Classifying Twitter Favorites: Like, Bookmark, or Thanks?

Genevieve Gorrell and Kalina Bontcheva
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP, United Kingdom. E-mail: g.gorrell@sheffield.ac.uk; k.bontcheva@sheffield.ac.uk

Since its foundation in 2006, Twitter has enjoyed a meteoric rise in popularity, currently boasting over 500 million users. Its short text nature means that the service is open to a variety of different usage patterns, which have evolved rapidly in terms of user base and utilization. Prior work has categorized Twitter users, as well as studied the use of lists and re-tweets and how these can be used to infer user profiles and interests. The focus of this article is on studying why and how Twitter users mark tweets as "favorites"—a functionality with currently poorly understood usage, but strong relevance for personalization and information access applications. Firstly, manual analysis and classification are carried out on a randomly chosen set of favorited tweets, which reveal different approaches to using this functionality (i.e., bookmarks, thanks, like, conversational, and self-promotion). Secondly, an automatic favorites classification approach is proposed, based on the categories established in the previous step. Our machine learning experiments demonstrate a high degree of success in matching human judgments in classifying favorites according to usage type. In conclusion, we discuss the purposes to which these data could be put, in the context of identifying users' patterns of interests.

Introduction

User-generated media (UGMs) facilitate the creation and sharing of content by users (e.g., Facebook, LinkedIn, and Twitter). Due to their widespread adoption, they have become an important social phenomenon. Three main types of UGM have evolved so far and can be categorized as follows:

- Interest-graph media (Ravikant & Rifkin, 2010) encourage users to form connections with others based on shared interests, regardless of whether they personally know the other user. They aim to provide the information to users that they will find most interesting. Twitter, a microblogging service in which users share short status updates, encourages this model. The "following" relationship is often one-way;
- Social-graph media (e.g., Facebook) encourage users to connect primarily with people they have real-life relationships with. Typically, short status updates are shared, either written by users themselves or linking to content of interest on the Internet. Friends have the option to "like" the status update and/or comment on it;
- Professional networking services (PNS), such as LinkedIn, address the work context, where connections are implicit professional endorsements, and it is also possible to recommend users and explicitly endorse their skills (Skeels & Grudin, 2009).

The focus of our studies is on the interest graph medium, Twitter. Previous work has studied the kinds of information exchanged (e.g., Ehrlich & Shami, 2010; Naaman, Boase, & Lai, 2010; Zhao & Rosson, 2009), the capacity in which users post on Twitter (e.g., personal vs. professional [Bontcheva, Gorrell, & Wessels, 2013b], information diffusion patterns [e.g., re-tweet vs. follower networks, Kwak, Lee, Park, & Moon, 2010]), and Twitter social networks (e.g., Huberman, Romero, & Wu, 2008; Kwak et al., 2010). Previous work has also studied re-tweets and users' own tweets as readily available data from which to automatically build models of user interests and tweet relevance (e.g., Abel, Gao, Houben, & Tao, 2011b), which are then used for personalized tweet recommendation (e.g., Yan, Lapata, & Li, 2012; Chen, Nairn, Nelson, Bernstein, & Chi, 2010). However, re-tweets are not only sparse data (only 6% of tweets get re-tweeted [Sysomos Inc, 2010]), but also only account for certain kinds of tweet relevance and interestingness (Rout, Bontcheva, & Hepple, 2013).

One little-studied, complementary, and increasingly important source of data on implicit user interests and tweet relevance comes from the favorites functionality. In particular, recent Twitter statistics show that mobile Twitter users are 76% more likely to favorite and 66% more likely to

Received January 13, 2014; revised April 22, 2014; accepted May 27, 2014

 DOI: 10.1002/asi.23352

re-tweet (Schreiner, 2013). Favoriting a tweet, which is done in Twitter by selecting "Favorite" via the web interface, or using a star icon or similar in other Twitter clients, results in that tweet appearing in the list of favorites for that user. Furthermore, the number of times a particular tweet has been favorited is made available, in a similar manner by which it is possible to see how many times a particular tweet has been re-tweeted. Analogous functionality in Facebook might be considered to be the "like" button. "Likes" are counted and made available on the post in a similar manner to favorites. On the other hand, researchers have suggested that Twitter users use favoriting as a way of maintaining a list of Tweets on their profile; the "bookmark" approach (Kwak, Chun, & Moon, 2011).

On this basis, we hypothesize that "bookmarking" users will have favorites lists that make sense out of context, since they favorite a tweet for later reference, whereas "liking" users may have favorites lists comprising many tweets that are not intelligible or useful out of the context of a particular interaction; for example, a comment such as "yeah, right" requires conversational context to interpret and is of little interest without it.

This observation raises the question of whether there are other ways in which use of favoriting differs between groups of Twitter users, and how this impacts other dimensions of Twitter use. Moreover, being able to distinguish automatically between different usages of favorites will improve the quality of user models derived automatically from UGM (e.g., Abel, Gao, Houben, & Tao, 2011a; Angeletou, Rowe, & Alani, 2011), as well as the performance of methods for personalized tweet recommendation (e.g., Abel et al., 2011b; Chen et al., 2010; Chen, Nairn, & Chi, 2011), and tweet summarization (e.g., Harabagiu & Hickl, 2011; Yan et al., 2012).

The first contribution of this work lies in identifying five categories of favorites usage (i.e., like, bookmark, thanks, conversational, and self-promotion), three of which have not been studied in related work. Moreover, this adds a new dimension of classifying Twitter users, because their favoriting behavior is consistent over time, that is, some users are likers, whereas others are bookmarkers. Another important contribution is an automatic method for favorites classification, which demonstrates a high degree of success in automatically matching human judgments in classifying favorites according to usage type. The Conclusion discusses the purposes to which these data could be put, in the context of identifying users' patterns of interests.

Related Work

Very little research has thus far been done on the ways in which Twitter users make use of the favorites functionality. Blau and Neuthal (2012) make use of favorites data as an activity measure and as a measure of gratification in their investigation into Twitter use. Kwak et al. (2011) use favorites data as an indicator of a tweet's quality, and the depth of interest the favoriter has in the tweet. They contrast favoriting and re-tweeting as ways of expressing appreciation, but with slightly different emphases. They suggest that favorites are typically used for personal reference (i.e., as bookmarks), whereas a re-tweet is a broadcast. In contrast, the study reported in this article is a much more in-depth study of the use of favorites in Twitter, coupled with a machine learning-based automatic classification method.

Other relevant research is on *categorizing Twitter users based on the content they post*. For instance, Ehrlich and Shami (2010) found that over 25% of tweets in their sample were directed posts, making Twitter into a short, public message service, in addition to a means of sharing status updates (11%) and information (29%). Zhao and Rosson (2009) emphasize the "water cooler conversation" angle, suggesting people tweet within cliques. With respect to message types, Naaman et al. (2010) found over 40% of their sample of tweets were "me now" messages; that is, posts by a user describing what they are currently doing. Next most common were statements and random thoughts, opinions, and complaints, and information sharing such as links, each taking over 20% of the total. Less common tweet themes were self-promotion, questions to followers, presence maintenance such as, "I'm back," anecdotes about oneself and anecdotes about another. Messages posted from mobile devices are more likely to be "me now" messages (51%). Females post more "me now" messages than males. A relatively small number of people undertake information sharing as a major activity; users can be grouped into "informers" and "meformers," where meformers mostly share information about themselves. Informers and meformers differ in various ways. Informers tend to be more conversational and have more contacts.

A second, complementary dimension of user classification is *purpose of Twitter use*. In earlier research, Bontcheva et al. (2013b) find that 38% of tweeters use Twitter both in a personal and professional capacity, 38% for personal use only and 24% for professional use only, making usage fairly balanced. 51.5% use Twitter to follow the status updates of friends, family, and celebrities. Forty-eight percent use it to get professional information from colleagues, 46.6% get news updates via Twitter, and 34.4% converse with friends. It was found that 15.2% ask questions and get help via Twitter. Those who use Twitter for personal use are more likely to follow friends, family, and celebrities (63.04% as opposed to 36.61%). Those who use Twitter for professional purposes are more likely to use Twitter to get professional information (73.21% as opposed to 42.75%).

Bontcheva et al. (2013b) also found evidence of different models of Twitter use, with self-publicizers, who are more likely to be "meformers" (Naaman et al., 2010), being only one possibility. Experts such as professional bloggers may be classic "informers" (Naaman et al., 2010), primarily attracting followers through the dissemination of interesting information, although these are rarer. Another type of professional use may take the form of workplace or workgroup cliques, where Twitter is used as the medium of communication within the group.

DOI: 10.1002/asi

Our research is complementary to both content-based and purpose-based user categorization, in that it adds a new dimension by categorizing users according to their favoriting behavior (e.g., likers, bookmarkers). Our analysis shows that this behavior is consistent over time and can thus be learnt and used in automatically derived user models.

Theoretical Framework

In devising a taxonomy of different motivations for favoriting, two approaches are possible; asking users their reason for favoriting particular tweets, or focusing on distinctions that are apparent from the data. For example, where a user favorites a link to a blog post, is his intention to publicize the blog post, use the blog post as a way of indicating something about himself, that is, that he likes or approves of it, or to make a note of the blog post for his own future reference? Some users may have a clear intention in this regard, whereas others may favorite the post for a mixture of these reasons, or for one of the reasons but at the same time being aware of the other benefits.

Asking users why they favored a particular post will yield additional information. However, it also limits the quantity of data that can be collected and analyzed. Therefore, our approach is to focus on distinctions that are apparent from the data, without consulting the user. This allows us to analyze a much larger number of tweets, which is required in order to reach reliable conclusions, as well as to develop and evaluate an automated approach to classification. This is also in keeping with previous work on classifying tweet message content (Naaman et al., 2010).

A grounded theory approach (e.g., Corbin & Strauss, 2008) was used to derive favorites usage patterns with reference to a corpus of data. Although a review of the literature, as described in the Introduction, predicted that certain usage patterns would be likely to emerge, that is, "like" and "bookmark" patterns along with some way of using twitter conversationally, in analyzing the data we hoped to be led primarily by a good fit to observation and a comprehensive, reliable, and repeatable fit to the data. At the same time, categorization needed to be sufficiently specific to be useful and interesting. We also needed to acknowledge that Twitter usage patterns have been shown to be evolving rapidly, and therefore previous literature does not necessarily apply to a current Twitter snapshot. Nonetheless, categories derived from the literature, specifically "like," "bookmark," and "conversational," provided information for an initial, theoretical coding, with other categories emerging as required.

Analysis was at first performed by one researcher, who expanded the initial coding as necessary to comprehensively accommodate observation. Following from this stage, six categories were identified, including the five that we will outline, plus a convenience category for tweets not entirely in English, which would later be excluded. A second researcher also reviewed the categories and raised a question on the value of distinguishing "like" from "conversational," because separating the two depends in many cases on knowing whether a user is personally acquainted with people mentioned in the tweet. This is sometimes impossible to deduce from data alone. However, it was decided to retain both categories, because although challenging from an annotation point of view, the two usage patterns indicate very different intent, and are interesting both sociologically and from the point of view of using the work to model users.

Therefore, the following five favorite categories were selected and validated with other annotators (see Schema Validation section):

- "**Like**"—the like category is so called because usage resembles the "like" functionality in Facebook. It is the largest and most generic category, and might be considered the "default" usage pattern. A liked tweet will contain content that the favoriter appreciated, rather than a reference to content elsewhere such as a link (e.g., a tweet with text "Taurus, Leo, Scorpio and Aquarius instantly know what they like and dont like. If they choose you, they want you." is authored by User456 and favorited by FavUser456[1]).
- "**Conversational**"—in many cases, a user will favorite a tweet as part of a conversation (e.g., a tweet with text "@FavUser123 Do you snore? I have to go." is authored by UserXYZ and favorited by FavUser123, who is mentioned explicitly in the text). In other words, if another user directs a tweet towards them, by naming them in the tweet, the user may choose to favorite that tweet rather than send a message in response. The appearance is of using favorites functionality to communicate with another user, since Twitter will notify the tweet author that his tweet has been favorited. In cases where the favoriter is not mentioned explicitly in the tweet body, conversational favorites can be hard to distinguish from "likes."
- Self-promotion, or "**selfpro**"—indicating the use of favoriting to promote oneself, this refers to the case in which a user favorites a tweet that he authored himself. The exception is where the user favorites a tweet that he authored himself, but that is clearly intended to promote another. An example might be a quote by an admired person. This would not be considered a self-promotion but a like, or a bookmark (see following bullet point). A self-promoting favorite does not refer to a tweet authored by another user in which the favoriter is mentioned favorably. This is a "thanks" (see following).
- "**Bookmark**"—in the case of "bookmark," a tweet is favorited in which attention is drawn to a third party, by describing or linking to content relating to them. Users might favorite such a tweet as a reminder to themselves, as a way of expressing their interests on their profile, or as a way of promoting the third party. In practice it is impossible to distinguish these without asking the favoriter, so this class covers all such usage. An example might be "Skeleton: Beautiful Boilerplate for Responsive, Mobile-Friendly Development http://t.co/r8emoQxD," in which a software product is advertised.
- "**Thanks**"—as mentioned earlier, a user may favorite a tweet in which they are mentioned favorably (e.g., a tweet with text "So excited to be following @ThanksUser123" is authored by a user and then favorited by ThanksUser123). A user might do this as a form of self-promotion, or to thank the author for his positive comment. In practice it is impossible to

[1]All user IDs have been anonymized.

DOI: 10.1002/asi

distinguish between these, without asking the favoriter, so therefore this category captures both uses.

As outlined earlier with respect to specific categories, we have in several cases conflated different reasons why the user might have favorited a particular tweet. In "bookmark," for example, promoting external content, using favoriting to convey one's personal interest profile, and noting the external content for future personal reference are combined into one category. Use of the name "bookmark" for this category does not imply that the only purpose was to make a note of the content for future reference. The name "bookmark" has been chosen for this category with reference to earlier research on this usage pattern rather than an attempt to fully describe the reasons why a favoriter might adopt that usage pattern. Similarly there are different reasons why a user might use the "like" favoriting pattern, and we do not attempt to distinguish them. The name "like" is chosen for this category in reference to a "Facebook"-style usage pattern.

Finding a balance between choosing concepts that are reliably able to be identified and still retaining concepts of interest and value is a central dilemma in grounded theory. The categories we have identified are of interest for several reasons. From a sociological perspective, they create the possibility to track evolving social media usage patterns over time. For example, "like" emerges from a paradigm created by Facebook, and usage in Twitter might be found to relate to the popularity of Facebook. "Conversational" usage may relate to the increasing use of Twitter as a messaging service. "Thanks" may interact with the use of Twitter as a professional outreach platform. From the point of view of modeling users with the intention of creating new technology, these categories also have potential. "Thanks" and "selfpro" might relate to professional usage, and could be used to differentiate between broadcasters and consumers, who have different needs. "Bookmark" usage, potentially relating to "informers" rather than "meformers," might be used to help identify material of wider interest in tracking informational trends.

Schema Validation

The next step in validating the taxonomy involved determining the extent to which classification is repeatable by other annotators. A random sample of 688 favorited tweets, from 10 favoriters, was selected on the basis of the time they were tweeted in order to validate the schema[2]. A time-based selection tends to avoid many problems with selection bias, although there may to varying extents be topic biases associated with certain times of the day or week. In our case, little evidence of time-based topic bias was evident. For each favoriter, three human annotators were allocated at random, who each annotated the favorites, in order for us to be able to estimate interannotator agreement (IAA) (also known as intercoder reliability). The ability of independent human annotators to agree on a classification for a tweet is taken as an indicator that the proposed categorization is reliable and the work is repeatable.

In keeping with best practice in natural language processing (Stede & Huang, 2012), annotation guidelines were prepared for the annotators[3], similar in structure and detail to those used by the Linguistic Data Consortium (2005) for annotating entities and Prasad et al. (2007) for annotating discourse. The principle behind the use of written annotation guidelines is that all annotators receive the same task explanation and the potential for differences in their understanding is reduced. We also make the annotation guidelines public, in order to ensure repeatability and consistency of future work. We prepared the guidelines, redrafted based on feedback from one of the annotators, and thereafter served as the source of guidance for all other annotators, although on one occasion a verbal query was responded to. The guidelines contain definitions of the five categories of favorites, positive examples of each category, as well as negative examples of when one category is confused for another. The grounding of the five categories in previous work was discussed in earlier sections.

GATE Teamware (Bontcheva, Cunningham, Roberts, & Tablan, 2010), a web-based, open-source text annotation tool was used by annotators to categorize manually all favorited tweets according to the five categories defined previously. Figure 1 shows a screenshot of the annotation interface, where the favoriter is highlighted in blue, the tweet text in yellow, etc.

Six annotators in total were used, of which two annotated each favoriter's favorites while the remaining four annotators divided the work between them. Observed agreement was found to be 0.92, with a Cohen's kappa (Cohen, 1960) and Scott's pi (Scott, 1955) of 0.83. Although there is a debate regarding the implications of particular kappas, it has been suggested that a figure of 0.83 indicates "almost perfect" (Landis & Koch, 1977) or "excellent" agreement (Fleiss, 1981). Note that the agreement between expert human annotators also provides a guide as to the maximum result that can reasonably be expected from automatic classification.

Research Questions

Having demonstrated the workability of the schema, the rest of the article investigates the following hypotheses:

1. Favorites functionality serves different purposes for different groups of users. It is possible to identify the purpose to which a user is putting the favorites functionality, and patterns reflect different characteristics of Twitter users in terms of the uses to which they put the service;

[2]The data set can be obtained by contacting the first author, Genevieve Gorrell.

[3]Available from http://www.dcs.shef.ac.uk/~genevieve/annotation-manual.pdf

DOI: 10.1002/asi

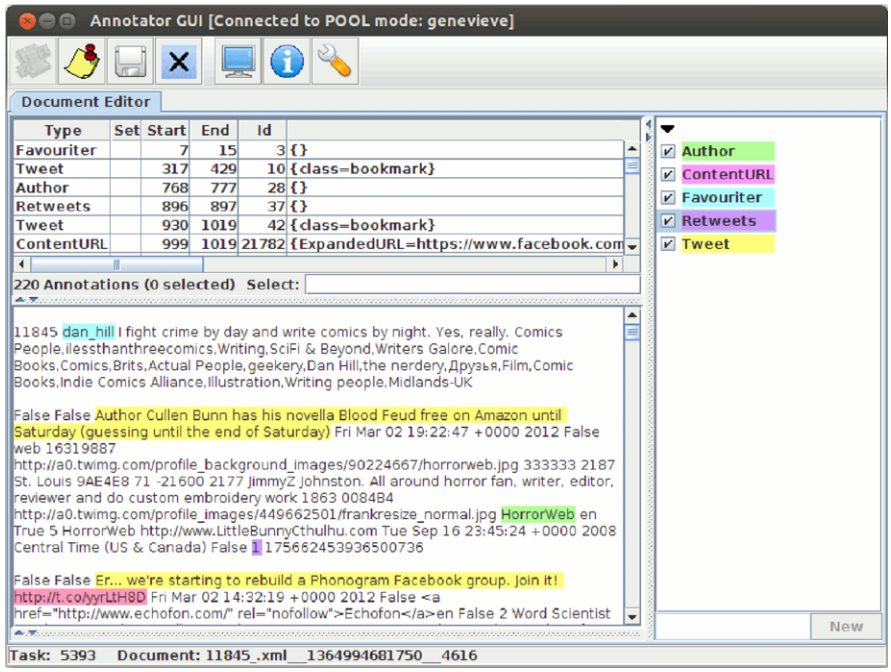

FIG. 1. Teamware annotation interface. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

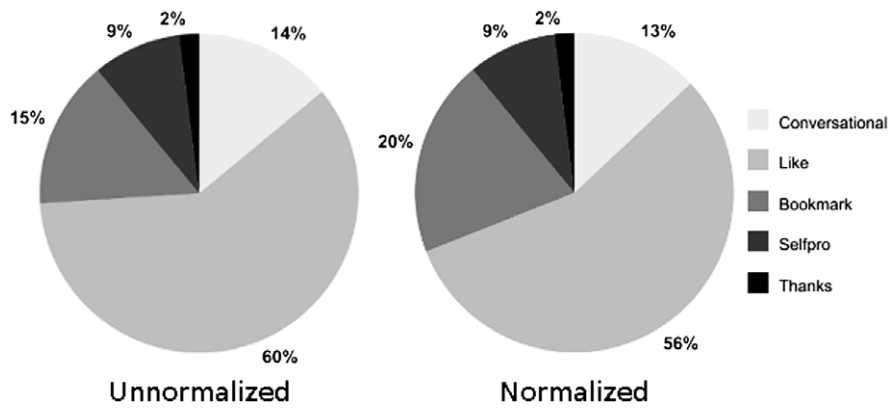

FIG. 2. Favorites by type.

2. Automatic classification of favorites types is possible, with high accuracy.

Data Analysis

A sample of the complete favorites list of a random set of 382 users was gathered from Twitter, comprising a total of 15,178 favorited tweets. Having validated the schema and established a high interannotator agreement (as explained earlier), manual annotation on this larger data set was performed by a single expert annotator.

Tweeters were broadly selected for their use of the English language; however many tweeters use multiple languages, so around 16% of non-English favorites were present at this stage. These were excluded for analysis purposes, but included for the machine learning work, where being able to automatically identify non-English favorites is useful. Figure 2 shows the breakdown into different classes following the exclusion of non-English favorites. In more detail, "conversational" favorites account for 14% of the total, "like" for 60%, "bookmark" for 15%, "selfpro" for 9%, and "thanks" for 2%. The first pie chart is un-normalized. In the second pie chart, each tweeter contributes the same size of input, meaning that those tweeters who favorite a lot more than others will not skew the overall statistics. The normalized graph shows a lower proportion allocated to "like" and a higher proportion to "bookmark," indicating that tweeters who "like" tend to favorite more than those who "bookmark."

DOI: 10.1002/asi

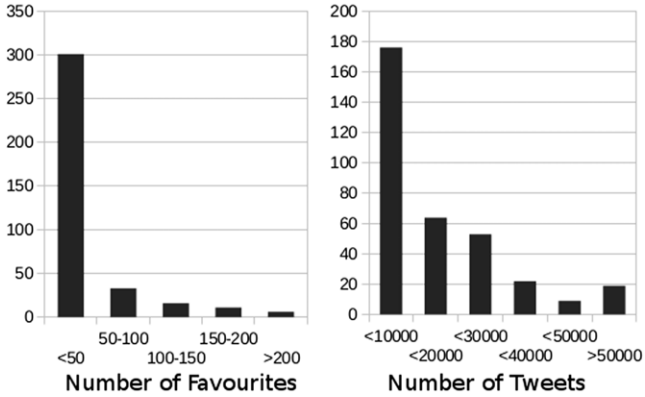

FIG. 3. Number of favorites/tweets.

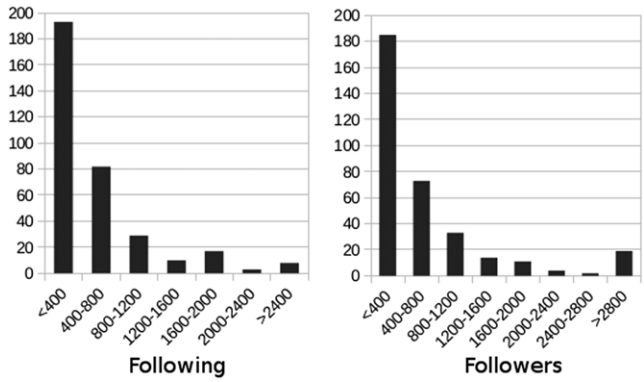

FIG. 4. Number of following/followers.

We also studied how usage of favorites differs according to purpose of Twitter usage. In this case, users were assigned manually into three categories based on the text of their Twitter profiles: personal users, professional users (those who use Twitter in their professional capacity [e.g., journalist] and focus their usage on topics related to that profession), and entities (companies, products, organizations, etc.). The majority (89%) of our sample are personal users, with professionals and entities comprising 7% and 4%, respectively.

Lastly, the frequency of favorites use was examined. Figure 3 shows the numbers of favorites users have, with the majority having fewer than 50. This is contrasted with the number of tweets users have, which is much higher, though again is a distribution skewed to the left, with a long tail of users with larger numbers of tweets. Note that the appearance of a slight tick at the end of this graph is caused by the final category containing all users with more than 50,000 tweets, which is a larger category. We assume that the underlying distribution tails off gradually as ever fewer users have ever larger numbers of tweets.

In contrast, the number of tweeters a person follows and the number by whom they are followed are broadly similar. In Figure 4, again we see long tails, but the majority of users follow and are followed by fewer than 800 users.

Independent samples t-tests showed differences in Twitter/favorites usage between personal and professional users and accounts representing commercial entities. For convenience, entities and professionals are grouped, since they account for only a small number of tweeters and show similar properties. Personal users may tweet more than professionals or entities, $p = 0.041$. Professionals and entities follow more and are followed by more other tweeters, $p < 0.001$ and $p < 0.001$, respectively; a result that remains significant after applying a Bonferroni correction (Welkowitz, Cohen, & Ewen, 2006) over the two conditions and three dependent variables (number of tweets, number of followers, and number of followees), thus reducing α to 0.008.

Personal users may "like" more often, $p = 0.049$, and professionals/entities "thank" more often, $p = 0.001$, $\alpha = 0.005$. For other favorite types, no significant differences were observed between these different types of tweeters.

Pearson's product moment correlation coefficient (Welkowitz et al., 2006) was used to determine relationships between Twitter usage and favoriting preferences. There is some indication that users who tweet more have more "conversational" favorites, $p = 0.019$, suggesting that this is the more prolific type. Applying a Bonferroni correction across the five favorites types reduces α to 0.01, rendering the correlation of less significance; however the Bonferroni correction is conservative and the result is still of interest.

Use of conversational favorites tends to correlate inversely with "bookmark," "like" or "selfpro" favorite types, $p < 0.001$, $p < 0.001$, and $p = 0.024$, respectively. Similarly, use of "like" correlates inversely with other types, $p < 0.001$ in all cases. The result is similar for bookmark, $p < 0.001$ in all cases except for "thanks," and "selfpro," $p < 0.001$ for "like" and "bookmark," $p = 0.024$ for "conversational." Results in most cases are significant even after a Bonferroni correction over the 20 comparisons reduces α to 0.0025. The general pattern is that users favor one type, though significant results are not obtained to the same extent for thanks, perhaps because fewer data are available for this type. Indeed half of all users assign at least 75% of their favorites to one preferred type, further reinforcing the message that favoriters have a tendency not to mix types. This suggests that users may approach Twitter with a particular mindset, or way in which they prefer to use it, from which they do not tend to deviate, perhaps because it pertains to personality factors, or perhaps because Twitter accounts are started with a particular purpose in mind, such as to disseminate a particular type of information. This lends further value to identifying usage patterns, because we learn something relatively consistent about the user, and constitutes a major finding of the work.

DOI: 10.1002/asi

TABLE 1. Results for three systems.

	Accuracy	Cohen's kappa	Scott's pi
SVM	**0.847**	**0.756**	**0.756**
Rule-based Baseline	**0.755**	**0.604**	**0.599**
Everything to "like"	**0.569**	**0**	**–0.159**

TABLE 2. SVM system, raw results.

	Book.	Conv.	Like	Selfpro	Thanks	Noneng	Total
Book.	**1,659**	**15**	**252**	**21**	**0**	**23**	**1,970**
Conv.	**59**	**1,268**	**457**	**7**	**2**	**6**	**1,799**
Like	**250**	**14**	**7,392**	**63**	**1**	**14**	**7,734**
Selfpro	**21**	**4**	**10**	**1,058**	**0**	**5**	**1,098**
Thanks	**11**	**141**	**59**	**0**	**1**	**0**	**212**
Noneng	**231**	**187**	**414**	**51**	**0**	**1,482**	**2,365**

TABLE 3. SVM system, normalized results.

	Book.	Conv.	Like	Selfpro	Thanks	Noneng	Total
Book.	**0.842**	**0.008**	**0.128**	**0.010**	**0**	**0.012**	**1**
Conv.	**0.033**	**0.704**	**0.254**	**0.004**	**0.001**	**0.003**	**1**
Like	**0.032**	**0.002**	**0.956**	**0.008**	**0**	**0.002**	**1**
Selfpro	**0.019**	**0.004**	**0.009**	**0.964**	**0**	**0.005**	**1**
Thanks	**0.052**	**0.665**	**0.278**	**0**	**0.005**	**0**	**1**
Noneng	**0.098**	**0.079**	**0.175**	**0.022**	**0**	**0.627**	**1**

TABLE 4. Baseline system, raw results.

	Book.	Conv.	Like	Selfpro	Thanks	Noneng	Total
Book.	**1,726**	**5**	**148**	**90**	**0**	**0**	**1,969**
Conv.	**173**	**1,154**	**459**	**10**	**0**	**0**	**1,796**
Like	**321**	**11**	**7,300**	**66**	**0**	**0**	**7,698**
Selfpro	**1**	**0**	**10**	**1,087**	**0**	**0**	**1,098**
Thanks	**28**	**125**	**59**	**0**	**0**	**0**	**212**
Noneng	**559**	**429**	**1,236**	**137**	**0**	**0**	**2,361**

Automatic Favorites Classification Method

Support vector machines (SVMs) were used to automatically classify the favorites into different types. SVMs (Cristianini & Shawe-Taylor, 2000) are used frequently in natural language processing tasks, because they are well-suited to this type of task, and typically give a superior performance. A linear kernel was applied in conjunction with a cost function of 0.7 and an uneven margins ratio of 0.4 (see Li & Shawe-Taylor, 2003, for more information about the utility of uneven margins with SVM). This binary classification approach was applied to the multiclass problem of differentiating between the five favorites types, by converting it into five binary classification problems, each of which differentiates between a favorites type and all the other cases. This approach is chosen for its superior efficiency compared with the approach of creating n! classifiers comparing each type with each other.

Features used were: the words contained within the tweet, the part of speech of the words contained within the tweet (created using the part-of-speech tagger from GATE [Cunningham et al., 2011]), whether or not the author of the tweet is also the favoriter, whether the favoriter is mentioned in the tweet, how many times the tweet has been re-tweeted (<5, 5–50 or >50), whether a URL features in the tweet, and the hostname of any URL contained in the tweet.

Fivefold cross-validation was used to ascertain the degree of success in automatically classifying favorites, on the data set of 15,178 manually labeled tweets.

A baseline system was also created that classified tweets according to the following simple rules:

- If the author of the tweet was also the favoriter, then the tweet is a "self-promotion."
- If there is a URL in the tweet then it is a "bookmark."
- If the favoriter is mentioned in the tweet then it is a "conversational."
- Otherwise it is a "like."
- "Noneng" and "thanks" are harder categories to provide simple rules for and were omitted from the baseline.

Results

As shown in Table 1, the accuracy obtained from automatically classifying the tweets according to favorite types using SVM was 0.847. Cohen's kappa was 0.756 and Scott's pi was also 0.756. This is a substantial improvement on the baseline system which produced the following results: accuracy of 0.755, Cohen's kappa of 0.604, and Scott's pi of 0.599. Were the tweets to be all classified as "like" we could expect to see an accuracy of 0.569 but a Cohen's kappa of 0 and a Scott's pi of –0.159, indicating results no better than chance. These figures need to be interpreted in the context of inter-annotator agreement of 0.92 (Cohen's kappa or Scott's pi 0.83). A combined approach was also tried, in which high confidence SVM results were used to override default classifications provided by the rule-based baseline system. This resulted in marginal improvements, compared against the simple SVM approach, but the difference was not statistically significant.

Confusion matrices are shown for the SVM system in Table 2, normalized in Table 3, and for the baseline in Tables 4 and 5 (normalized) for comparison. Normalization is achieved by dividing each figure by the total number of actual instances for that category (not the total number of instances that the system hypothesized) to present a figure that indicates the proportion of the instances for that category that were allocated to each category by the system.

In absolute terms, the largest numbers of misclassifications come from three factors. First, "like" being the dominant type, there is a tendency to misclassify other types as "like" and a tendency to misclassify "like" as other types.

DOI: 10.1002/asi

TABLE 5. Baseline system, normalized results.

	Book.	Conv.	Like	Selfpro	Thanks	Noneng	Total
Book.	0.877	0.003	0.075	0.0457	0	0	1
Conv.	0.096	0.643	0.256	0.006	0	0	1
Like	0.042	0.001	0.948	0.009	0	0	1
Selfpro	0.001	0	0.009	0.990	0	0	1
Thanks	0.132	0.590	0.278	0	0	0	1
Noneng	0.237	0.182	0.524	0.058	0	0	1

Second, "noneng" is difficult to classify. Recall that the noneng category contains all the tweets not in English, and as described in the annotation manual[4], even tweets only partially in English. Distinguishing such tweets essentially requires memorizing which words are and are not English words; a task beyond the scope of the machine learner. Third, "thanks" and "conversational" are highly confusable, and since "conversational" is the dominant type, "thanks" tends to be misclassified as "like."

After normalization it becomes clearer that the first two sources of error are large in numerical terms but not indicative of real type confusion, whereas "thanks" poses a real problem for classification, being simply too hard. Distinguishing "thanks" from "conversational" or even "like" often requires quite complex human judgment. Since "thanks" is a small class, depending on the application we might choose to amalgamate it with "conversational." In terms of maximizing overall performance, however, the greatest gains will be had in further improving performance with regards to "bookmark," "conversational," "like," and "noneng." The latter could be achieved by first pre-filtering all tweets with a tweet language identification algorithm (e.g., Derczynski, Maynard, Aswani, & Bontcheva, 2013) and then only classifying the favorites of the English tweets. We plan to address this in future work.

Summary and Discussion

Our data analysis revealed five different patterns of favorites usage, which are reliably distinguished by human annotators. This diversity differentiates interest graph UGM from social networks and their "like" functionality. An important result is that Twitter users tend to prefer only one of the five approaches when it comes to favoriting. Users who are likers will tend to just use favorites this way; similarly bookmarkers will show a tendency to only bookmark and so on. Some variation from the pattern is shown but the strong tendency is to prefer one approach to favorites usage only.

The "like" approach tends to be a feature of personal Twitter usage, with professionals and entities using this to a much lesser degree. Professionals and entities are more likely to use "thanks," as indeed you might expect where an account exists primarily to promote a product.

[4]Available from http://www.dcs.shef.ac.uk/~genevieve/annotation-manual.pdf

With respect to message usage patterns, our results support the findings of Ehrlich and Shami (2010), that is, that a quarter of Twitter usage now comprises directed messages. Moreover, a new insight of our work is to show that this evolution in usage also extends to favorites functionality, which has been subverted to serve this purpose. Conversational favorites comprise 14% of the total.

With regards to automatic classification, the SVM-based method proposed here achieves an accuracy of 0.85, which is high enough to allow the favorites categories to be used as a feature in modelling UGM users, tweet recommendation and summarization.

In ongoing work, we are investigating which favorite types are most useful in determining user interests. Preliminary data analysis has indicated that "bookmark" is the most useful favorite category for this purpose, and that possibly the content of the page linked to is of value, whereas "conversational" may be less useful.

With regards to improving automatic classification performance of favorites and bringing it closer to the 0.92 human levels, we plan to experiment first with classifying users as personal or professional/entity, based on the text of their Twitter profiles. This information could then be used as an input feature to our SVM classifier. The rationale here stems from our finding that personal users tend to use favorites differently from professional or entity ones.

Additionally, the finding that users tend to prefer one particular approach to favoriting will be used to improve performance. This knowledge could be integrated into the automatic classification approach in several ways. First, we will experiment with temporally-based classifiers such as some neural net approaches, which make use of the previous n classifications in allocating future examples. This would also accommodate preference drift over time. The second strand of experiments will include approaches such as conditional random fields, which also take neighboring data points into consideration.

Acknowledgments

This work was partially supported by the European Union under grant agreements No. 287863 TrendMiner (http://www.trendminer-project.eu) and the UK EPSRC grant No. EP/I004327/1.

References

Abel, F., Gao, Q., Houben, G.J., & Tao, K. (2011a). Semantic enrichment of twitter posts for user profile construction on the social web. In Antoniou, Grobelnik, Simperl, Parsia, Plexousakis, Leenheer and Pan (Eds.), The semanic web: Research and applications (pp. 375–389). Berlin Heidelberg: Springer.

Abel, F., Gao, Q., Houben, G.J., & Tao, K. (2011b). Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In Proceedings of the 3rd International Web Science Conference (pp. 2–10), Koblenz, Germany. New York: ACM.

Angeletou, S., Rowe, M., & Alani, H. (2011). Modelling and analysis of user behaviour in online communities. In Proceedings of the 10th

DOI: 10.1002/asi

International Semantic Web Conference (ISWC) 2011 (pp. 35–50), Bonn, Germany. Berlin Heidelberg: Springer.

Blau, I., & Neuthal, T. (2012). Tweeting educational technology: A tale of professional community in practice. International Journal of Cyber Society and Education, 5(1), 75–80.

Bontcheva, K., Cunningham, H., Roberts, I., & Tablan, V. (2010). Web-based collaborative corpus annotation: Requirements and a framework implementation. In Witte, Cunningham, Patrick, Beisswanger, Buyko, Hahn, Verspoor and Coden (Eds.), Proceedings of the Workshop New Challenges for NLP Frameworks (pp. 20–27). Valletta, Malta: ELRA.

Bontcheva, K., Gorrell, G., & Wessels, B. (2013b). Social Media and Information Overload: Survey Results. arXiv:1306.0813 [cs.SI] Retrieved from http://arxiv.org/abs/1306.0813

Chen, J., Nairn, R., Nelson, L., Bernstein, M., & Chi, E. (2010). Short and tweet: experiments on recommending content from information streams. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1185–1194). New York: ACM.

Chen, J., Nairn, R., & Chi, E. (2011). Speak little and well: recommending conversations in online social streams. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 217–226). New York: ACM.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.

Corbin, J., & Strauss, A. (Eds.). (2008). Basics of qualitative research: Techniques and procedures for developing grounded theory. Thousand Oaks, California: Sage.

Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines. Cambridge, UK: Cambridge University Press.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., et al. (2011). Text processing with GATE (Version 6). Sheffield, UK: University of Sheffield.

Derczynski, L., Maynard, D., Aswani, N., & Bontcheva, K. (2013). Microblog-genre noise and impact on semantic annotation accuracy. In Proceedings of the 24th ACM Conference on Hypertext and Social Media (pp. 21–30). New York: ACM.

Ehrlich, K., & Shami, N.S. (2010). Microblogging inside and outside the workplace. In Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM) (pp. 42–49), Washington, DC. Menlo Park, CA: AAAI Press.

Fleiss, J.L. (1981). Statistical methods for rates and proportions (2nd ed., pp. 212–225). New York, NY: John Wiley and Sons.

Harabagiu, S.M., & Hickl, A. (2011). Relevance Modeling for Microblog Summarization. In Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM) (pp. 514–517), Barcelona, Spain. Menlo Park, CA: AAAI Press.

Huberman, B.A., Romero, D.M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. arXiv preprint arXiv:0812.1045.

Kwak, H., Chun, H., & Moon, S. (2011). Fragile online relationship: a first look at unfollow dynamics in twitter. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1091–1100). New York: ACM.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In Proceedings of the 19th International Conference on World Wide Web (pp. 591–600). New York: ACM.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

Li, Y., & Shawe-Taylor, J. (2003). The SVM with uneven margins and Chinese document categorization. In Proceedings of The 17th Pacific Asia Conference on Language, Information and Computation (PACLIC17) (pp. 216–227). Cambridge, MA: MIT Press.

Linguistic Data Consortium (2005). Annotation Guidelines for Entity Detection and Tracking (EDT). Retrieved from http://catalog.ldc.upenn.edu/docs/LDC2005T09/guidelines/EnglishEDTV4-2-6.PDF

Naaman, M., Boase, J., & Lai, C.H. (2010). Is it really about me?: message content in social awareness streams. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (pp. 189–192). New York: ACM.

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., & Webber, B.L. (2007). The penn discourse treebank 2.0 annotation manual.

Ravikant, N., & Rifkin, A. (2010). "Why Twitter Is Massively Undervalued Compared To Facebook". TechCrunch, 2010. Retrieved from http://techcrunch.com/2010/10/16/why-twitter-is-massively-undervalued-compared-to-facebook/

Rout, D., Bontcheva, K., & Hepple, M. (2013). Reliably evaluating summaries of twitter timelines. In Proceedings of the AAAI Workshop on Analyzing Microtext (pp. 64–71). Palo Alto, CA: AAAI Press.

Schreiner, T. (2013). New Compete study: Primary mobile users on Twitter. 11 Feb 2013. Retrieved from https://blog.twitter.com/2013/new-compete-study-primary-mobile-users-on-twitter

Scott, W.A. (1955). Reliability of content analysis: The case of nominal scale coding. Public Opinion Quarterly, 19(3), 321–325.

Skeels, M.M., & Grudin, J. (2009). When social networks cross boundaries: A case study of workplace use of facebook and linkedin. In Proceedings of the ACM 2009 International Conference on Supporting Group Work (pp. 95–104). New York: ACM.

Stede, M., & Huang, C.-R. (2012). Interoperability and reusability: The science of annotation. Language Resources and Evaluation, 46(1), 91–94. doi: 10.1007/s10579-011-9164-x

Sysomos Inc (2010). Replies and Retweets on Twitter. Retrieved from https://www.sysomos.com/insidetwitter/engagement/

Welkowitz, J., Cohen, B.H., & Ewen, R.B. (2006). Introductory statistics for the behavioral sciences. Hoboken, NJ: John Wiley & Sons.

Yan, R., Lapata, M., & Li, X. (2012). Tweet recommendation with graph co-ranking. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers (Vol. 1, pp. 516–525). Jeju, Korea: Association for Computational Linguistics.

Zhao, D., & Rosson, M.B. (2009). How and why people Twitter: The role that micro-blogging plays in informal communication at work. In Proceedings of the ACM 2009 International Conference on Supporting Group Work (pp. 243–252). New York: ACM.