Temporal Modeling of Twitter Posting Behavior: An Empirical Study

Matthew Michelson
InferLink Corporation
2361 Rosecrans Avenue, Suite 348
El Segundo, California 90245

John Bumgarner
U.S. Cyber Consequences Unit
P.O. Box 690087
Charlotte, NC 28227-7001

***Abstract*—This paper describes our approach for temporally modeling the posting behavior of users on the Twitter micro-blogging service. While other researchers have analyzed general trends of Twitter behavior (for instance, few users contribute most of the posts), we instead focus on modeling the temporal behavior of individual users. Specifically, we determine whether we can model the day-of-the-week posting behavior (e.g., this person is a "weekend warrior") and the within-day posting behavior (e.g., "night owl" or "morning person") for individuals. This "individual life pattern" analysis is useful for medical and public health applications. For instance, significant deviations from normal behavior might indicate that a person is potentially ill. Our contribution focuses on probabilistically modeling the posting behavior and then using Kullback-Leibler divergence to demonstrate how well the model captures a user's posting behavior.**

I. Introduction

With the growing popularity of sharing on social networks comes the increasing possibility of using social network data for social science and public health research (e.g., [1], [2], [3], [4]). However, much of this work has focused on research of aggregate behavior.

In this paper, we instead focus on modeling "individual life patterns" of social media users, specifically focusing on the Twitter micro-blogging service. Intuitively, these patterns capture aspects such as person's tendency to post information during the weekend or early in the morning. Our intention is that these individual life patterns capture the baseline behavior, identifying normal activities (e.g., usually posting messages at lunch), which we can then use to uncover deviations in normal behavior. Such deviations, in turn, could potentially identify life-changing events such as major illness, which might then warrant additional investigation (e.g, by a health provider, public health analyst, etc.). In fact, at large such an approach could potentially help signal early indications of community illness if a significant number of users have deviations in behavior.

Previous work has identified other such life-pattern information, such as geolocating Twitter users by analyzing the content of their Tweets [5] (instead of their geo-tagged posts which may be error prone). In this paper, we focus on the temporal aspect of the posting behavior, which can be combined with the previous approach to locate a user in space and associate the temporal behavior for improved health monitoring.

Specifically, our goal is to model users temporal posting behavior, noting the preferred days of the week and times within a day in which a user tends to post his or her messages. We model a specific user's posting behavior at two levels of granularity. First, we model day-of-the-week (DoW) behavior, which reflects a user's preference for posting on different days of the week. For instance, a user may do most of their posting during the week (when they have access to their work computer) or on the weekend (when he or she has more free time). Second, at one level of granularity deeper, we model the within-the-day (WtD) behavior, that is, the time of day that a user prefers to post. This captures the intuition that some people are morning posters, while others are "night owls." In this study we focus on modeling the behavior probabilistically and one of our key results is exploring Kullback-Leibler divergence as a means to characterize how well the model captures the user's behavior behavior.

The rest of this paper is organized as follows. Section II describes our probabilistic model for a user's posting behavior. Section III details our empirical study. Section IV outlines future research, and Section V presents our conclusions.

II. Modeling a User's Posting Behavior

Our main contribution is to model a user's posting behavior. At its core, our approach builds a *distribution* representing the probability that a Twitter user will send a message at a point in time (at some level of granularity). For instance, we model the probability that someone will send a Tweet on Monday, Tues., Wed., etc.[1] We model the behavior probabilistically (e.g., instead of rule-based) to capture the uncertainty for any particular data point in a user's history (for instance, someone may simply not Tweet on a day because he or she gets stuck in traffic).

Formally, given a set of granular time periods $t_i \in T$, we first define the set of observations (e.g, Tweets) for each time period as:

$$o_t = \sum o_i^t; \quad t \in T$$

We then can define the full set of observations as:

$$O_T = \sum o_t; \quad t \in T$$

[1]We then test this model on hold-out data in our empirical study to verify that the model indeed can capture the characteristic distribution in the hold out data as well.

And then we can define the distribution (model) as

$$P(O_T) = \left\{ \frac{o_t}{o_T}; \quad t \in T \right\}$$

We build this behavioral distribution for each user, noting that we can flexibly define the time periods T (e.g., they could be weekdays, hour periods within a day, etc.).

III. Empirical Study

Here we describe an empirical study of our modeling of posting behavior. We gathered data on 10,621 Twitter users who sent at least 100 messages during our collection period of 8/11/2010 through 2/8/2013. From this data, we built two models, a day-of-the-week model (DoW), which models user posting behavior for each day of the week, and a within-the-day model (WtD), which characterizes behavior during the course of a given day of the week. To construct the WtD model, rather than model each hour in a day, we instead define three disjoint time periods, "morning," "afternoon" and "night." The posts are then bucketed into each of these categories according to their posting time. The model then defines the distribution over these time periods rather than individual hours in the day. By binning the data into these groups, rather than individual hours, we felt that the model was more intuitive (the distribution reflects a "night owl" which is easier to interpret than preferring "9pm" and "11pm") and we could improve the accuracy since less data points can sufficiently capture the distribution.

We have two main motivations in this paper. First, we develop a probabilistic model of posting behavior, and second we explore the use of Kullback-Leibler divergence [6] as a means for understanding how well the model characterizes the user's posting behavior.

To that end, our procedure starts with an individual user's posts (e.g., observations), O_T^u and separates the posts into "historic" data (O_T^{hist}) and "hold out" data, which we mark as O_T^{test} (using "test" simply in order to avoid confusion with "h" in the superscript). We then build the behavioral model using the observations from O_T^{hist}, as defined above (e.g., $P(O_T^{hist})$) and measure how well that distribution describes the observations in O_T^{test} (e.g., how well $P(O_T^{hist})$ describes $P(O_T^{test})$). If it describes the distribution well, then the model can be thought to have sufficiently described the behavior.

More specifically, we run Monte Carlo simulations to build the model by randomly sampling (without replacement) from O_T^u to construct O_T^{hist}, and then using $O_T^{test} = \{O_T^u \setminus O_T^{hist}\}$ as the hold out set. These random samples then provide the observations to construct $P(O_T^{hist})$ and $P(O_T^{test})$. We note that in this study we ran 1,000 simulations per user.

We then examine how well $P(O_T^{hist})$ characterizes $P(O_T^{test})$ using the Kullback-Leibler divergence (KL) between the distributions. KL measures how well one distribution ($P(O_T^{hist})$ in this case) approximates another ($P(O_T^{test})$, and can intuitively be thought of as the information lost (in information theory terms) when you try to do the approximation. If the value of KL is zero, then $P(O_T^{hist})$ perfectly approximates $P(O_T^{test})$, so values closer to zero mean that the measured distribution from the historic data generalizes well in explaining the distribution from the hold out data.

Kullback-Leibler divergence is defined as:

$$KL(P(O_T^{test})||P(O_T^{hist})) = \sum_{t_i \in T} \log\left(\frac{P(O_{t_i}^{test})}{P(O_{t_i}^{hist})}\right) * P(O_{t_i}^{test})$$

To gain a more intuitive understanding of KL, consider Figure 1 which shows three example distributions A, B and C. The distributions A and B are constructed to be quite similar. As such, their KL value is quite small since the deviations between the distributions are pretty minimal. On the other hand, distribution C is a mirror image of distribution A. Therefore, as shown in the figure, it has a much higher value for KL divergence.

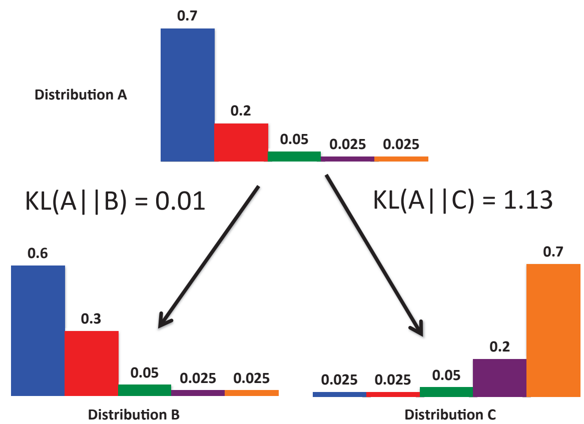

Fig. 1. Different distributions and their KL divergences

Given this experimental procedure, we then built models for each user and explored how well those models characterized the behavior. Our first experiment focused on users who posted frequently, to ensure that we had sufficient data for our analysis (later in this section we relax this condition to analyze the effects of the amounts of data). Specifically, we sub-sampled our users down to those that posted at least 5,000 times during our collection period. This resulted in 555 distinct users for the analysis. Then, we computed the percentage of those users whose distribution had an average KL below a specific threshold. For instance, for a threshold of 0.05 (quite small), we measure the percentage of the users whose KL between $P(O_T^{hist})$ and $P(O_T^{test})$ is less than 0.05 on average, across the 1,000 Monte Carlo trials. Further, we vary the size of O_T^{hist} to measure its effect as well. For instance, if we set the size of O_T^{hist} to be 30%, this means that 30% of O_T^u was used for O_T^{hist} in each of the 1,000 runs (and 70% for O_T^{test}). We tested sizes of 10% through 70% (as shown in the figure), and with 1,000 simulations per user, per threshold. The results for this DoW model are shown in Figure 2

At its peak, the models appear to explain almost 85% of the users' behavior with a KL at or below 0.1. This means that the model is accurate for a large percentage of the users.

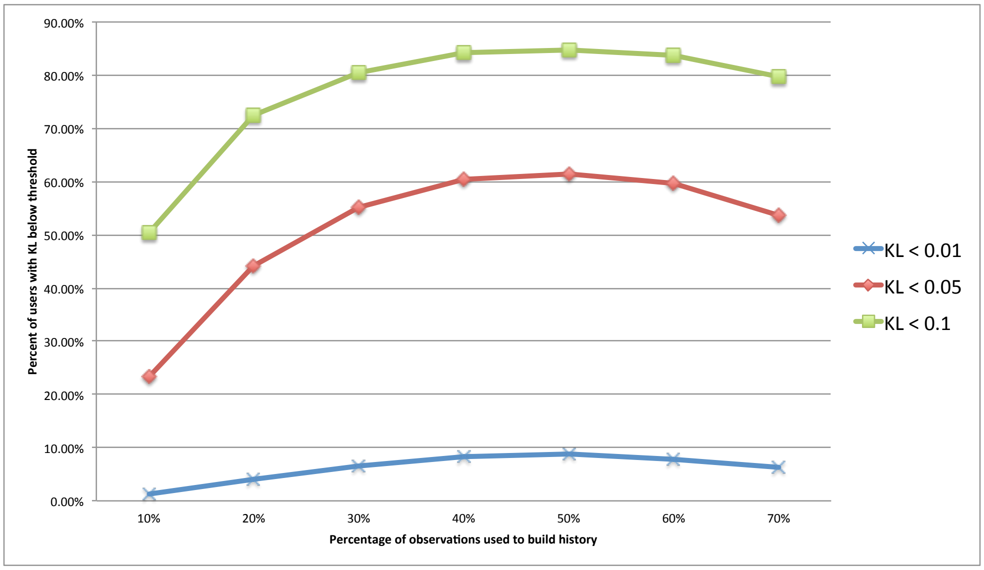

Fig. 2. Fitting distributions of posting behavior for a DoW model

TABLE I
COHORT DEFINITIONS

Cohort Membership	Number of Users	% of overall users
$\geq$ 5,000 posts	555	5.2%
$\geq$ 2,500 posts	903	8.5%
$\geq$ 1,000 posts	1,636	15.4%
$\geq$ 500 posts	2,775	26.1%
$\geq$ 250 posts	4,875	45.9%
$\geq$ 100 posts	10,621	100.0%

Interestingly after 50% the model begins to degrade. It seems that after 50% of the training data is used for modeling, the uncertainty around the data points when there is less data starts being reflected in the training data (which is why it climbs to a point at 50%) and then begins to decline as the less data in the test set begins to exhibit more variation and uncertainty. Nonetheless, we believe that these results indicate that the model captures the main characteristics of the posting behavior.

As we described above, another aspect of this study is to not only understand the effects on the model of the amount of history for each user, but also to understand how the model is affected by absolute size of a user's overall posts. Intuitively, we aim to understand whether the model holds for users who might post less frequently overall. To that end, we divided our users into (overlapping) cohorts, defined by the minimum number of posts he or she made during the collection period. At the highest end, we have our 555 users who made at least 5,000 posts during that period. At the lowest end, we have 10,621 users who made at least 100 Tweets (e.g., everyone in our collection). The full set of cohorts and their data description are given in Table III.

We then ran the same procedure and analysis as before. For each user in the cohort, for each amount of historic data, we ran 1,000 simulations and recorded the average KL for the constructed DoW model. The results are shown in Figure 3. We note, for clarity in the figure we only show the results for each cohort with the KL threshold set to 0.1.

The first result to note from the figure is that the more data we have overall for a user, the better we can model the distribution. This is fairly obvious and is reflected by the fact that the cohorts align in order vertically in the figure (e.g., the model is the best with the cohort whose users have a minimum of 5,000 posts and the worst when the cohort only requires at least 100). Perhaps less obvious are that the gaps between 2,500 and 5,000 are much smaller than the other gaps between cohorts. So it appears that after some point, a sufficient amount of data is seen to be considered "large." Also, when a user has at least 500 posts, the results seem to begin to become useful in a practical sense (for instance in this case we can characterize more than 50% of the users with this low KL).

Finally, we demonstrate that we can build within-the-day (WtD) models as well. Figure 4 shows the results for the same procedure as above, except for that the time granularity is set for the WtD model, rather than the DoW model. Interestingly, not only can we build models that explain the WtD behavior, these models outperform the DoW models. This is especially clear for the cohorts with smaller amounts of data. This is likely due to the fact that we are modeling fewer categories (three time periods instead of five days), so we have more observations for fewer categories. Regardless, this type of WtD analysis is quite useful and our results demonstrate that it is possible.

IV. Future Work

This current effort reflects our early work in temporally modeling the posting behavior on social media. We currently model at daily units (e.g., a particular day of the week or

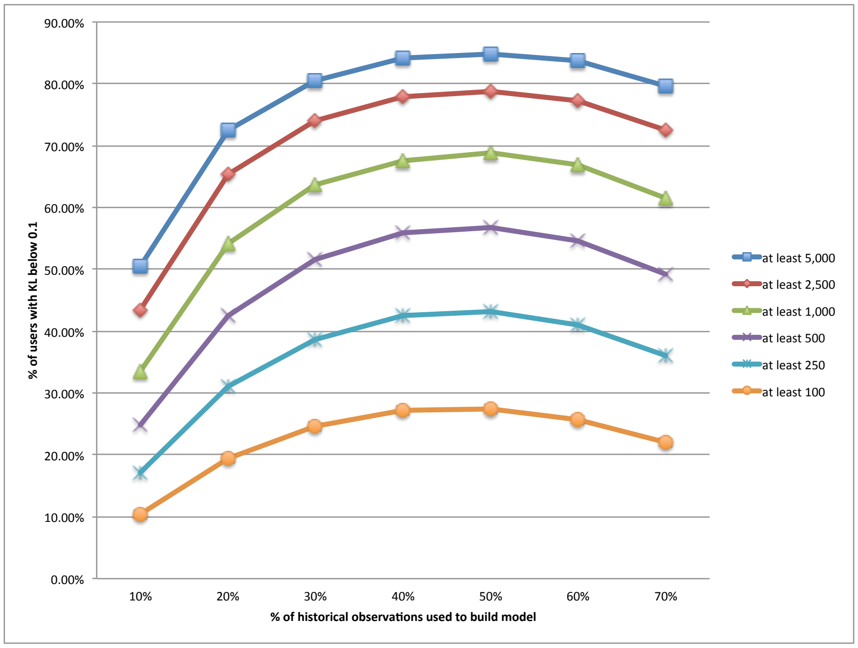

Fig. 3. Cohort analysis for fitting the DoW model

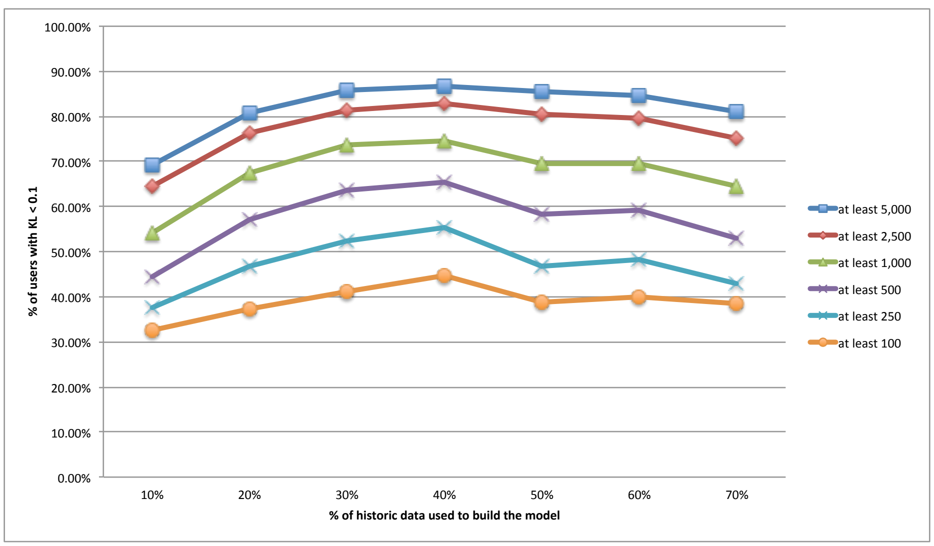

Fig. 4. Fitting distributions of posting behavior for WtD model

a period within the day), but we do not yet differentiate between specific days at larger scale (e.,g, particular day(s) of the month) or special days, such as holidays or historic events (such as during a natural disasters) which should be treated specially. Also, we didn't yet investigate high levels of granularity (e.g., weeks within a month or months within a year) or much lower levels (e.g, minutes within a day).

Therefore, in the future we intend to build more sophisticated models that provide deeper insight through various levels of temporal granularity. For instance, we may model day of the week, within the day, day of the month, and month of the year, all at once, and incorporate them into a single model. However, since we have to incorporate time series data at different levels, some of which may be parse, this is a statistical challenge to model. Further, this model should take into account the special days and model the behavior accordingly. However, these future enhancements could help build a more robust individual life pattern to provide deeper insights.

V. Conclusion

In this paper we demonstrated how to build models of user posting behavior on the micro-blogging service Twitter. This type of analysis is useful for modeling an individual life pattern for a user which can then be used for applications such as determining when someone might be sick, as deviations from this pattern become obvious. Our contributions in this paper were two-fold. First, we demonstrated a probabilistic model for posting behavior, modeling the posting time as a distribution. Second, we performed an empirical study to understand how well the model explains the user's behavior and also studying how the amount of data used to build the models effects its overall performance.

References

[1] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors," in *Proceedings of the 19th International Conference on World Wide Web*, 2010.

[2] S. A. Macskassy and M. Michelson, "Why do people retweet? anti-homophily wins the day!" in *Proceedings of the International Conference on Weblogs and Social Media*, 2011.

[3] D. Scanfeld, V. Scanfeld, and E. L. Larson, "Dissemination of health information through social networks: Twitter and antibiotics," *American Journal of Infection Control*, vol. 38, no. 3, pp. 182 – 188, 2010.

[4] C. Chew and G. Eysenbach, "Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak," *PLoS ONE*, vol. 5, no. 11, p. e14118, 11 2010.

[5] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: A content-based approach to geo-locating twitter users," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 2010.

[6] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 03 1951. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177729694