Extracting Human Temporal Orientation from Facebook Language

H. Andrew Schwartz,[1,4] **Gregory J. Park,**[2] **Maarten Sap,**[2] **Evan Weingarten**[3], **Johannes Eichstaedt,**[2]
Margaret L. Kern,[5] **David Stillwell,**[6] **Michal Kosinski,**[7] **Jonah Berger,**[3] **Martin Seligman,**[2] **Lyle H. Ungar**[1]
[1]Computer & Information Science, [2]Psychology, [3]Wharton, University of Pennsylvania
[4]Computer Science, Stony Brook University [5]Graduate School of Education, University of Melbourne
[6]Psychometrics Centre, Cambridge University [7]Graduate School of Business, Stanford University
hansens@seas.upenn.edu, gregpark@sas.upenn.edu

Abstract

People vary widely in their temporal orientation—how often they emphasize the past, present, and future—and this affects their finances, health, and happiness. Traditionally, temporal orientation has been assessed by self-report questionnaires. In this paper, we develop a novel behavior-based assessment using human language on Facebook. We first create a past, present, and future message classifier, engineering features and evaluating a variety of classification techniques. Our message classifier achieves an accuracy of 71.8%, compared with 52.8% from the most frequent class and 58.6% from a model based entirely on time expression features. We quantify a users' overall temporal orientation based on their distribution of messages and validate it against known human correlates: conscientiousness, age, and gender. We then explore social scientific questions, finding novel associations with the factors openness to experience, satisfaction with life, depression, IQ, and one's number of friends. Further, demonstrating how one can track orientation over time, we find differences in future orientation around birthdays.

1 Introduction

How much one emphasizes the past, present, or future is predictive of many human factors such as occupational and educational success, engagement in risky behavior, financial stability, depression, and health (Boyd and Zimbardo, 2005; Zimbardo and Boyd, 1999). However, studies on the human experience of time are filled with diverse measurement methods (Strathman and Joireman, 2005), mostly involving questionnaires which are expensive to administer multiple times or at scale and can be subject to confounds when compared to other questionnaire based assessments.

Text mining and language processing techniques can provide a more objective and scalable measurement of *temporal orientation*, one's tendency to emphasize the past, present, or future. Whereas most prior computational linguistics and text mining temporal studies have focused on events, there has been a lack of work looking at the temporal orientation of people. Such measures, which were not practical before the growth of social media, can open many avenues of large-scale psychological discovery into the consequences of temporal orientation and yield applications such as targeted marketing, loan repayment forecasting, understanding economic patterns, or even quantified self-help tools to encourage more future-mindedness.

In this paper, we develop a temporal orientation measure based on language in social media. The measure uses a message-level classifier of past, present, and future, aggregated over users to create user-level assessments. We evaluate the message-level classifier over hand annotated data and the derived user-level model against known human correlates of temporal orientation: *conscientiousness*, age, and gender. To the best of our knowledge, this represents the first paper to study a language-based measure of user-level temporal-orientation.

Our contributions include: (a) the introduction of the task of extracting human temporal orientation from their language use, (b) methodological evaluation and feature engineering for the task, and (c) novel social scientific applications and findings. To-

Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, pages 409–419,
Denver, Colorado, May 31 – June 5, 2015.

ward (a) and (b), we find that achieving the task is non-trivial as we build on and diverge from related computational linguistics tasks (e.g. time expression recognition) and utilize a classifier capturing non-linear relationships and interactions. Towards (c), we show how our measure usefully informs psychological theory by relating our human assessments to other psychological variables at a scale not easily explored, and by tracking changes in temporal orientation over time.

2 Background

Researchers and philosophers have long been interested in the subjective experience of time: how individuals relate to their past, are mindful of their present, and envision their futures (James, 1890; Lewin, 1942). Similarly, computational studies have a rich history on extracting temporal relationships beginning decades ago (Allen, 1983). Here, we provide some background on temporal orientation's broader relevance, on computational techniques used to extract temporal information from text, and on related user-level prediction tasks.

Temporal orientation and its correlates. Studies on the human subjective experience of time are filled with diverse measurement methods, varying in their emphasis on cognitive, affective, and/or motivational aspects.[1] Decisions are influenced by the past, present, and mental simulations of possible futures (Seligman et al., 2013).

One widely studied aspect of subjective time is *temporal orientation*, or an individual's tendency to habitually emphasize past, present, or future temporal frames (Boyd and Zimbardo, 2005). Understanding how and why individuals differ in their temporal orientation, can, for example suggest how they can achieve favorable outcomes in areas of life that require substantial long-term planning, including education, higher status occupations, and physical health (Zimbardo and Boyd, 1999; Boyd and Zimbardo, 2005; Steinberg et al., 2009).

Consistent links have been established between temporal orientation and a psychological factor associated with planning, health, and risky behaviors: the personality trait of *conscientiousness*. Conscientious individuals are characterized as self-disciplined, orderly, planful, and reliable (Roberts et al., 2013). Past research has established that highly conscientious people exhibit more future- and less present-oriented (Zimbardo and Boyd, 1999; Webley and Nyhus, 2006; Adams and Nettle, 2009). We use a measure of conscientiousness from the well-established "Big-five" or Five Factor Model of personality (Goldberg, 1990; McCrae and John, 1992). The other four factors, *extraversion* (e.g. active, outgoing, talkative), *agreeableness* (e.g. kind, trusting, generous), *neuroticism* (e.g. touchy, anxious, depressive), and *openness* (e.g. intellectual, artistic, insightful), have been found to have little connection with temporal orientation (Zimbardo and Boyd, 1999).

Other studies have established consistent links between temporal orientation and demographic characteristics. In particular, as one ages they think less about the immediate present and more about the future (Friedman, 2000; Nurmi, 2005; Steinberg et al., 2009), and females tend to think a bit more about the future than males (Keough et al., 1999). However, detailed age trends are not well understood, with studies mostly focusing on adolescents or college-aged students.

For many other important outcomes, such as happiness or well-being, past research leaves us unclear as to the relationship with temporal orientation. Some suggest future-oriented individuals are happier as they engage in more provident behaviors such as saving money and establishing healthier habit (Desmyter and De Raedt, 2012; Diener et al., 2013). This is supported by the connection between future orientation and less depression (Zimbardo and Boyd, 1999). However, others argue that emphasis on the future inhibits ones ability to reflect wisely on the past and savor present experiences (Boniwell et al., 2010). Our study explores this relationship at an unprecedented scale, utilizing the Satisfaction with Life Scale (Diener et al., 1985) and the Center for Epidemiologic Studies Depression Scale, the CES-D (Radloff, 1977). We also look at previously unexplored variables, IQ and number of friends, for which links with temporal orientation seem plausible (e.g. one might suspect it is smart to think about the future, or wonder if one's reflection

[1] For a review, see Strathman et al., 2005.

on the past is related to their popularity as measure by number of friends).

Related work. Studying temporal language is by no means new to the field of computational linguistics (or NLP). Most recently, time annotation has gained greater interest with a successive sequence of three SemEval tasks (TempEval-1, -2 and -3).

The SemEval competitions have provided data sets that facilitate the comparison of different methods for evaluating time expressions, events, and temporal relations (Verhagen et al., 2007; Verhagen et al., 2010; UzZaman et al., 2013). Such research on temporal text analysis generally focuses on determining when events start and end or how they relate temporally to each other; specific goals include information extraction of time-dependent facts from news media (Ling and Weld, 2010; Talukdar et al., 2012), or extracting personal histories in social media (Wen et al., 2013). In contrast, our goal is to find the temporal orientation of people.

Of the numerous TempEval tasks, we build upon those which identify time expressions and resolve their expressed time and date relative to the time of writing (e.g. the time expression 'yesterday' in a document written on *January 15, 2014* is resolved as *January 14, 2014*). Many methods have been used, ranging from hand-crafting rules to machine learning models. Unlike other areas of natural language processing where stochastic techniques dominate, rule-based systems have been quite competitive in time expressions recognition, especially in less domain dependent settings or for relaxed matching tasks (UzZaman et al., 2013).

A number of useful toolkits have been produced for temporal text analysis (Verhagen et al., 2005; Ling and Weld, 2010; Chang and Manning, 2012). In this work, we use Stanford University's rule-based temporal tagger, SUTime, which geve accuracy in line with the state-of-the-art systems at identifying time expressions at TempEval (Chang and Manning, 2012).[2] SUTime, built on top of Stanford's part-of-speech and named entity taggers, labels times, durations, intervals, and relative times compared to the time at which the document was written.

Our work fits a growing tradition of computational work to better understand people based on their online behavior. Much of this type of work uses human properties to better perform traditional computational linguistics tasks, while others focus particularly on predicting user attributes. User network information has been used for tweet summarization or filtering (Panigrahy et al., 2012; Chang et al., 2013; Feng and Wang, 2013).

Others utilize psychological knowledge about people, such as exploiting the human tendency to report more positive extreme feelings than negative in order to improve on sentiment analysis (Guerra et al., 2014). Toward attribute prediction, a large proportion of works have focused on demographics (Argamon et al., 2009; Goswami et al., 2009; Burger et al., 2011; Al Zamal et al., 2012; Bergsma et al., 2013; Sap et al., 2014). and personality prediction (Mairesse et al., 2007; Iacobelli et al., 2011; Schwartz et al., 2013; Park et al., 2015).

Human temporal orientation, as we study it here, differs from previous studies of user attribute prediction in that temporal orientation calls for consideration of additional language features (some more sophisticated, such as time expressions), and exploration of classification techniques (e.g. that can capture non-linear relationships or interactions). We also add multidisciplinary applications, showing not just how accurately our models predict, but also studying how temporal orientation relates to other factors, for example, by weighing in on conflicting literature as to whether people who are more future-oriented are more satisfied with their life.

3 Method

We develop a methodology for measuring a given social media user's temporal orientation. First, we build a classifier to label whether a message discusses the past, present, or future, and then we quantify users' temporal orientation as the percentage of their messages in each category.

We train a variety of supervised classifiers and explore many features in order to label the temporal

[2]Our goals differ slightly from the TempEval accuracy criteria. For example, when SUTime fails to distinguish "one and a half weeks" from "one week", it does not affect our performance. However, other errors, such as confusing the verb 'march' with the month March will harm our accuracy.

class of a social media message. Because this task is new, it is not clear what classification technique is ideal (for example, it is possible that present orientation is best captured with non-linear relationships), so we explore four techniques:

logR: (*logistic regression*). We use regularized logistic regression (equivalent to maximum entropy) (Fan et al., 2008; Bishop, 2006). From cross-validation over the training data, we chose L1 penalization ($\alpha||\beta||_1$).

lSVC, rSVC: (*support vector classification*). Compared to logR, support vector machines offer non-linear kernel functions (Cortes and Vapnik, 1995), and large-margin optimization for class split. We consider both a linear kernel (lSVC) and a radial basis function kernel (rSVC). From cross-validation over the training data, we chose L1 penalization for lSVC and L2 ($\alpha||\beta||^2$) for rSVC.

ERTs: (*forest of extremely randomized trees*). This technique uses an ensemble of decision trees in which both the feature and cut-point are chosen at each node from a randomly generated set of possible options (Geurts et al., 2006). Such an approach can handle both interactions and non-linear relationships, at the expense of a larger search space. From cross-validation over our training data, we set the following algorithm parameters: we build 1,000 decision trees, using the Gini impurity measure when choosing splits (as opposed to entropy), and selecting each node's feature threshold from among square-root of the total features.

All classifications algorithms were implemented using the scikit-learn toolkit (Pedregosa et al., 2011). Multi-classication over binary classifiers (logR, lSVC, rSVC) was achieved using a series of *one-v-rest* classifiers.

We explore five language-based features:

ngrams: *1 to 3 token sequences*. Messages are tokenized using the *happierfuntokenizing* tool[3] which captures common social media tokens such as emoticons, hashtags, and user handles. Features are encoded simply as binary indicators for whether the ngram appears in the message.

time exs: *The mean difference between the resolved date-time of all time expressions and the date-time in which the message was posted.* Time expressions are labeled via Stanford's SUTime annotator (Chang and Manning, 2012), discussed previously. Specific features recorded include the temporal difference itself (e.g. -2.5 for "two and half days ago"), its base 2 log ($log(1 + value)$), its absolute value, total number of time expressions, and binary variables indicating if any past, present, or future expressions appear in the text. We also include binary features for each of the named-entity time tags for the time expression provided by SUTime (e.g. "future_ref", "present_ref", "next_immediate").

POS tags: *The relative frequency of each part-of-speech tag.* Tagging is done via Stanford's part-of-speech tagger (Toutanova et al., 2003). Stanford's tagger does not have explicit social media tags, but we are most interested in capturing tense which it does well.[4] Also, it is already being used as part of SUTime. Each part-of-speech tag is encoded as the frequency of tag usage ($freq(tag, msg)$) divided by the total number of tokens in the message ($tokens_{msg}$):

$$p(tag|msg) = \frac{freq(tag, msg)}{|tokens_{msg}|}$$

lexica: *The relative frequency of categories, based on the Linguistic Inquiry and Word Count (LIWC) dictionary* (Pennebaker et al., 2007). We use the 2007 version of *LIWC* which includes 64 categories of psychologically-relevant language, including past, present, and future verb categories. The features are encoded as the frequency with which a word from a category (cat) appeared in the message (msg) divided by the total tokens in the message ($tokens_{msg}$):

$$p(cat|msg) = \frac{\sum_{token \in cat} freq(token, msg)}{|tokens_{msg}|}$$

[3] available here: wwbp.org/public_data/happierfuntokenizing.zip

[4] The Stanford Tagger has well documented errors on microblog text (Derczynski et al., 2013). However, we manually evaluated 49 verbs across 20 randomly selected statuses, and all verb tenses were correctly tagged while 4 non-verbs were incorrectly tagged as base-form verbs.

Status	R1	R2	R3	Maj
:) today was actually pretty good	pa	pa	pa	pa
is listening to The Sad Cafe by The Eagles!	pr	pr	pr	pr
considering checking out base jumping and parkour some time in the future XP	fu	fu	fu	fu
I just watched Oprah and am posting what it was about.	pa	pr	pa	pa
really wanted a snow day, but probably not going to get one tomorrow. now homework.	pr	fu	fu	fu
Another day of great restraint.	pa	pa	pr	pa

Table 1: Examples of statuses annotated for temporal classes: past (pa), present/none (pr), and future (fu). R1, R2, R3: judgements from each rater; Maj: choice from majority voting. The bottom three examples illustrate difficult cases.

lengths: *mean size of 1grams and number of tokens in the post.*

We found it useful to use a modest variety of feature types and to build on existing work that labels time expressions. While one might expect time expression features to be extremely valuable for this task, we found only 15% of Facebook messages contain them, even though many more communicate a focus on the past or future through other means (e.g. tense or semantic information). All features were limited to those mentioned in at least 0.05% of messages.

At the user-level, we produce three categories of temporal orientation, defined simply as the proportion of a user's total messages ($msgs(user)_{all}$) classified in the given temporal category ($tc \in \{past, present, future\}$):

$$orientation_{tc}(user) = \frac{|msgs_{tc}(user)|}{|msgs_{all}(user)|}$$

We generate three separate variables (summing to one), rather than a single variable temporal index, in order to capture non-linear relationships (i.e., the potential for the present to correlate in the opposite direction of both the past and future). All of our user analyses are based on 100 randomly selected messages from each user.

4 Data collection and labeling

We use three social media datasets: the *training set*, *test set*, and *user set*. The *training set* consists of 4,302 Twitter and Facebook annotated messages. The *test set* is a random subset of 500 annotated Facebook messages, representative of messages we will apply our model. Finally, the *user set* contains 531,893 messages from Facebook users with known age, gender, personality, satisfaction with life, depression, IQ and number of Facebook friends. We derived the *test set* from the *user set* in order to establish accuracies of our model over the application domain.

Training set. Our training data consists of both Facebook and Twitter messages. For Facebook, 3,000 status updates, sent between March 2009 and October 2011, were randomly sampled from users of the MyPersonality application (Kosinski and Stillwell, 2012; Quercia et al., 2012), who also provided their age and gender. For Twitter, 3,000 messages were sampled from the 1% *random* stream provided by Twitter during September 2012.

Three annotators, undergraduate students at the University of Pennsylvania, independently labeled the temporal orientation of each message. Messages were labeled in units of days past or future (adapted from Liberman et al. (2007)). For example, -7 would be a week ago, $^{-1}/_{24}$ would be an hour ago, 0 would be now (present), and 365 would be a year from now. Inter-annotator agreement, as the intraclass correlation coefficient (Shrout and Fleiss, 1979), was 0.85. Ratings were averaged into a single "time from now" index. For the purposes of this study we then discretized the data into past (mean rating < 0), present (mean rating $= 0$), or future (mean rating > 0). Annotation of the 6,000 messages took approximately 150 human hours.

When rating, messages were marked 'NA' when they appeared to come from a bot or were composed of song lyrics or quotations. (Removing unoriginal content was desired for the consumer behavior research for which the messages were first labeled.) For our purposes, in order to maximize the training set size, we only removed messages when all three raters chose 'NA', such that there was no average rating available for the message. The resulting final *training set* consisted of 4,302 total messages (2,009 tweets; 2,293 Facebook status updates). Since our application of the data does not include a manual filtering of messages, we created a separate message test set with no filtering in order to accurately evaluate our classifier in the application's setting (below).

Test set. Evaluating our classifiers over our annotated training set would not yield an accurate assessment of the performance when applied to the *user set* (described next). Therefore, we randomly selected 500 statuses from the *user set* as our message *test set*.[5] Statuses exclude

[5]While we desired a large training set, the test set only needed to be large enough to evaluate differences in accuracy.

	mfc	logR	lSVC	rSVC	ERTs	*msgs*
Accuracy	.528	.686	.708	.684	**.718**	*500*
past (p, r, f1)	(.00, .00, .00)	(.56, .68, .62)	(.63, .67, .65)	(.63, .56, .59)	(.73, .67, .71)	*131*
present (p, r, f1)	(.53, 1.0, .69)	(.80, .74, .77)	(.78, .78, .78)	(.70, .85, .77)	(.74, .84, .79)	*264*
future (p, r, f1)	(.00, .00, .00)	(.60, .56, .58)	(.61, .56, .58)	(.69, .43, .53)	(.60, .47, .53)	*105*

Table 2: Accuracy (percentage classified correct) message classifiers based on different learning algorithms (identified in section 3). Temporal class results are broken down by precision (p), recall (r), and f1 score for each of past (pa), present (pr), and future (fu). Number of messages (*msgs*) are listed on the far right. The most frequent class baseline (mfc) indicates accuracy if only predicting the present class.

reposts of others' statuses and comments on other people's posts, and we found only 2 of the 500 random messages were made by apps (users still choose whether or not to post these to their walls). Three annotators independently classified each status message as predominantly talking about the past, present, or future. The overall rating for each message was determined by a majority vote (when there was a tie: i.e., one of each class, present was used). Agreement among these raters, calculated as the intraclass correlation coefficient, was 0.83. One might suggest some messages do not have a temporal class (e.g. does "I like Selena Gomez" have a predominant temporal class?).[6] Such messages would be marked as 'present' in our annotation scheme. Thus, one might consider our present class to encompass both a present and "non-temporal" class.

User set. Human-level data is used to evaluate our model toward understanding the relationship between human temporal orientation and individual characteristics (e.g. demographics, personality). Thus, this data spans both message and user levels, from consenting participants, in the MyPersonality Facebook study (Kosinski and Stillwell, 2012).

We used five subsets of the MyPersonality data in order to capture various psychological and behavioral variables: *user subset 1*: gender, age, and personality; *user subset 2*: satisfaction with life, *user subset 3*: depression, *user subset 4*: IQ, and *user subset 5*: number of friends. For our first subset, gender, age, and personality variables are well represented in the dataset, so we created a stratified sample over 1,520 users. We sampled equal proportions of males and females across 4-year age bins from 13 to 60 (i.e., ages *[13,16], [17,20], . . . ,[57, 60]*), which provides gender- and age-controlled correlations for each personality factor (openness, conscientiousness, extraversion, agreeableness, and neuroticism).

Other variables are more limited. Thus, instead of creating stratified samples, these three subsets include users for whom gender and age information is also available: 1,565 in the case of satisfaction with life, 268 for the CES-D depression scale, 898 for IQ and, 1,000 in the case of number of friends. The gender and age data is then included as covariates in regression analyses to find the relationship between these variables and temporal orientation, controlled for demographics. In all five subsets of the *user set*, we randomly sample 100 messages from each user in order to determine their temporal orientation.

Table 1 shows example status updates along with ratings. As evidenced from the rater agreement, most status updates were fairly easy to determine. Some messages have explicit temporal phrases (e.g. "in the future") while others are more subtle (e.g. relying on verb tense: "is listening ..."). Others, such as the bottom three examples, might reference multiple temporal classes or not include clear verb tense and thus rely on the raters' judgements for what is most dominant.

5 Evaluation

We evaluate our past, present, and future message classifier as well as its features. All models were trained over our *training set* and evaluated over the *test set*.

Table 2 compares accuracy of various types of classifiers: logistic regression (logR), linear support vector classifier (lSVC), support vector classifier with rbf kernel (rSVC), and a forest of extremely randomized trees (ERTs). We saw best results from the ERTs classifier, suggesting some of its benefits (capturing non-linear relationships or interactions among features) may help this problem. We also see, from the F1 scores, that all classifiers found the future class most difficult to predict; this was the smallest class and likely subject to the most bias against. All classifiers performed significantly better than the most frequent class baseline with an error reduction of 41% in the case of ERTs ($p < 0.001$ from paired t-test on absolute errors). We selected the ERTs classifier for the remaining experiments.[7]

[6] We attempted to include a non-temporal class and found disagreement. Some argue that every message has a temporal class (e.g. "I like Selena Gomez" is truly signalling present).

[7] This trained classifier is available at: wwbp.org/data.html.

Past		Present		Future	
type	feature	type	feature	type	feature
POS	*verb, past tense*	time exs -	*num of timexs*	lexica	*relativity*
ngrams	*was*	POS -	*verb, past tense*	time exs	*num of timexs*
lexica -	*common present verbs*	lexica -	*relativity (in, on, at, ...)*	POS	*verb, base form*
lengths	*num of tokens*	lexica	*common present verbs*	ngrams	*tomorrow*
ngrams	*had*	POS	*verbs, 3rd pers singular*	POS	*to*

Table 4: Top five most correlated features for each of the temporal classes. '-' indicates negative correlations; positive otherwise. Correlation absolute strengths ranged from Pearson $r = 0.08$ to 0.40.

Features	Accuracy	Features	Accuracy
mfc baseline	.528	all features	**.718**
ngrams alone	.688	w/o *ngrams*	.672
time exs alone	.586	w/o *time exs*	.708
POS tags alone	.614	w/o *POS tags*	.712
lexica alone	.684	w/o *lexica*	.702
lengths alone	.544	w/o *lengths*	.718

Table 3: Accuracy of our full past, present, future message classifier (top) and an ablation analysis of accuracies when removing each feature type (bottom). The most frequent class baseline (mfc) indicates accuracy if only predicting the present class. The full classifier significantly out-performed using time expressions alone ($p < 0.05$; bolded accuracy).

We did feature ablation analyses as shown in Table 3. Every feature type produced improvement over the baseline and, with the exception of $lengths$, removing any feature resulted in reduced performance (though none strong enough to meet significance at $p < 0.05$). The limited reduction implies that while each feature type may contain temporal information, there is also substantial redundancy across the feature types.

Results when using time-expression alone can be considered another baseline, representing a model based entirely on previous time expression work. A reason for the large advantage of using additional features is that many temporally indicative messages did not contain any time expressions (instead expressing orientation through verb tense or semantics). Indeed, we see from Table 4, which lists the top features for each class, that time expression features were very useful when then occurred. All feature types made it into these top ten lists.

User-level temporal orientation is trivially defined: percentage of a given user's messages that are classified as past, present, or future oriented. Thus, accuracy for each proportion is directly tied to message accuracy. Still, we validate that our approach is in line with psychological theory (discussed in Section 2) by correlating user-level temporal orientations with outcomes whose associations have been previously established: conscientiousness, age, and gender number of users; results controlled for age and gender. In particular, future orientation should be positively correlated with conscientiousness, age, and being female, while present orientation should be negatively correlated. Our results which which are consistent with the literature, can be seen in the top half of Table 5. Among users with personality scores, we found positive correlations between future orientation and seemingly future-oriented questionnaire items: "I make plans and stick to them" ($r = .16$) and "I finish what I start" ($r = .12$). To the best of our knowledge, psychology literature has not established standard correlates of past orientation, so the correlation with age and past orientation, though not surprising, is somewhat novel.

Correlations with questionnaire measures help to establish convergent validity — i.e. our measure is empirically related to other measures in a way that is consistent with theory about the underlying constructs. However, self report questionnaires are often used for convenience, not necessarily because they are most valid (Paulhus and Vazire, 2007). In fact, more objective or behavior-based measures in social science have been called for (Baumeister et al., 2007).

6 Exploration

Here we use our language-based user measure to explore behavioral and psychological correlates of temporal orientation. Figure 1 illustrates the user-level distributions of the message classes, broken down by gender, within the stratified sample. All experiments in this section applied our measure over the *user set*. Within users, the mean proportions of past, present, and future messages were 0.24, 0.61, and 0.15 respectively. Among most users, the majority of messages were classified as present, while future-oriented messages were least frequent.

We compare these temporal orientations to user personality, satisfaction with life, IQ, and their number of Facebook friends. We use both Pearson correlation and linear regression (OLS) to estimate relationships between

Attribute	*N*	Past	Present	Future
validation				
conscientiousness	*1520*	.02	**-.08**	**.12**
age	*1520*	**.30**	**-.30**	**.15**
gender	*1520*	**.10**	**-.15**	**.14**
exploration				
openness	*1520*	.05	.04	**-.12**
extraversion	*1520*	-.04	.03	.00
agreeableness	*1520*	.00	-.02	.04
neuroticism	*1520*	-.01	-.01	.04
satisfaction w/ life	*1565*	.00	-.05	**.08**
depression	*268*	-.14	**.21**	**-.17**
IQ	*898*	**.14**	**-.14**	.05
# of friends	*1000*	**-.15**	**.13**	-.05

Table 5: Correlations between user temporal orientation and human attributes. The attributes conscientiousness, age, and gender are well-established in previous literature to be associated with temporal orientation. Gender is coded as 0 = male, 1 = female. **bold**: $p < .01$ after Benjamini-Hochberg multiple comparison correction; *N*: number of users; results controlled for age and gender.

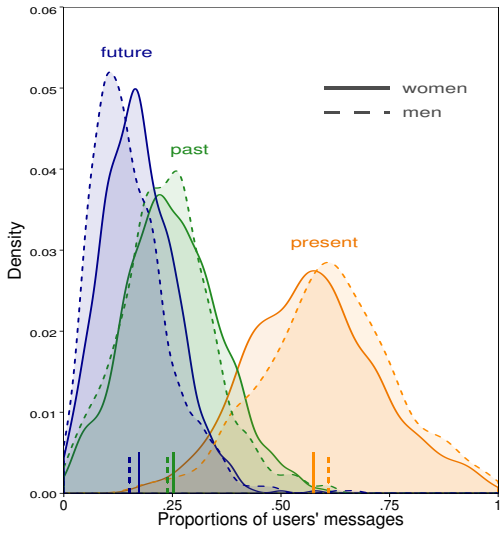

Figure 1: Kernel density estimates of user-level proportions of past, present, and future classified messages, broken down by gender. Vertical bars represent means.

temporal orientation and other variables. In our age and gender stratified sample (1520 users), we calculate Pearson correlations between temporal orientation and age, gender, and measures of openness, conscientiousness, extraversion, agreeableness, and neuroticism. Because fewer users completed measures of satisfaction with life, IQ, and number of friends, we were not able to produce sufficient stratified samples, so we used ordinary least squares linear regression to fit the standardized outcome of interest to standardized temporal orientation also including standardized age and gender as covariates to adjust for their effects. The coefficient, often denoted β, for temporal orientation then represents the strength of the relationship, controlled for age and gender.

Table 5 lists the correlation coefficients between temporal orientation and user attributes. We found the strongest effects for age, with patterns that are consistent with the psychological literature (Steinberg et al., 2009). Figure 2 illustrates trends from age 13 to 60. Most notably, present orientation decreases steadily across age; past orientation steadily increases, and future orientation increases quickly throughout adolescence, slows in early adulthood, and finally levels off in late adulthood. Female users were significantly more future-oriented, slightly more past-oriented, and significantly less present-oriented than males.

For personality, we found the expected patterns of correlations with conscientiousness, but more interestingly, openness to experience was correlated with lower future orientation (beyond conscientiousness, there was no support in the literature for any of the five personality factors to correlate with temporal orientation). This is surprising when considering openness to experience is characterized by creativity and intellect (McCrae and John, 1992), yet IQ instead correlates with less present and more past orientation, suggesting future orientation characterizes a difference between the two.[8]

We found a modest yet significant positive correlation between future orientation and satisfaction with life, with future orientation associated with higher life satisfaction. On the other hand, we found a stronger negative correlation between future orientation and depression as well as a positive correlation between present orientation and depression. As previously noted, past literature was conflicting on the relationship between these factors, so our study weighs in with a behavior-based assessment in support of a future-oriented people being more satisfied in life and less depressed.

Lastly, we consider whether the use of our language-based measure of temporal orientation can track changes over time. As a proof-of-concept, we focus on patterns around birthdays; excluding messages containing birthday terms and users turning 21, we calculated the standardized proportion of messages which were future-

[8]One interpretation is that our classifier may be more accurate on messages authored by those with higher IQ (i.e. more grammatical sentences); however no significant difference in error was found when spitting messages by the authors' IQ, age, or gender.

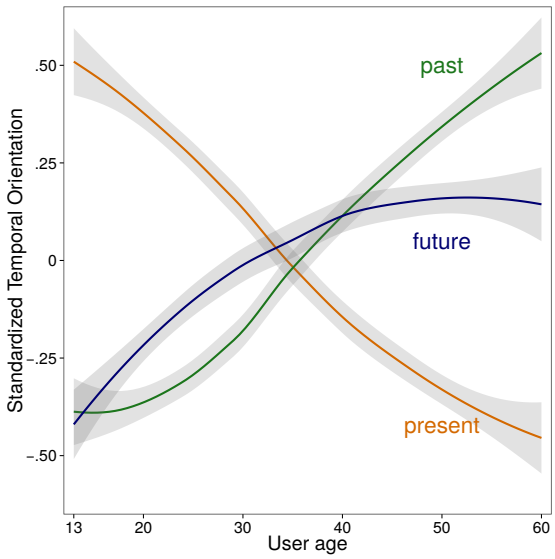

Figure 2: Standardized temporal orientation of users over their age. Shaded area indicates ± 1 standard error around loess smoothed estimates.

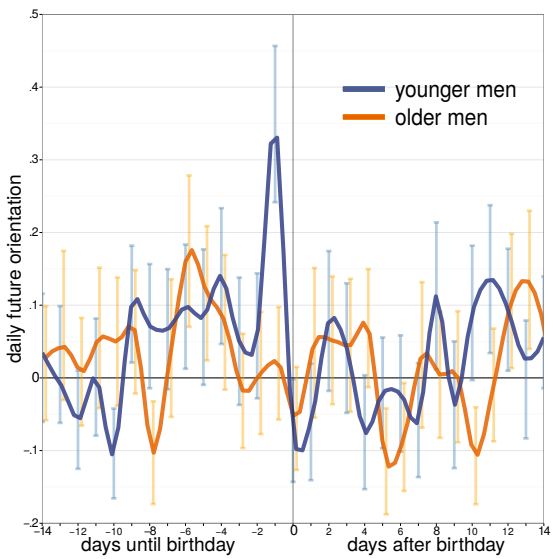

Figure 3: Daily proportion of future-oriented messages leading up to and following birthdays. Younger men: $<$ 23, N = 423; Older men: $\geq$ 23, N = 402.

oriented over two weeks before and after one's birthday. We controlled for individual differences and date effects (e.g. if more people happen to be born near a holiday) by standardizing users' future orientation over all their messages and standardizing daily future orientation scores over all messages from each day. Figure 3 shows the patterns for men $<$ 23 and $\geq$ 23, suggesting that younger men look forward to their birthdays while older men do not (patterns were less pronounced for women).

7 Conclusions

This is, to our knowledge, the first study to automatically assess individual temporal orientation through language, and it constitutes one of the largest (perhaps the largest) studies of temporal orientation. Our message-level past, present, and future classifier achieved an accuracy of 71.8%, well-above most frequent sense, pasts-of-speech only, and time expression only baselines. The associations we found between user-level temporal orientation and conscientiousness, gender, and age validated our novel method against well-established correlates. We then explored novel links with other personality factors, satisfaction with life, depression, IQ, and number of friends.

Our automatic labeling of temporal orientation yielded strong accuracy. There are, however, several ways in which our analysis might be improved or extended. We used a coding scheme that did not allow a "no temporal orientation" class. Separate handling of "non-temporal" messages might provide more nuanced profiles of user orientation. Another simplification was our binning of temporal relations into past, present, and future. One might want to instead use a continuous measure of temporal distance, indicating how far into the past or future a message is oriented, providing a more detailed profile of individual orientations. Among the future-oriented, for example, there may be important differences between those who plan for tomorrow versus those who plan years in advance. Assessing additional characteristics of messages, such as sentiment, may also allow further insight and more nuanced characterizations (e.g. "I will be fine" vs. "I will never be ok").

Whereas prior computational linguistics work with temporal relations has focused on classifying events, we focus on people and open many avenues for social scientific investigations that were previously not very feasible. For instance, temporal orientation was generally treated as a stable trait (with little change through the lifetime); For men, we found age differences in temporal orientation leading up to one's birthday. Learning how, why, and when people become more future-oriented may inform planning and risk assessment interventions, and lead to better understanding of health and economic prosperity.

8 Acknowledgments

We thank George Wan for his assistance running prototype classifiers. This work was supported by the Templeton Religion Trust under grant #TRT0048.

References

Jean Adams and Daniel Nettle. 2009. Time perspective, personality and smoking, body mass, and physical activity: An empirical study. *British Journal of Health Psychology*, 14(1):83–105.

Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. In *ICWSM*.

James F Allen. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

Shlomo Argamon, Moshe Koppel, James W. Pennebaker, and Jonathan Schler. 2009. Automatically profiling the author of an anonymous text. *Commun. ACM*, 52(2):119–123, February.

Roy F Baumeister, Kathleen D Vohs, and David C Funder. 2007. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2(4):396–403.

Shane Bergsma, Mark Dredze, Benjamin Van Durme, Theresa Wilson, and David Yarowsky. 2013. Broadly improving user classification via communication-based name and location clustering on twitter. In *Anuual Meeting of HLT-NAACL*, pages 1010–1019.

Christopher M Bishop. 2006. *Pattern recognition and machine learning*, volume 4. Springer New York.

Ilona Boniwell, Evgeny Osin, P Alex Linley, and Galina V Ivanchenko. 2010. A question of balance: Time perspective and well-being in british and russian samples. *The Journal of Positive Psychology*, 5(1):24–40.

John N Boyd and Philip G Zimbardo. 2005. Time perspective, health, and risk taking. In *Understanding behavior in the context of time*, pages 85–107. Lawrence Erlbaum, Mahwah.

John D Burger, John Henderson, George Kim, and Guido Zarrella. 2011. Discriminating gender on twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1301–1309. Association for Computational Linguistics.

Angel X Chang and Christopher Manning. 2012. Sutime: A library for recognizing and normalizing time expressions. In *LREC*, pages 3735–3740.

Yi Chang, Xuanhui Wang, Qiaozhu Mei, and Yan Liu. 2013. Towards twitter context summarization with user influence models. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 527–536. ACM.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Leon Derczynski, Diana Maynard, Niraj Aswani, and Kalina Bontcheva. 2013. Microblog-genre noise and impact on semantic annotation accuracy. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 21–30. ACM.

Fien Desmyter and Rudi De Raedt. 2012. The relationship between time perspective and subjective well-being of older adults. *Psychologica Belgica*, 52(1):19–38.

ED Diener, Robert A Emmons, Randy J Larsen, and Sharon Griffin. 1985. The satisfaction with life scale. *Journal of personality assessment*, 49(1):71–75.

Ed Diener, Louis Tay, and Shigehiro Oishi. 2013. Rising income and the subjective well-being of nations. *Journal of Personality and Social Psychology*, 104(2):267–276.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

Wei Feng and Jianyong Wang. 2013. Retweet or not?: personalized tweet re-ranking. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 577–586. ACM.

William J Friedman. 2000. The development of children's knowledge of the times of future events. *Child Development*, 71(4):913–932.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine learning*, 63(1):3–42.

Lewis R Goldberg. 1990. An alternative" description of personality": the big-five factor structure. *Journal of Personality and Social Ssychology*, 59(6):1216–1229.

Sumit Goswami, Sudeshna Sarkar, and Mayur Rustagi. 2009. Stylometric analysis of bloggers age and gender. In *Third International AAAI Conference on Weblogs and Social Media*.

Pedro Calais Guerra, Wagner Meira Jr, and Claire Cardie. 2014. Sentiment analysis on evolving social streams: How self-report imbalances can help. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 443–452. ACM.

Francisco Iacobelli, Alastair J. Gill, Scott Nowson, and Jon Oberlander. 2011. Large scale personality classification of bloggers. In *Proc of the 4th int conf on Affect comput and intel interaction*, pages 568–577. Springer-Verlag.

William James. 1890. *The principles of psychology.* Henry Holt and Company.

Kelli A Keough, Philip G Zimbardo, and John N Boyd. 1999. Who's smoking, drinking, and using drugs? Time perspective as a predictor of substance use. *Basic and Applied Social Psychology*, 21(2):149–164.

M Kosinski and D Stillwell. 2012. myPersonality research wiki. http://www.mypersonality.org/wiki/.

Kurt Lewin. 1942. Time perspective and morale. In *Resolving social conflicts and Field theory in social science*, pages 80–93. APA, Washington, D. C.

Nira Liberman, Yaacov Trope, Sean M McCrea, and Steven J Sherman. 2007. The effect of level of construal on the temporal distance of activity enactment. *Journal of Experimental Social Psychology*, 43(1):143–149.

Xiao Ling and Daniel S Weld. 2010. Temporal information extraction. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence*.

F. Mairesse, M.A. Walker, M.R. Mehl, and R.K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30(1):457–500.

Robert R McCrae and O P John. 1992. An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215.

Jari-Erik Nurmi. 2005. Thinking about and acting upon the future: Development of future orientation across the life span.

In Alan Strathman and Jeff Joireman, editors, *Understanding behavior in the context of time*, pages 31–57. Lawrence Erlbaum, Mahwah.

Rina Panigrahy, Marc Najork, and Yinglian Xie. 2012. How user behavior is related to social affinity. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 713–722. ACM.

Greg Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, David J Stillwell, Michal Kosinski, Lyle H Ungar, and Martin EP Seligman. 2015. Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*.

Delroy L Paulhus and Simine Vazire. 2007. The self-report method. *Handbook of research methods in personality psychology*, pages 224–239.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12:2825–2830.

James W. Pennebaker, C.K. Chung, M. Ireland, A. Gonzales, and R.J. Booth. 2007. The development and psychometric properties of liwc2007. *Austin, TX, LIWC. Net*.

Daniele Quercia, Renaud Lambiotte, David Stillwell, Michal Kosinski, and Jon Crowcroft. 2012. The personality of popular facebook users. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 955–964. ACM.

Lenore Sawyer Radloff. 1977. The ces-d scale a self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401.

Brent W Roberts, Carl Lejuez, Robert F Krueger, Jessica M Richards, and Patrick L Hill. 2013. What is conscientiousness and how can it be assessed? *Developmental Psychology*, (online advance publication).

Maarten Sap, Gregory Park, Johannes C. Eichstaedt, Margaret L. Kern, David J. Stillwell, Michal Kosinski, Lyle H. Ungar, and H. Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *EMNLP 2014*, pages 1146–1151. Association for Computational Linguistics.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Martin EP Seligman, Peter Railton, Roy F Baumeister, and Chandra Sripada. 2013. Navigating into the future or driven by the past. *Perspectives on Psychological Science*, 8(2):119–141.

Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin*, 86(2):420–428.

Laurence Steinberg, Sandra Graham, Lia OBrien, Jennifer Woolard, Elizabeth Cauffman, and Marie Banich. 2009. Age differences in future orientation and delay discounting. *Child Development*, 80(1):28–44.

Alan Strathman and Jeff Joireman, editors. 2005. *Understanding behavior in the context of time: Theory, research, and application*. Lawrence Erlbaum, Mahwah.

Partha Pratim Talukdar, Derry Wijaya, and Tom Mitchell. 2012. Coupled temporal scoping of relational facts. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 73–82. ACM.

Kristina Toutanova, Dan Klein, Christopher D Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 173–180. Association for Computational Linguistics.

Naushad UzZaman, Hector Llorens, Leon Derczynski, Marc Verhagen, James Allen, and James Pustejovsky. 2013. SemEval-2013 task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of the 7th International Workshop on Semantic Evaluation*.

Marc Verhagen, Inderjeet Mani, Roser Sauri, Robert Knippen, Seok Bae Jang, Jessica Littman, Anna Rumshisky, John Phillips, and James Pustejovsky. 2005. Automating temporal annotation with TARSQI. In *Proceedings of the ACL 2005 on interactive poster and demonstration sessions*, pages 81–84. Association for Computational Linguistics.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 75–80. Association for Computational Linguistics.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. SemEval-2010 task 13: TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 57–62. Association for Computational Linguistics.

Paul Webley and Ellen K Nyhus. 2006. Parents influence on childrens future orientation and saving. *Journal of Economic Psychology*, 27(1):140–164.

Miaomiao Wen, Zeyu Zheng, Hyeju Jang, Guang Xiang, and Carolyn Penstein Rosé. 2013. Extracting events with informal temporal references in personal histories in online communities. In *Proceedings of 51st annual meeting of the ACL*, pages 836–842.

Philip G Zimbardo and John N Boyd. 1999. Putting time in perspective: A valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, 77(6):1271–1288.