [20]. In English texts, the dimension can easily reach 10K even the techniques of stop word removing and stemming are employed.

To deal with the extreme sparsity and high dimensionality problem in the short social media data, the application of low-rank semantic kernels is proposed. It first builds semantic kernels based on term-term correlations [7]. Then, a Nyström approximation [6] is applied to extract low-rank representations. In our following experiments, the low rank number is set to 500. Through these two steps, the original document-term space is transformed into a dense semantic-enhanced space. The dimensionality and sparsity are greatly alleviated while reserving the most important bases of semantic meanings in the data.

3.2 Novelty Measure by Kernel Mean Matching

The kernel mean matching method for novelty detection proceeds as follows. Let $r(x)$ be a normality score for a data point x , defined as the density ratio between the reference and test data:

$$r(x) = \frac{p_{\text{rf}}(x)}{p_{\text{ts}}(x)}, \quad (1)$$

where $p_{\text{rf}}(x)$ and $p_{\text{ts}}(x)$ are the PDFs of reference and test data, respectively. For a data point x , if the test density is relatively higher than the reference density, then the normality score will be very small, and this point is more likely to represent a novelty.

To derive a formula for $r(x)$, we use the theorem of Kernel Mean Matching (KMM) [13] and minimize the Maximum Mean Discrepancy (MMD) between the weighted distribution $r(x)p_{\text{ts}}(x)$ and the reference distribution $p_{\text{rf}}(x)$ in a Reproducing Kernel Hilbert Space (RKHS) $\phi(x) : x \rightarrow \mathcal{F}$,

$$\text{MMD}^2(\mathcal{F}, (r(x), p_{\text{ts}}(x)), p_{\text{rf}}(x)) = \left\| E_{x \sim p_{\text{ts}}(x)} [r(x) \cdot \phi(x)] - E_{x \sim p_{\text{rf}}(x)} [\phi(x)] \right\|^2. \quad (2)$$

Using the empirical means of \mathcal{X}_{rf} and \mathcal{X}_{ts} to replace the expectations, and defining a vector $\mathbf{r} = [r(x_1), \dots, r(x_{n_{\text{ts}}})]^T$, we can obtain the following quadratic optimization problem:

$$\begin{aligned} \hat{\mathbf{r}} &= \operatorname{argmin}_{\mathbf{r}} \left\| \frac{1}{n_{\text{ts}}} \sum_{j=1}^{n_{\text{ts}}} r(x_j) \phi(x_j) - \frac{1}{n_{\text{rf}}} \sum_{i=1}^{n_{\text{rf}}} \phi(x_i) \right\|^2 \\ &= \operatorname{argmin}_{\mathbf{r}} \left[\frac{1}{n_{\text{ts}}} \sum_{i,j=1}^{n_{\text{ts}}} r(x_i) \kappa(x_i, x_j) r(x_j) \right. \\ &\quad \left. - \frac{2}{n_{\text{ts}} n_{\text{rf}}} \sum_{i=1}^{n_{\text{ts}}} \sum_{j=1}^{n_{\text{rf}}} r(x_i) \kappa(x_i, x_j) + \frac{1}{n_{\text{rf}}^2} \sum_{i,j=1}^{n_{\text{rf}}} \kappa(x_i, x_j) \right] \\ &= \operatorname{argmin}_{\mathbf{r}} \left[\frac{1}{2} \mathbf{r}^T K_{x_{\text{ts}}, x_{\text{ts}}} \mathbf{r} - \frac{n_{\text{ts}}}{n_{\text{rf}}} \mathbf{r}^T K_{x_{\text{ts}}, x_{\text{rf}}} \mathbf{1} \right], \quad (3) \end{aligned}$$

where

$$\begin{aligned} K_{x_{\text{ts}}, x_{\text{ts}}}(i, j) &= \kappa(x_i, x_j), \{x_i, x_j \in \mathcal{X}_{\text{ts}}\}, \\ K_{x_{\text{ts}}, x_{\text{rf}}}(i, j) &= \kappa(x_i, x_j), \{x_i \in \mathcal{X}_{\text{ts}}, x_j \in \mathcal{X}_{\text{rf}}\}, \\ \mathbf{1} &= [1, \dots, 1]^T. \end{aligned} \quad (4)$$

The optimal solution of Eq. 3 without imposing constraints on \mathbf{r} can be analytically obtained as:

$$\hat{\mathbf{r}} = \frac{n_{\text{ts}}}{n_{\text{rf}}} K_{x_{\text{ts}}, x_{\text{ts}}}^{-1} K_{x_{\text{ts}}, x_{\text{rf}}} \mathbf{1}.$$

For a test point $x \in \mathcal{X}_{\text{ts}}$,

$$\hat{r}(x) = \frac{n_{\text{ts}}}{n_{\text{rf}}} \sum_{x_i \in \mathcal{X}_{\text{rf}}} \kappa(x, x_i) - \sum_{x_j \in \mathcal{X}_{\text{ts}} \setminus x} r(x_j) \kappa(x, x_j). \quad (5)$$

This means the normality index $r(x)$ is the difference between terms r_1 and r_2 which are defined as

$$\begin{aligned} r_1(x) &= \frac{n_{\text{ts}}}{n_{\text{rf}}} \sum_{x_i \in \mathcal{X}_{\text{rf}}} \kappa(x, x_i), \\ r_2(x) &= \sum_{x_j \in \mathcal{X}_{\text{ts}} \setminus x} r(x_j) \kappa(x, x_j). \end{aligned}$$

These two terms affect the novelty of x as follows: The first term r_1 captures the similarity between the test instance x and all reference instances \mathcal{X}_{rf} (based on $\kappa(x, x_i)$).

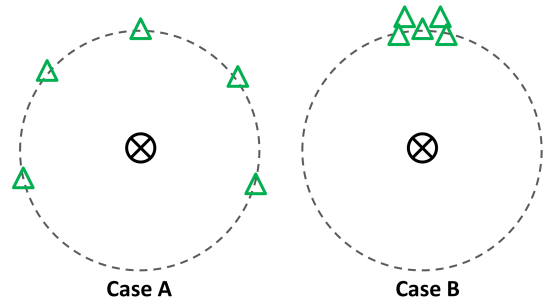


Figure 1: Two cases with identical r_1 .

This quantifies how x is different from the previously-seen instances and, accordingly, measures how emerging it is. If x is very dissimilar to all reference instances, then x is an emerging novelty and the value of r_1 will be very small. On the other hand, if x is very similar to many reference instances, then x is a normal instance and the value of r_1 will be very large.

The second term r_2 captures the similarity between the test instance x and other test instances \mathcal{X}_{ts} . This quantifies how x is a novelty relative to other new instances, depending on how similar these instances are (based on $\kappa(x, x_j)$) and how novel they are (based on $r(x_j)$). This term is a key indicator for detecting evolving concepts. To understand how, let us consider the case where a test instance x appears very close to reference data \mathcal{X}_{rf} . In this case, using r_1 only results in the conclusion that this point is not novel. However, if x appears within a tight cluster of other test instances, the large score of r_1 will propagate through the calculation of r for the instances of this cluster, and result in a very large value for r_2 . Accordingly, r_2 will reduce the overall score $r(x)$ and lead to the conclusion that x is an evolving novel.

3.3 Limitations of Density Ratio Measure

Although $r(x)$ captures the emerging and evolving aspects of novelty, the score of emerging novelty r_1 is very inaccurate, as it mainly depends on the absolute similarity between x and other data instances. For instance, supposing that x has an average similarity of s to all the reference instances, we cannot conclude anything about how novel x is unless we learn about how similar reference instances are to each other. If s is a common similarity in the subspace of reference instances, then x should be considered normal regardless of the absolute value of s .

To illustrate this argument, Fig. 1 shows two cases for a test instance x (black cross) and a set of reference instances (green triangles). For the two cases, the value of r_1 is exactly the same. However, comparing with their neighboring structures, x should be considered normal in Case A and novel in Case B. A similar argument was discussed by Breunig et al. [3].

3.4 Locally Adaptive Kernel

In order to address the aforementioned limitations, we need to incorporate information about the similarity of x and y to other neighboring instances in the calculation of how novel is x with respect to y . Since the novelty score is mainly based on the kernel function between x and y , one indirect way to modify the novelty score is to adjust the kernel function to

reflect how x is truly dissimilar (i.e., novel) to y with respect to their neighborhood. For instance, in the case of Gaussian kernels, we can adjust the kernel width σ to be adaptive to the local neighborhood of each pair of instances. In the rest of this section, we will focus developing a locally adaptive Gaussian kernel based on this idea. The same idea can be directly extended to other types of kernels by normalizing the value of the kernel function $\kappa(x, y)$ using the kernel between x and y and other neighboring points.

The Gaussian kernel function between two samples x_i and x_j is defined as

$$\kappa_\sigma(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2\sigma^2}\right), \quad (6)$$

where $d(x_i, x_j)$ is a distance function between the two samples x_i and x_j , and σ is the scaling factor (bandwidth) that decides the smoothness of the kernel. Instead of choosing one scaling factor σ for measuring the similarity between all data points in the kernel space, we propose the use of a locally adaptive kernel that captures the local density statistics of x_i and x_j . One locally adaptive kernel which was successfully used for enhancing spectral clustering [25] is defined as:

$$\kappa_l(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2d(x_i, N_k(x_i))d(x_j, N_k(x_j))}\right), \quad (7)$$

where $N_k(x_i)$ is the k -th nearest neighbor of x_i . In other words, the bandwidth of $\kappa_l(x_i, x_j)$ is the geometric mean of the k -th nearest neighbor distances for x_i and x_j :

$$\sigma_{ij} = \sqrt{d(x_i, N_k(x_i))d(x_j, N_k(x_j))}.$$

Using this locally adaptive kernel to calculate $K_{x_{ts}, x_{ts}}$ and $K_{x_{ts}, x_{rf}}$ of Eq. 3, the new normality score takes into consideration two factors: relativity to normality and relativity to neighborhood.

3.5 Approximate with Diagonal Shifting

According to [18], a valid reproducing kernel should satisfy the following two properties: 1) *Hermitian*. For finite data observations of real entries, the Gram matrix should be symmetric ($K_{ij} = K_{ji}$); and 2) *Positive Semi-Definite (PSD)*. For finite data observations, the Gram matrix should be positive semi-definite.

As seen from Eq. 7, the local kernel satisfies the symmetric condition, i.e. $\kappa_l(x_i, x_j) = \kappa_l(x_j, x_i)$. But, because the matrix $K_{x_{ts}, x_{ts}}$ is constructed from a locally adaptive kernel (Eq. 7), the Positive Semi-Definite (PSD) of the kernel might be violated. In order to address this issue and enforce PSD, we adopt the following approximation using diagonal shifting [5]:

$$\tilde{K}_{x_{ts}, x_{ts}} = K_{x_{ts}, x_{ts}} + (\delta + \epsilon) I, \quad (8)$$

where δ is the absolute value of the minimum negative eigenvalue, and ϵ is a small value for compensating numerical error. Through diagonal shifting, the only changes to the kernel matrix are elements representing self-similarity, while all other pairs of similarities are not affected. Therefore, our intuitive idea of expressing similarity based on neighborhood structure is still preserved in $\tilde{K}_{x_{ts}, x_{ts}}$; meanwhile, the PSD and symmetry are both satisfied.

Thus, after applying diagonal shifting to $K_{x_{ts}, x_{ts}}$, the optimization problem of Eq. 3 becomes a convex quadratic problem. Our implementation includes a boundary constraint on

Algorithm 1 Locally Adaptive Kernel Mean Matching

Input: $\mathcal{X}_{rf}, \mathcal{X}_{ts}, k, \epsilon, b$

Output: $\mathbf{r} = [r(x_1), \dots, r(x_{n_{ts}})]^T$

Steps:

- 1: $\sigma_{ij} = \sqrt{d(x_i, N_k(x_i))d(x_j, N_k(x_j))}, \forall x_i \in \mathcal{X};$
 - 2: $K_{x_{ts}, x_{rf}}(i, j) = \kappa_l(x_i, x_j), \forall x_i \in \mathcal{X}_{ts}, x_j \in \mathcal{X}_{rf}$
 - 3: $K_{x_{ts}, x_{ts}}(i, j) = \kappa_l(x_i, x_j), \forall x_i, x_j \in \mathcal{X}_{ts}$
 - 4: **if not** PSD($K_{x_{ts}, x_{ts}}$) **then**
 - 5: $\tilde{K}_{x_{ts}, x_{ts}} = K_{x_{ts}, x_{ts}} + (\delta + \epsilon) I$ (Eq. 8);
 - 6: **else**
 - 7: $\tilde{K}_{x_{ts}, x_{ts}} = K_{x_{ts}, x_{ts}}$
 - 8: **end if**
 - 9: $\mathbf{r} \leftarrow \text{QP_solver}(\tilde{K}_{x_{ts}, x_{ts}}, K_{x_{ts}, x_{rf}}, \epsilon, b)$.
-

Table 1: The tweet dataset and test scenario.

Scenario		Topic	n_{rf}	n_{ts}
S1	Stable	fashion	1000	1000
	Emerging	investing	0	1000
	Evolving	media	100~200	1000
S2	Stable	music, religion, shopping	3000	3000
	Emerging	sports, technology	0	2000
	Evolving	travel, video-games	200~400	2000

r and uses the well-known ‘interior-point-convex’ algorithm in the Matlab toolbox as the Quadratic Programming (QP) solver. The complete locally adaptive Kernel Mean Matching (*locKMM*) algorithm is outlined in Algorithm 1.

4. EXPERIMENTS

4.1 Dataset and Experimental Setup

The tweet dataset used in this experiment is described in [26], which includes 369K tweets spreading over 10 topics. The tweets were labeled using the content of the page which the URL refers to, and the categories of the Open Directory Project (ODP) as their labels. As listed in Table 1, two test scenarios are formulated. In each scenario, the emerging novelty is the category of tweets that appears in the test collection only. The evolving novelty is the category that is rare in the reference collection but shows high frequency in the test collection. The stable content are tweets that keep the same level of intensity in both the reference and test collection.

Comparison methods.

Besides the proposed method, state-of-the-art novelty detection algorithms dynamic NMF, OSVM, LOF, uLSIF and KMM are included as comparison. Their implementation details are set as follows.

- **Dynamic NMF:** Dynamic Non-negative Matrix Factorization [21]. This method is used for detecting novel topics by expanding the topic bases¹. The bandwidth of new topics follows the original paper, which is set as 4.
- **OSVM:** One-class Support Vector Machine [22]. The LibSVM library [4] is used with the parameter ν set to 0.1. The Gaussian kernel is adopted, the kernel width is set to the median distance between samples.

¹We implemented the dynamic NMF approach according to the details provided in [21].

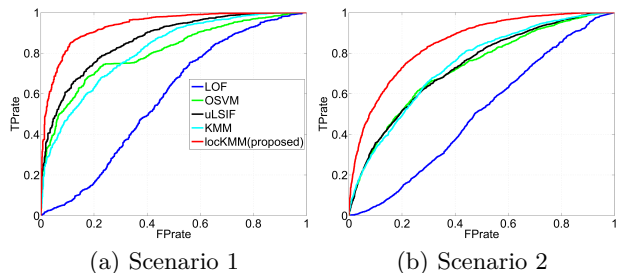


Figure 2: The ROC curves of different novelty detection methods for the two test scenarios.

Table 2: AUC and Prec@t of 10 runs on the tweet data.

Scenario	S1		S2	
	AUC	Prec@t	AUC	Prec@t
LOF	60.01±01.63	73.09±00.80	51.90±01.02	59.53±00.99
OSVM	82.55±01.52	81.00±01.41	71.73±00.96	70.69±00.68
uLSIF	86.76±00.67	84.94±00.79	70.87±03.23	70.18±02.31
KMM	81.70±00.83	81.74±00.56	75.53±00.84	73.87±00.63
locKMM	93.39±00.43	90.16±00.69	84.21±01.66	79.75±01.61

- **LOF**: Local Outlier Factor method [3]. The neighborhood size k is set to 7, the same as *locKMM* method.
- **uLSIF**: The unconstrained Least Squares Importance Fitting [12]. This is a well-known density-ratio estimation method. The Gaussian kernel is used and its width is selected using 10-fold cross validation.
- **KMM**: Kernel Mean Matching [13]. This is another well-known density-ratio estimation method. As in [13], the Gaussian kernel is used and its kernel width is set to the median distance between samples.
- **locKMM**: Locally adaptive Kernel Mean Matching. This is the proposed method. The neighborhood size k is set to 7. For KMM and *locKMM*, we set $b = 1000$ and $\epsilon = 1e - 10$.

4.2 Results

We use the Receiver Operating Characteristic (ROC) curve, Area Under the ROC Curve (AUC) [8], and the Precision-at-t (Prec@t) [23] as performance evaluation measures, which are commonly adopted in the literature of novelty detection. It should be noted that the dynamic NMF method is designed to identify novel topics only and does not own the ability of generating novelty scores for test instances. Therefore, the NMF method is included as a comparison in last experiment of this study (the key words detection task), but is not shown in the other performance measurement.

Fig. 2 plots the ROC curves of different novelty detection algorithms for the two test cases. As can be observed, the proposed *locKMM* method outperforms other methods over the entire space. The quantitative detection results in terms of AUC and Prec@t on the two test scenarios are reported in Table 2, which show large margin of performance improvement of the proposed method in all cases.

Further, we evaluate the detection results by precision, recall, F1, and F0.5. Table 3 reports the average detection performance of 10 runs. Each method is by varying the threshold of novelty scores and reports the best result

Table 4: Top 20 words of novel topics being detected by different methods (*: words of emerging novel topic; ^: words of evolving novel topic; underlined: mis-targeted words).

Method	Top 20 key words
Truth	<i>investing</i> (emerging-novel*): forex, usd, market, stock, trade, euro, new, otc, gold, dollar, jpy, outlook, price, march, business, report, daily, data, home, rise. <i>media</i> (evolving-novel^): tip, helium, und, guide, save, money, korea, help, article, effect, plan, learn, treatment, teach, home, care, idea, child, overview, social.
LOF	forex*, usd*, market*, stock*, <u>look</u> , <u>fashion</u> , tip^, home*^, trade*, new*, euro*, <u>spring</u> , <u>week</u> , daily*, gold*, dollar*, <u>trend</u> , <u>launch</u> , helium^, company.
OSVM	forex*, usd*, market*, stock*, trade*, new*, euro*, home*^, tip^, business*, look, gold*, otc*, helium^, dollar*, outlook*, daily*, week, price*, jpy*.
uLSIF	forex*, usd*, market*, stock*, trade*, tip^, new*, home*^, business*, euro*, helium^, otc*, dollar*, gold*, <u>look</u> , jpy*, march*, <u>invest</u> , company, plan^.
KMM	forex*, usd*, market*, stock*, trade*, tip^, home*^, new*, <u>look</u> , euro*, gold*, helium^, dollar*, jpy*, business*, <u>fashion</u> , <u>week</u> , price*, otc*, outlook*.
NMF	usd*, forex*, euro*, jpy*, outlook*, stock*, market*, trade*, aapl, data*, nyse, wfc, big, fail, jpm, otc*, gold*, new*, link, johansson.
locKMM	forex*, usd*, market*, stock*, trade*, tip^, new*, home*^, business*, euro*, helium^, gold*, dollar*, otc*, jpy*, price*, march*, money^, daily*, <u>health</u> .

of F0.5. From Table 3, the superiority of *locKMM* method is obvious. The main reason lies in the fact that tweets from emerging topics may also demonstrate the evolving aspects, which include overlaps with previous concepts. Another observation is that the *locKMM* method maintains high levels of accuracy with reasonable recalls. This is a preferred feature in information retrieval applications as the novelty detection. The relative low recalls mean a number of tweets from novel topics are not distinguishable from existing topics. This is likely caused by the limitation of the representation model.

Table 4 presents the top 20 weighted words of scenario ‘S1’, which are extracted by different detection methods, as well as the ground truth. We also compare with a recent work [21] that uses dynamic NMF to learn evolving and emerging topics. The results clearly show that our approach can effectively detect both emerging and evolving novel topics with the minimal number of mis-targeted words. The only mis-targeted word ‘health’ in fact is the 23rd ranked word in the target topic ‘media’.

5. CONCLUSION

The ability to track both emerging and evolving novel content in the social media is important to help us understand the full view of social events, while a number of challenges lay ahead. Though traditional algorithms achieve acceptable performance, they are limited to detecting either emerging novelties or evolving novelties. In this work, a novel locally adaptive kernel mean matching algorithm was proposed, which is built on the success of the idea of using density ratio as a measure of evolving novelty and augments the estimation of density ratio with information about the neighborhood structure of each data instance to capture the emerging novelty. A comparison to different methods demonstrates the superiority of the proposed approach.

Table 3: Detection performance in terms of precision, recall, F1, and F0.5.

Scenario	S1				S2			
	P	R	F1	F0.5	P	R	F1	F0.5
LOF	71.48±00.66	91.79±02.25	80.36±01.08	74.78±00.74	60.03±00.61	84.89±01.45	70.32±00.73	63.76±00.61
OSVM	84.28±00.82	78.80±02.64	81.42±01.53	83.11±00.94	72.58±00.89	65.85±01.38	69.05±01.01	71.12±00.89
uLSIF	88.82±01.16	75.53±03.66	81.58±01.97	85.76±00.91	66.50±03.05	84.66±04.45	74.35±01.23	69.41±02.08
KMM	83.24±00.55	77.30±01.88	80.15±01.01	81.97±00.56	75.15±00.72	69.63±01.04	72.28±00.66	73.97±00.61
locKMM	95.58±00.41	75.93±01.54	84.62±00.96	90.87±00.52	82.61±01.87	72.28±01.02	77.10±01.33	80.31±01.63

6. ACKNOWLEDGMENTS

This publication was made possible by a grant from the Qatar National Research Fund through National Priority Research Program (NPRP) No. 06-1220-1-233. Its contents are solely the responsibility of the authors.

7. REFERENCES

- [1] L. M. Aiello, G. Petkos, C. Martin, D. Corney, S. Papadopoulos, R. Skraba, A. Goker, I. Kompatsiaris, and A. Jaimes. Sensing trending topics in twitter. *IEEE Transactions on Multimedia*, 15(6):1268–1282, 2013.
- [2] J. Allan. Introduction to topic detection and tracking. In *Topic detection and tracking*, pages 1–16. Springer, 2002.
- [3] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. LOF: identifying density-based local outliers. In *ACM Sigmod Record*, volume 29, pages 93–104, 2000.
- [4] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, 2011.
- [5] I. Dhillon, Y. Guan, and B. Kulis. A unified view of kernel k-means, spectral clustering and graph cuts. Technical Report TR-04-25, University of Texas at Austin, 2005.
- [6] A. K. Farahat, A. Ghodsi, and M. S. Kamel. A novel greedy algorithm for Nyström approximation. In *AISTATS*, pages 269–277, 2011.
- [7] A. K. Farahat and M. S. Kamel. Statistical semantics for enhancing document clustering. *Knowledge and Information Systems*, 28(2):365–393, 2011.
- [8] T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [9] D. Filev and F. Tseng. Novelty detection based machine health prognostics. In *the 2006 International Symposium on Evolving Fuzzy Systems*, pages 193–199. IEEE, 2006.
- [10] A. Guille and C. Favre. Mention-anomaly-based event detection and tracking in twitter. In *the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 375–382. IEEE, 2014.
- [11] H. Hachiya, M. Sugiyama, and N. Ueda. Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition. *Neurocomputing*, 80:93–101, 2012.
- [12] S. Hido, Y. Tsuboi, H. Kashima, M. Sugiyama, and T. Kanamori. Statistical outlier detection using direct density ratio estimation. *Knowledge and Information Systems*, 26(2):309–336, 2011.
- [13] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf. Correcting sample selection bias by unlabeled data. In *NIPS*, pages 601–608, 2007.
- [14] M. Karkali, F. Rousseau, A. Ntoulas, and M. Vazirgiannis. Efficient online novelty detection in news streams. In *Web Information Systems Engineering–WISE 2013*, pages 57–71. 2013.
- [15] M. Kemmler, E. Rodner, E.-S. Wacker, and J. Denzler. One-class classification with gaussian processes. *Pattern Recognition*, 46(12):3507–3518, 2013.
- [16] Y.-Q. Miao, A. K. Farahat, and M. S. Kamel. Auto-tuning kernel mean matching. In *Workshop on Incremental Clustering, Concept Drift and Novelty Detection at the 2013 IEEE ICDM*, pages 560–567, 2013.
- [17] Y.-Q. Miao, A. K. Farahat, and M. S. Kamel. Discriminative density-ratio estimation. In *2014 SIAM International Conference on Data Mining (SDM)*, pages 830–838, 2014.
- [18] C. A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. *The Journal of Machine Learning Research*, 7:2651–2667, 2006.
- [19] S. Middleton, L. Middleton, and S. Modafferi. Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, 29(2):9–17, 2014.
- [20] Oxford University Press. Rt this: OUP dictionary team monitors twitterer’s tweets. <http://blog.oup.com/2009/06/oxford-twitter/>, Retrieved in January 2015.
- [21] A. Saha and V. Sindhwani. Learning evolving and emerging topics in social media: a dynamic nmf approach with temporal regularization. In *WSDM*, pages 693–702. ACM, 2012.
- [22] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt. Support vector method for novelty detection. In *NIPS*, pages 582–588, 1999.
- [23] A. J. Smola, L. Song, and C. H. Teo. Relative novelty detection. In *AISTATS*, pages 536–543, 2009.
- [24] X. Yan, J. Guo, S. Liu, X. Cheng, and Y. Wang. Learning topics in short texts by non-negative matrix factorization on term correlation matrix. In *SDM*, pages 749–757, 2013.
- [25] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *NIPS*, pages 1601–1608, 2004.
- [26] A. Zubiaga and H. Ji. Harnessing web page directories for large-scale classification of tweets. In *WWW’2013 Companion*, pages 225–226. ACM, 2013.