Twitter Trend Detection and Classification

Abstract

There are three review cycles for ICML 2013, with full paper submissions due on October 1, 2012; December 15, 2012; and February 15, 2013. Reviewing will be blind to the identities of the authors, and therefore identifying information must not appear in any way in papers submitted for review. Submissions must be in PDF, with an 8 page length limit (upto 9 pages including references).

1. Introduction

2. Related Work

The paper (Nikolov, 2012) introduced a new approach to detecting Twitter trends in a nonparametric method. It presented a general time series analysis method that can be used to detect the future trend on twitter. This paper used both positive trending examples and negative, or non-trending example to vote for the current observation sequence. Then based on the voting result to find out whether it's a trend or not. The result of this paper is fairly good. In 79% percent of cases, this paper detected trending topics earlier than Twitter (1.43 hours earlier), and it managed to keep an error rate of around 95% (4% false positive rate, 95% true positive rate). As this paper's experiment result shows that it is fairly good to use time series data to predict trends in twitter, we believe the time series prior of the peak would be a good indictor of the trend pattern.

The paper (Asur et al., 2011) made a detail analysis on different kinds of factors that may affect the persistence and decay of twitter trend based on history data and experiments. It showed that the resonance of the content with the users of the social network plays a major role in causing trends.It also demonstrate empirically how factors such as user activity and number of followers do not contribute strongly to trend creation and its propagation. Based on the conclusion of

───────────

this paper, we believe that content of twitter activity has a major influence to the type of twitter trends.

3. Dataset

We extract nearly 200 million tweets from Brendan's "Gardenhose" archive from June 2009 to February 2010 as same time period as described in the paper (Yang & Leskovec, 2011). And we further extract 2.5 million all the tweets containing the top 998 most frequently appeared hashtags. This data size is about 600MB compressed and >3GB uncompressed.

Besides that, we also build our time series based on the number of tweets in each one hour unit. For the known peak time series, we build time series using the truncated time series of 128 hours with peak located at the $1/3$ of the 128 hours. For the unknown peak time series, we build time series throughout the whole eight months.

4. Method

Twitter trends are measured by the activity of Twitter hashtags or topics. In our work, we use the activity of hashtag to measure Twitter trend. We want to predict the future trend of an emerging hashtag/topic based on either content or some previous temporal information from tweets containing that hashtag. For example, we want to predict future trend of hashtag "#protest" based on some previous tweets containing that hashtag, like `RT Iranian women must liberate themselves http://rurl.org/1rlv #protest #nytimes`.

The problem can be approached as a supervised learning problem. We first assign each hashtag a trending pattern based on its time series. Although this problem is a supervised learning problem, we label the instances through clustering. From then on, we can train a classifier based on either text content or time series from hashtag histogram. And we could then classify the hashtag to different patterns using our classifier.

The content and temporal information maybe beneficial in predicting the trend pattern of a hashtag. We use content information because some words or

phrases reveal their trend to some extent. For example, the trend pattern of "#olympics" is different from "#apple" that "#olympics" is reoccurring every four years and it will only be a trend when it is going on, while "#apple" is a daily trend and can trigger a trend very often like it has new product to release or new dispute with competitors. The context words associate with "#olympics", like athlete names "Michael Phelps" and medals "Gold Medal", and "#apple", like product name "iPhone" and competitor name "Samsung", are quite informative when training and testing using classifiers. We also use temporal information like the time series (histogram volume) as an another way to predict trend. The ups and downs in different time series can also indicate different patterns.

In our paper, we have used K-SC algorithm to do Trend Clustering of time series (Yang & Leskovec, 2011). We have tried four different approaches to do Trend Prediction.

4.1. Trend Clustering

We do Trend Clustering mainly to label the data, which assign a label to a time series according to which cluster it belongs. We use these labeled data to do training and testing on our classifiers.

We use the K-SC algorithm (Yang & Leskovec, 2011) to do clustering of the 128 hours time series with peak located at 1/3 of the 128 hours. This algorithm is very similar to K-means algorithm. It has for each data points calculate which nearest cluster centroid it belongs and then compute the new cluster centroid again. The only two differences with K-means are K-SC uses the special designed similarity metric, which is similar to Time Wrapping Distance (compute distance after shifting and scaling). Another difference is K-SC uses a new method to compute cluster centroid, which is essentially select the centroid that minimize the squared distance over all data points to their closest centroid using the new similarity metric.

We choose to K=6 and the clustering result in Figure 1

4.2. Trend Prediction

For predicting Twitter trend, we can consider using both content information and temporal information as the clue for the trending activity.

4.2.1. Bag-of-Word Naive Bayes Trend Prediction

The Naive Bayes Method is the first one we come up with in this project. Since this is a classification problem and tweets are composed of words and tokens, the
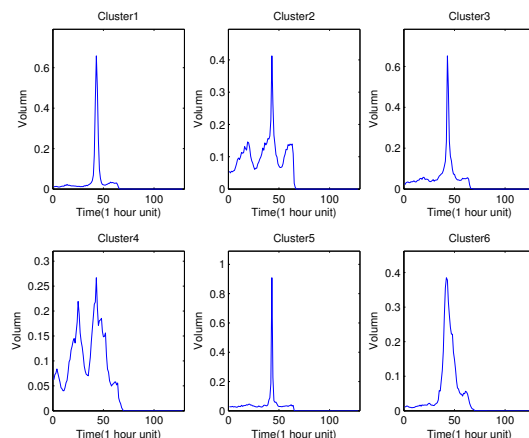

Figure 1. All model performance using different metrics.

use of bag of word model is very intuitive. The Naive Bayes algorithm we implemented is similar to the one we did in homework. For the training set, each tweet has a label (1 to 6), so each of them can be treated as a training sample. We first utilize LingPipe to tokenize the tweets by considering lowercase, stemming, etc. Then with these separate tokens, we count the number of co-occurrence and other metrics needed to build the Naive Bayes model. The whole process was done very quickly and efficiently using Hadoop. After getting all the model parameters we need, we start working on the test data. The log likelihood of each test sample can be written as:

$$(\prod_{j=1}^{d} \frac{C(X_j = x_j \wedge Y = y') + mq_j}{C(Y = y') + m}) \frac{C(Y = y') + mq_j}{C(Y = ANY) + m} \tag{1}$$

Using this equation, we can easily classify our test tweets data into one of the six categories. But here we made some changes on the test process. We first combine several tweets of the same hashtag into one document, then apply the NB classifiers. The reason we did so is because the tweets are often very short, thus making the token distribution very sparse which may lead to bad result.

In addition to this traditional Naive Bayes method, we also take the weight of token into account. We believe that some of the words must be more important than others when apply to certain situations or categories. For example, people won't talk about Hurricane too often usually, but when we have a real Hurricane in the world, the world Hurricane becomes important. To discover such important tokens and give different weight on them, we first calculate both

background and foreground token distributions, and use Chi-square method to find those important words. The Chi-Square method can be written as

$$\frac{(fg - bg)^2}{bg} \qquad (2)$$

By using this model, we successfully find out those important words give certain cluster. For example, in cluster 3, the top 10 words include mac, apple, ipad, ipod, chrome, android, etc., which are all relevant to IT fields. Though this token weight makes sense, when apply to Naive Bayes Model, we didn't find good improvement on the final results. Maybe our combine method is not good enough, which we may dig deeper later.

4.2.2. BAG-OF-WORD LOGISTIC REGRESSION TREND PREDICTION

In trend prediction using bag-of-word Logistic Regression, we treat all words in the tweets containing the hashtag as a document (bag-of-word). The document's label is the hashtag's label assigned in Trend Clustering part based on the clustering result from hashtag's time series. For this specific Logistic Regression problem, we want to maximize log likelihood $L(w) = \Pi_l P(y^l|x^l, w)$, where x^l represents the lth training document (BOW of all words in tweets containing the same hashtag), y^l represents the class label of lth training document and w is the model weight. Using stochastic gradient descent, we can derive the training rules as

$$w_{ij} = w_{ij} + \lambda(y^l - \hat{y}^l)x^l - 2\lambda\mu w_{ij} \qquad (3)$$

where w_{ij} is the ith classifier's jth element in weight vector, y^l is the true class label of lth training document, $\hat{y}^l$ is the predicted probability of chewing this label as the class label of lth training document, x^l is the bag-of-words in lth training document, λ is the gradient step size, μ is the regularization constant. To make our algorithm more fast and space efficient for processing large data, we use the delayed update stochastic gradient descend with hash trick that introduced in class.

Since this specific Logistic Regression classifier is a multi class classifier, we train multiple binary classifier and choose the label with the largest weight in the test phase. The testing document consists of bag-of-words all words in tweets containing the same hashtag. The test rules is

$$y^l = \arg\max_i \sum_j w_{ij}x_j^l \qquad (4)$$

where w_{ij} is the ith classifier's jth element in weight vector, y^l is the predicted class label for lth testing instance, x_j^l is the jth word (feature after hash trick) in lth testing document.

4.2.3. TIME SERIES SIMILARITY MEASUREMENT

After working on content based measurement of Twitter trend, we turn to use temporal information with hashtag to classify and predict the type of Twitter trend. Based on the former research(Nikolov, 2012) about the twitter trend, we believe time series prior of the peak would be a good indictor of the trend pattern. So we add this part time series information to do classification and prediction. Different from (Nikolov, 2012), we not only need to predict if the sequence of tweets will be a trend, but also decide which trend pattern it would be. The meaning of doing this is find the spread pattern of twitter trend before the trend reaches its peak. Because of time limitation, we applied a parametric model to predict and classify the trend pattern instead of using a nonparametric prediction. The whole process could be divided into two steps. Trend prediction and Trend classification.

The trend prediction is find the twitter activity that could become a trend. We can assume that a normal trend model of tweet activity is roughly constant with occasional jumps. So the detection process is similar with a filter, which is used to filter the occasional jump and find the true jump that become a twitter trend.

Here we define a trend by two variables, jump rate and popularity duration. Jump rate is

$$\frac{\alpha * standardvariance}{mean} \qquad (5)$$

The standard variance and mean refer to the twitter activity of the according time interval. Popularity duration is the total volume in a time interval. If the popularity duration is beta times higher than the mean popularity duration value. We will consider this is a trend. As this is a parametric model, alpha and beta needs to be estimated by experiment. When monitoring the twitter activity, we use a sliding window to check if current trend would become a twitter trend. A acceleration step applied here that we don't need to compute the parametric model every hour but only when the twitter activity has a large increase. The above detection process is illustrated in Figure 2. When the sliding window comes to a beginning of trend, we can see in the red box, the jump rate
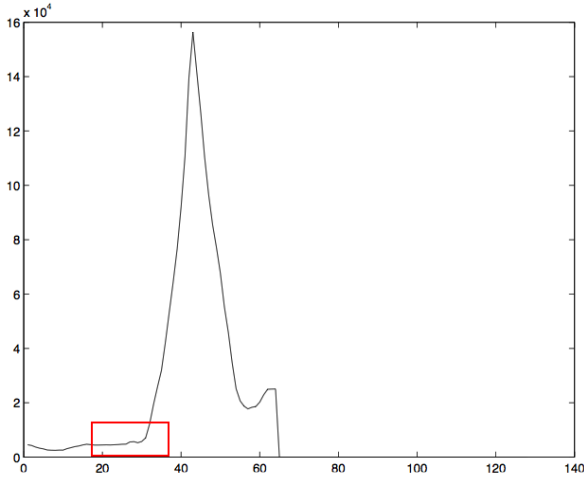
Figure 2. The detection of a big jump

and popularity duration is satisfied by the parametric model. We will consider this is a big jump, which imply the follow twitter activity will become a trend. Then we will come to the classification step to decide the type of the possible trend.

In the above step, we already get the trend patterns. So we apply a similarity measurement to classify the pattern of possible trend. As this twitter time series activity is discrete data, we compute the similarity of current twitter activity to each patterns for each hour. In another words, each hour in the time interval will vote for each trend pattern based on similarity between the instance in detected time interval and trend patterns. The similarity is defined by a decaying exponential, in this form, the nearest time series data has more weights while the non-similar data has less weight. Similarity defined as follows:

$$Similarity(instance, pattern) = e^{-d(i,p)} \quad (6)$$

In the above equation, i means the time series data at time i. p means the detected pattern time series data in the according time i. $d(i, p)$ is the euclidean distance between i and p. The similarity measurement process is illustrated in figure 3 The left subfigure in figure 3 is a section of twitter activity of certain hashtag, which is the time series data of 8 month. We can see that there could be several small jumps leading up to a big jump. The right subfigure in figure 3 is a detected pattern of twitter trend, which is the time series data of 128 hours of all twitter activity in this pattern.
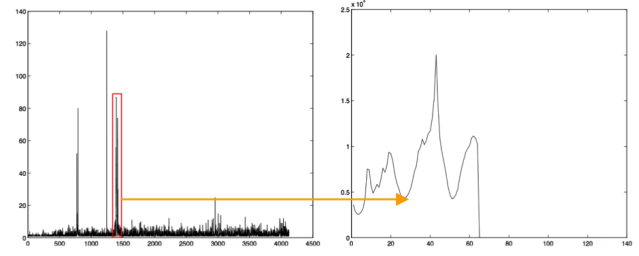

Figure 3. Illustration of similarity measurement process

4.2.4. Time Series Logistic Regression Trend Prediction

Although bag-of-words model can give us some hints about future trend, using content information to predict temporal pattern is somewhat limited. Because words alone are not always correlated with trend. Therefore, we use directly a time interval in time series (hashtag histogram) to predict the future trend pattern.

Since we have clustered the time series using the 128 hours time series, placing the peak at one third in the whole 128 hours time series, we assume that the time interval prior to a trend (0-43 hour) is a good indicator of future pattern of trend. We define trend as a sudden jump in the time series here. Therefore, we still model the problem as a classification problem that we can train and test using Logistic Regression based on the data points in a short interval that in the very beginning of time series.

We always choose the interval prior to a trend (sudden jump in time series) to train and test, since we want to predict the future pattern of trend as quickly as possible with least amount of previous known data. Therefore, the problem becomes to detect trend and use the interval of time series when the trend just begin gaining popularity to train and test our classifier. We can then try to solve the transformed problems with two assumptions here: predict trend patterns with known trend and predict trend patterns with unknown trend.

Predict Trend Patterns with Known Trend: We use the only trend that leads to peak (sudden jump in highest time series globally) in time series as the best trend as the interval for training and testing. With known most salient trend, we could get more good training instances and thus good classifier which leads to better testing results.

Predict Trend Patterns with Unknown Trend: We now need to detect a relative good trend (sudden jump in highest time series locally) by our trend de-

Table 1. Model performance using different metrics

Model Name	Precision	Recall	F-measure
BOW_NB	0.22	0.22	0.22
BOW_LR	0.34	0.13	0.19
TS_LR_KT	**0.57**	**0.57**	**0.57**
TS_LR_UNKT	0.31	0.34	0.33
TS_SM_UNKT	0.19	0.19	0.19
TS_LR_UNKT	0.31	0.34	0.33

tection algorithm. The detection algorithm needs to find real potential trend instead of small fluctuations. The algorithm is parametric that it needs parameter tuning to avoid this.

5. Experiment

5.1. Content (BOW) vs Temporal (Time Series)

We can see from Table 1 the final evaluation result from the table below. Note that "BOW_NB" and "BOW_LR" represents Naive Bayes and Logistic Regression using Bag-of-Word model respectively. "TS_LR_KT" and "TS_LR_UNKT" represents Logistic Regression with known trend and unknown trend using time series model respectively. The time series model only have trained and tested using only a interval of nine data points (first nine hours of a trend which the popularity is about to go up).
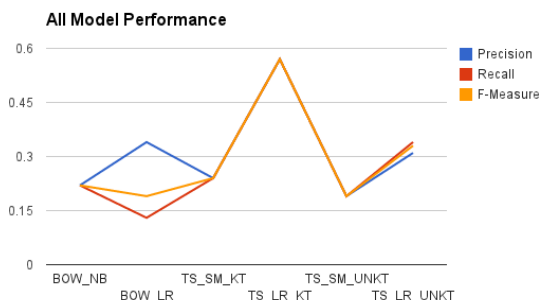

Figure 4. All model performance using different metrics.

As we can see from Figure 4, we can see that the time series based model gives us way more better performance than content based model (TS is much better than BOW). The reason might be tweets content can only provide little information about the trend of the underlying topic they talk about. Or we haven't found a better way to use the content data.

Besides, Logistic Regression (LR) classifier works better than Naive Bayes (NB) classifier in content based Bag-of-Word model. This is not surprising that LR works better than NB, since NB assume the oder of words does't not matter, namely independence among a sequence of words.

5.2. Known vs Unknown Trend

As shown in Figure 4, we can run our classification algorithm on both previous known trend (tested on 128 hours time series and regard the interval around peak as most salient trend) and unknown trend (tested on 8 months time series and detected most salient trend by our detection algorithm). We can see from the figure below that the Precision, Recall and F-measure became worse when we switch to use our own trend detection algorithm.

This indicates that with a better trend detection algorithm, we can use the real time streaming data to predict a trend and a future hottest emerging trending pattern to a specific hashtag/topic.

6. Conclusion

Tweet content has a big connection with tweet trend patterns. But the key information for this is difficult to extract. Temporal information is the direct representation of tweet trend and the time interval prior to a trend is a good indicator of future pattern of trend.

6.1. Templates for Papers

Electronic templates for producing papers for submission are available for LaTeX and Microsoft Word. Templates are accessible on the World Wide Web at: `http://icml.cc/2013/`

Send questions about these electronic templates to `program@icml.cc`.

The formatting instructions below will be enforced for initial submissions and camera-ready copies.

- The maximum paper length is 8 pages excluding references, and 9 pages including references.

- Do not alter the style template; in particular, do not compress the paper format by reducing the vertical spaces.

- Do not include author information or acknowledgments in your initial submission.

- Place figure captions *under* the figure (and omit titles from inside the graphic file itself). Place table captions *over* the table.

- References must include page numbers whenever possible and be as complete as possible. Place multiple citations in chronological order.

Please see below for details on each of these items.

6.2. Submitting Papers

Submission to ICML 2013 will be entirely electronic, via a web site (not email). The URL and information about the submission process are available on the conference web site at

`http://icml.cc/2013/`

Paper Deadline: The deadline for paper submission to ICML 2013 is at 23:59 Universal Time (3:59 Pacific Daylight Time) on the due date (October 1, December 15, or February 15, depending on the review cycle). If your full submission does not reach us by this time, it will not be considered for publication. There is no separate abstract submission.

Anonymous Submission: To facilitate blind review, no identifying author information should appear on the title page or in the paper itself. Section 6.7 will explain the details of how to format this.

Simultaneous Submission: ICML will not accept any paper which, at the time of submission, is under review for another conference or has already been published. This policy also applies to papers that overlap substantially in technical content with conference papers under review or previously published. ICML submissions must not be submitted to other conferences during ICML's review period. Authors may submit to ICML substantially different versions of journal papers that are currently under review by the journal, but not yet accepted at the time of submission. Informal publications, such as technical reports or papers in workshop proceedings which do not appear in print, do not fall under these restrictions.

To ensure our ability to print submissions, authors must provide their manuscripts in **PDF** format. Furthermore, please make sure that files contain only Type-1 fonts (e.g., using the program `pdffonts` in linux or using File/DocumentProperties/Fonts in Acrobat). Other fonts (like Type-3) might come from graphics files imported into the document.

Authors using **Word** must convert their document to PDF. Most of the latest versions of Word have the facility to do this automatically. Submissions will not be accepted in Word format or any format other than PDF. Really. We're not joking. Don't send Word.

Those who use LaTeX to format their accepted pa-

pers need to pay close attention to the typefaces used. Specifically, when producing the PDF by first converting the dvi output of LaTeX to Postscript the default behavior is to use non-scalable Type-3 PostScript bitmap fonts to represent the standard LaTeX fonts. The resulting document is difficult to read in electronic form; the type appears fuzzy. To avoid this problem, dvips must be instructed to use an alternative font map. This can be achieved with something like the following commands:

dvips -Ppdf -tletter -G0 -o paper.ps paper.dvi
ps2pdf paper.ps

Note that it is a zero following the "-G". This tells dvips to use the config.pdf file (and this file refers to a better font mapping).

Another alternative is to use the **pdflatex** program instead of straight LaTeX. This program avoids the Type-3 font problem, however you must ensure that all of the fonts are embedded (use `pdffonts`). If they are not, you need to configure pdflatex to use a font map file that specifies that the fonts be embedded. Also you should ensure that images are not downsampled or otherwise compressed in a lossy way.

Note that the 2013 style files use the `hyperref` package to make clickable links in documents. If this causes problems for you, add `nohyperref` as one of the options to the `icml2013` usepackage statement.

6.3. Reacting to Reviews

We will continue the ICML tradition in which the authors are given the option of providing a short reaction to the initial reviews. These reactions will be taken into account in the discussion among the reviewers and area chairs.

6.4. Submitting Final Camera-Ready Copy

The final versions of papers accepted for publication should follow the same format and naming convention as initial submissions, except of course that the normal author information (names and affiliations) should be given. See Section 6.7.2 for details of how to format this.

The footnote, "Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute." must be modified to "*Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28. Copyright 2013 by the author(s)."

For those using the LaTeX style file, simply change

`\usepackage{icml2013}` to

`\usepackage[accepted]{icml2013}`

Authors using **Word** must edit the footnote on the first page of the document themselves.

Camera-ready copies should have the title of the paper as running head on each page except the first one. The running title consists of a single line centered above a horizontal rule which is 1 point thick. The running head should be centered, bold and in 9 point type. The rule should be 10 points above the main text. For those using the LATEX style file, the original title is automatically set as running head using the `fancyhdr` package which is included in the ICML 2013 style file package. In case that the original title exceeds the size restrictions, a shorter form can be supplied by using

`\icmltitlerunning{...}`

just before `\begin{document}`. Authors using **Word** must edit the header of the document themselves.

All submissions must follow the same format to ensure the printer can reproduce them without problems and to let readers more easily find the information that they desire.

6.5. Length and Dimensions

Papers must not exceed eight (8) pages, including all figures, tables, and appendices, but excluding references. When references are included, the paper must not exceed nine (9) pages. Any submission that exceeds this page limit or that diverges significantly from the format specified herein will be rejected without review.

The text of the paper should be formatted in two columns, with an overall width of 6.75 inches, height of 9.0 inches, and 0.25 inches between the columns. The left margin should be 0.75 inches and the top margin 1.0 inch (2.54 cm). The right and bottom margins will depend on whether you print on US letter or A4 paper, but all final versions must be produced for US letter size.

The paper body should be set in 10 point type with a vertical spacing of 11 points. Please use Times Roman typeface throughout the text.

6.6. Title

The paper title should be set in 14 point bold type and centered between two horizontal rules that are 1 point thick, with 1.0 inch between the top rule and the top edge of the page. Capitalize the first letter of content words and put the rest of the title in lower case.

6.7. Author Information for Submission

To facilitate blind review, author information must not appear. If you are using LATEX and the `icml2013.sty` file, you may use `\icmlauthor{...}` to specify authors. The author information will simply not be printed until `accepted` is an argument to the style file. Submissions that include the author information will not be reviewed.

6.7.1. SELF-CITATIONS

If your are citing published papers for which you are an author, refer to yourself in the third person. In particular, do not use phrases that reveal your identity (e.g., "in previous work (Langley, 2000), we have shown . . . ").

Do not anonymize citations in the reference section by removing or blacking out author names. The only exception are manuscripts that are not yet published (e.g. under submission). If you choose to refer to such unpublished manuscripts (Author, 2011), anonymized copies have to be submitted as Supplementary Material via CMT. However, keep in mind that an ICML paper should be self contained and should contain sufficient detail for the reviewers to evaluate the work. In particular, reviewers are not required to look a the Supplementary Material when writing their review.

6.7.2. CAMERA-READY AUTHOR INFORMATION

If a paper is accepted, a final camera-ready copy must be prepared. For camera-ready papers, author information should start 0.3 inches below the bottom rule surrounding the title. The authors' names should appear in 10 point bold type, electronic mail addresses in 10 point small capitals, and physical addresses in ordinary 10 point type. Each author's name should be flush left, whereas the email address should be flush right on the same line. The author's physical address should appear flush left on the ensuing line, on a single line if possible. If successive authors have the same affiliation, then give their physical address only once.

A sample file (in PDF) with author names is included in the ICML2013 style file package.

6.8. Abstract

The paper abstract should begin in the left column, 0.4 inches below the final address. The heading 'Abstract' should be centered, bold, and in 11 point type. The abstract body should use 10 point type, with a vertical spacing of 11 points, and should be indented 0.25 inches more than normal on left-hand and right-hand margins. Insert 0.4 inches of blank space after

the body. Keep your abstract brief and self-contained, limiting it to one paragraph and no more than six or seven sentences.

6.9. Partitioning the Text

You should organize your paper into sections and paragraphs to help readers place a structure on the material and understand its contributions.

6.9.1. SECTIONS AND SUBSECTIONS

Section headings should be numbered, flush left, and set in 11 pt bold type with the content words capitalized. Leave 0.25 inches of space before the heading and 0.15 inches after the heading.

Similarly, subsection headings should be numbered, flush left, and set in 10 pt bold type with the content words capitalized. Leave 0.2 inches of space before the heading and 0.13 inches afterward.

Finally, subsubsection headings should be numbered, flush left, and set in 10 pt small caps with the content words capitalized. Leave 0.18 inches of space before the heading and 0.1 inches after the heading.

Please use no more than three levels of headings.

6.9.2. PARAGRAPHS AND FOOTNOTES

Within each section or subsection, you should further partition the paper into paragraphs. Do not indent the first line of a given paragraph, but insert a blank line between succeeding ones.

You can use footnotes[1] to provide readers with additional information about a topic without interrupting the flow of the paper. Indicate footnotes with a number in the text where the point is most relevant. Place the footnote in 9 point type at the bottom of the column in which it appears. Precede the first footnote in a column with a horizontal rule of 0.8 inches.[2]

6.10. Figures

You may want to include figures in the paper to help readers visualize your approach and your results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background.

---

[1]For the sake of readability, footnotes should be complete sentences.

[2]Multiple footnotes can appear in each column, in the same order as they appear in the text, but spread them across columns and pages if possible.
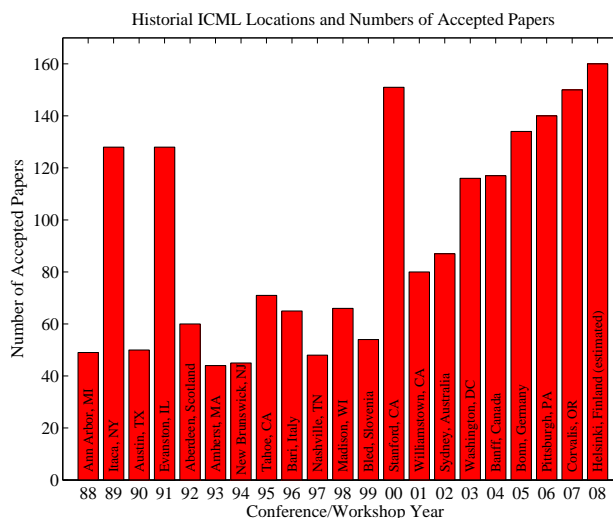

Figure 5. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead estimated.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes each curve. Do not include a title inside the figure; instead, the caption should serve this function.

Number figures sequentially, placing the figure number and caption *after* the graphics, with at least 0.1 inches of space before the caption and 0.1 inches after it, as in Figure 5. The figure caption should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left. You may float figures to the top or bottom of a column, and you may set wide figures across both columns (use the environment figure* in LaTeX), but always place two-column figures at the top or bottom of the page.

6.11. Algorithms

If you are using LaTeX, please use the "algorithm" and "algorithmic" environments to format pseudocode. These require the corresponding stylefiles, algorithm.sty and algorithmic.sty, which are supplied with this package. Algorithm 1 shows an example.

6.12. Tables

You may also want to include tables that summarize material. Like figures, these should be centered, legi-

---

Algorithm 1 Bubble Sort

  Input: data x_i, size m
  repeat
    Initialize $noChange = true$.
    for $i = 1$ **to** $m - 1$ **do**
      if $x_i > x_{i+1}$ **then**
        Swap x_i and x_{i+1}
        $noChange = false$
      end if
    end for
  until $noChange$ is $true$

Table 2. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9± 0.2	96.7± 0.2	√
CLEVELAND	83.3± 0.6	80.0± 0.6	×
GLASS2	61.9± 1.4	83.8± 0.7	√
CREDIT	74.8± 0.5	78.3± 0.6	
HORSE	73.3± 0.9	69.7± 1.0	×
META	67.1± 0.6	76.5± 0.5	√
PIMA	75.1± 0.6	73.9± 0.5	
VEHICLE	44.9± 0.6	61.5± 0.4	√

ble, and numbered consecutively. However, place the title *above* the table with at least 0.1 inches of space before the title and the same after it, as in Table 2. The table title should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left.

Tables contain textual material that can be typeset, as contrasted with figures, which contain graphical material that must be drawn. Specify the contents of each row and column in the table's topmost row. Again, you may float tables to a column's top or bottom, and set wide tables across both columns, but place two-column tables at the top or bottom of the page.

6.13. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the LaTeX bibliographic facility, use `natbib.sty` and `icml2013.bst` included in the style-file package to obtain this format.

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel's pioneering work (1959). Otherwise place the entire reference in parentheses with the authors and year sepa-

rated by a comma (Samuel, 1959). List multiple references separated by semicolons (Kearns, 1989; Samuel, 1959; Mitchell, 1980). Use the 'et al.' construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (Michalski et al., 1983).

Authors should cite their own work in the third person in the initial version of their paper submitted for blind review. Please refer to Section 6.7 for detailed instructions on how to cite your own papers.

Use an unnumbered first-level section heading for the references, and use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 10 points. The references at the end of this document give examples for journal articles (Samuel, 1959), conference publications (Langley, 2000), book chapters (Newell & Rosenbloom, 1981), books (Duda et al., 2000), edited volumes (Michalski et al., 1983), technical reports (Mitchell, 1980), and dissertations (Kearns, 1989).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

6.14. Software and Data

We strongly encourage the publication of software and data with the camera-ready version of the paper whenever appropriate. This can be done by including a URL in the camera-ready copy. However, do not include URLs that reveal your institution or identity in your submission for review. Instead, provide an anonymous URL or upload the material as "Supplementary Material" into the CMT reviewing system. Note that reviewers are not required to look a this material when writing their review.

Acknowledgments

Do not include acknowledgements in the initial version of the paper submitted for blind review.

If a paper is accepted, the final camera-ready version can (and probably should) include acknowledgements. In this case, please place such acknowledgements in an unnumbered section at the end of the paper. Typically, this will include thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

References

Asur, Sitaram, Huberman, Bernardo A., Szab, Gbor, and Wang, Chunyan. Trends in social media : Persistence and decay. *CoRR*, abs/1102.1402, 2011. URL http://dblp.uni-trier.de/db/journals/corr/corr1102.html#abs-1102-1402.

Author, N. N. Suppressed for anonymity, 2011.

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

Kearns, M. J. *Computational Complexity of Machine Learning*. PhD thesis, Department of Computer Science, Harvard University, 1989.

Langley, P. Crafting papers on machine learning. In Langley, Pat (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). *Machine Learning: An Artificial Intelligence Approach, Vol. I*. Tioga, Palo Alto, CA, 1983.

Mitchell, T. M. The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University, New Brunswick, MA, 1980.

Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (ed.), *Cognitive Skills and Their Acquisition*, chapter 1, pp. 1–51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1981.

Nikolov, Stanislav. *Trend or No Trend: A Novel Nonparametric Method for Classifying Time Series*. PhD thesis, Massachusetts Institute of Technology, 2012.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):211–229, 1959.

Yang, Jaewon and Leskovec, Jure. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining (WSDM 2011)*, pp. 177–186, New York, NY, USA, 2011. ACM.