How are you doing?
Emotions and Personality in Facebook

Golnoosh Farnadi[1,2,3], Geetha Sitaraman[2], Mehrdad Rohani[2], Michal Kosinski[4], David Stillwell[4], Marie-Francine Moens[3], Sergio Davalos[2], and Martine De Cock[1,2]

[1]Dept. of Appl. Math., Comp. Science and Statistics, Ghent University, Belgium
[2]Center for Data Science, University of Washington Tacoma, WA, USA
[3]Dept. of Computer Science, Katholieke Universiteit Leuven, Belgium
[4]The Psychometrics Centre, University of Cambridge, Cambridge, UK
{gfarnadi,sgeetha,rohanim,sergiod,mdecock}@uw.edu
{mk583,ds617}@cam.ac.uk
{golnoosh.farnadi,sien.moens}@cs.kuleuven.be
{golnoosh.farnadi,martine.decock}@ugent.be

Abstract. User generated content on social media sites is a rich source of information about latent variables of their users. Proper mining of this content provides a shortcut to emotion and personality detection of users without filling out questionnaires. This in turn increases the application potential of personalized services that rely on the knowledge of such latent variables. In this paper we contribute to this emerging domain by studying the relation between emotions expressed in approximately 1 million Facebook (FB) status updates and the users' age, gender and personality. Additionally, we investigate the relations between emotion expression and the time when the status updates were posted. In particular, we find that female users are more emotional in their status posts than male users. In addition, we find a relation between age and sharing of emotions. Older FB users share their feelings more often than young users. In terms of seasons, people post about emotions less frequently in summer. On the other hand, December is a time when people are more likely to share their positive feelings with their friends. We also examine the relation between users' personality and their posts. We find that users who have an open personality express their emotions more frequently, while neurotic users are more reserved to share their feelings.

Keywords: Emotion detection, NRC lexicon, User modelling, Big Five Personality model, Facebook, Social media, Time factor

1 Introduction

As more and more users are creating their own content on the web, there is a growing application potential for personalization in human computer systems such as personalized information access services, recommender systems, and tailored advertisements. To provide personalization, a wide variety of user information can be processed. Previous personalization research has focused on explicit user characteristics such as demographics, e.g., age, gender, location, or language [1]. Implicit user behavior has also been used, but existing research and applications focus mostly on users' online activities such as clicking behavior, web search history, mouse movement, or the amount of time that a user is looking at a web page [2].

Social media websites provide a unique opportunity for personalized services to use other aspects of user behavior. Besides users' structured information contained in their profiles, e.g., demographics, users produce large amounts of data about themselves in variety of ways including textual (e.g., status updates, blog posts, comments) or audiovisual content (e.g., uploaded photos and videos). Many latent variables such as personalities, emotions and moods — which are typically not explicitly given by users — can be extracted from user generated content (see e.g. [6, 7]).

In this study, we examine the relationship between users' emotions and other characteristics on Facebook. Little work has been done that examines the relation between a user's emotions and other characteristics in social media. In [3] the authors extract emotions from Twitter posts and find correlations with major events in politics and popular culture during a specific time frame, but they focus on the public emotion as a whole and not on feelings or other characteristics of individual users. We detect emotions from users' status updates using the NRC word-emotion lexicon [13], and determine the relation between users' feelings and their demographics (age and gender) and personality. We also extract time features from the time stamp of the status updates to find the relation between users' emotions and time. To the best of our knowledge, no work has been done to find the relations between different emotions and personality with respect to time factors. In [14] the authors study the relation between emotions and time, however their work is based on a questionnaire and not based on social media content. In [12], the authors use SVM classifiers to predict personality using emotion expression in text. For their experiments, they use essays from psychology students, while in this work we focus on emotion expression in Facebook status updates and its relation with users' personality.

1.1 Personality and Emotions

Personality is a fundamental differentiating factor of human behavior. Research in the psychology literature has led to a well established model for personality recognition and description, called the Big Five Personality Model. Five traits can be summarized in the following way [5]:

- **Openness to experience** (Openness) is related to imagination, creativity, curiosity, tolerance, political liberalism, and appreciation for culture. People scoring high on Openness like change, appreciate new and unusual ideas, and have a good sense of aesthetics.
- **Conscientiousness** measures preference for an organized approach to life in contrast to a spontaneous one. Conscientious people are more likely to be well organized, reliable, and consistent. They enjoy planning, seek achievements, and pursue long-term goals. Non-conscientious individuals are generally more

easy-going, spontaneous, and creative. They tend to be more tolerant and less bound by rules and plans.

- **Extroversion** measures a tendency to seek stimulation in the external world, the company of others, and to express positive emotions. Extroverts tend to be more outgoing, friendly, and socially active. They are usually energetic and talkative; they do not mind being at the center of attention, and make new friends more easily. Introverts are more likely to be solitary or reserved and seek environments characterized by lower levels of external stimulation.
- **Agreeableness** relates to a focus on maintaining positive social relations, being friendly, compassionate, and cooperative. Agreeable people tend to trust others and adapt to their needs. Disagreeable people are more focused on themselves, less likely to compromise, and may be less gullible. They also tend to be less bound by social expectations and conventions, and more assertive.
- **Emotional Stability** (reversely referred to as Neuroticism) measures the tendency to experience mood swings and emotions such as guilt, anger, anxiety, and depression. Emotionally unstable (neurotic) people are more likely to experience stress and nervousness, while emotionally stable people (low Neuroticism) tend to be calmer and self-confident.

Personality can affect the decision making process and has been shown to be relevant in the selection of music, movies, TV programs and books. It has been shown that personality affects preference for websites [10], language used in online social media [17], choice of Facebook Likes [11], music taste [16], and content such as movies, TV shows, and books [4].

In addition, it has been shown that users' *emotions* can also be used to detect users' taste at any moment, e.g., sad users are more likely to prefer action movies to watch [9]. Going yet one step further, personalized services can even have an impact on users' feelings. A nice example of this is that watching movies can change users' emotion, e.g., people feel joy when watching comedies or sadness when watching a late night romantic movie [9].

An interesting difference between personality and emotion is that personality is a stable characteristic and emotions are of short term duration. Emotion can be a momental feeling with respect to an object, person, event, or situation. As a consequence, people express a variety of different emotions over a period of time which is not the case for users' personality.

2 Material and Methods

2.1 Dataset

Our results are based on a data set from the myPersonality project [11]. MyPersonality was a popular Facebook application introduced in 2007 in which users took a standard Big Five Factor Model psychometric questionnaire [8] and gave consent to record their responses and Facebook profile. The data set consists of information about users' demographics, friendship links, Facebook activities (e.g., number of group affiliations, page likes, education and work history), status updates and Big Five Personality Scores. However, not all of this information is available for all users. From this data, we produce a data set of 5,865 users for which we have complete information about their age, gender, personality scores and status updates. Table 1 provides details about the data set characteristics.

Table 1: (Table on the left) Characteristics of female and male users in the data set. The entire data set contains 969,035 status updates written by 5,865 users. (Table on the right) Score threshold and number of users for each personality trait. Note that the same user can exhibit more than one personality trait at once.

	Female	Male
# users	3,446	2,419
Average age	26	25
# posts	625,921	343,114
Avg # posts/user	182	142
Min # posts/user	1	1
Max # posts/user	2,428	1,453

Personality	Threshold	# of users
Extroversion	3.60	2,971
Openness	3.80	3,284
Agreeableness	3.55	3,110
Conscientiousness	3.50	3,071
Neuroticism	2.80	2,631

The data set contains a personality score ranging from 1 to 5 for each user and each personality trait. To facilitate further analysis, for each personality trait we split the set of users into those that clearly exhibit the trait and those who do not. To this end we use the same thresholds that were used in the WCPR13 data set.[1] The score threshold and the number of users for each personality trait is presented in Table 1. For instance, in the remainder of this paper, we call a user an extrovert if his Extroversion score is at least 3.60; there are 2,971 such users in our data set. Note that such a binary split of users along the 5 personality dimensions is a fairly crude approach, and that a more fine grained study that considers the sliding scale from Introversion to Extroversion could provide further insights.

2.2 Emotion detection

To detect users' emotions from their status updates, we use the NRC hashtag emotion lexicon [13]. This lexicon contains a 10-dimensional binary emotion vector for 14,177 English words. The 10 dimensions or emotion categories are: *positive*, *negative*, *anger*, *anticipation*, *disgust*, *fear*, *joy*, *sadness*, *surprise*, and *trust*. In the NRC lexicon, *positive* and *negative* are actually referred to as sentiments instead of emotions, but in our study we use the terms emotions and feelings loosely and interchangeably to refer to all 10 categories of the NRC lexicon.

A word can convey several emotions at the same time. For instance, according to the NRC lexicon, "happy" represents positive, anticipation, joy, and trust emotions, while "birthday" represents positive, anticipation, joy, and surprise

[1] http://mypersonality.org/wiki/doku.php?id=wcpr13

emotions. In the remainder of this paper, we say that a status update conveys an emotion if it contains at least one term from the lexicon that is associated with that emotion. For example, the status update "thanks to everyone who wished me a happy birthday today" conveys positive, anticipation, joy, trust, and surprise emotions because of the presence of the words "happy" and "birthday". The other words in this particular status update do not convey any emotion according to the NRC lexicon.

Figure 1 presents the frequency of emotions in the posts in our data set. Almost 60% of the status updates express at least one kind of emotion, and the positive emotion is clearly the most prominent one. For completeness, we

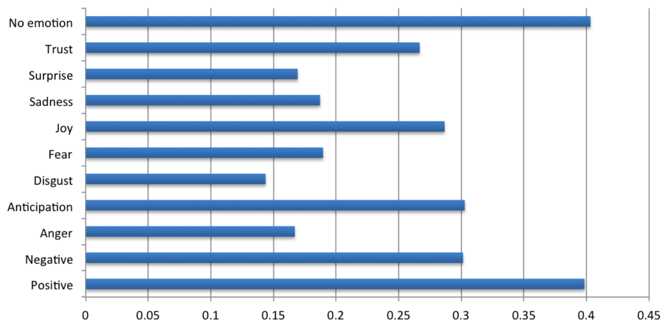

Fig. 1: Emotion frequency in Facebook status updates. Almost 60% of the status updates express at least one kind of emotion from the NRC lexicon, and many posts convey more than one emotion.

point out that to detect emotions we only scan the status updates for exact occurrences of words from the NRC lexicon. We use a bag of words approach and do not consider any misspellings (e.g., hapy or haaaappy), negation (e.g., not good), strength of the emotions using adjectives or adverbs (e.g., very happy vs. happy) or combined words (e.g., long-awaited vs. long awaited). Moreover, any emotions expressed with words that are not present in the NRC lexicon will remain undetected.

3 Results

In the remainder of this paper, let S denote the set of the 969,035 status updates in our study. Furthermore, for each of the 10 emotions 1:positive, 2:negative, 3:anger, 4:anticipation, 5:disgust, 6:fear, 7:joy, 8:sadness, 9:surprise, and 10:trust, let S_i, $i = 1, \ldots, 10$, be the set of status updates that contain at least one word associated with the respective emotion according to the NRC lexicon. As explained in Section 2.2, the sets S_1, S_2, …, S_{10} are not necessarily disjoint. In addition, we also introduce S_0 as the set of status updates that do not contain a term from the NRC lexicon, i.e. S_0 is the set of status updates that do not convey any emotion. It holds that $S = S_0 \cup S_1 \cup S_2 \cup \ldots \cup S_{10}$.

3.1 Emotion and gender

Let S_f denote the set of status updates written by female authors and S_m the set of status updates by male authors. From Table 1 we know that women post more frequently than men. The probability that a status update is written by a woman is $P(S_f) \approx 0.65$ while the probability that it is written by a man is $P(S_m) \approx 0.35$. To determine the probability that a post conveys a particular emotion, given that it is written by a man or a woman, we calculate

$$P(S_i|S_m) = \frac{|S_i \cap S_m|}{|S_m|} \text{ and } P(S_i|S_f) = \frac{|S_i \cap S_f|}{|S_f|}$$

for $i = 0, 1, \ldots, 10$. The results are visualized in Figure 2.

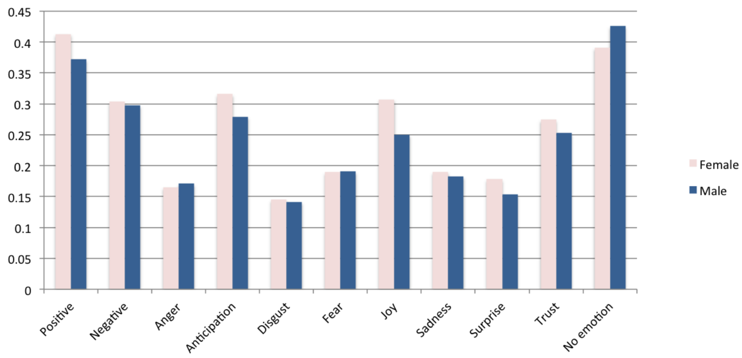

Fig. 2: Probability of occurrences of emotions in status updates from female and male users

Although the differences between both genders are small, we do observe that female users in general express more emotions in their posts. In particular, women are more likely than men to post about positive feelings, joy and anticipation, while men are more likely than women to post status updates that convey anger or no emotion at all.

3.2 Emotion and age

To assess the relation between different age groups and their emotion expression in Facebook, we use five age groups: users younger than 21, users between 21 and 30, users between 31 and 40, users between 41 and 50, and users older than 51. The average age of users in our data set is 26 years old with a standard deviation of 10, suggesting many young users in Facebook. For each age group a, let S_a be the set of status updates written by users from that age group. We calculate the probability of emotion expression for each age group a as $P(S_i|S_a) = \frac{|S_i \cap S_a|}{|S_a|}$ with S_i (for $i = 0, 1, \ldots, 10$) defined as in the beginning of Section 3.

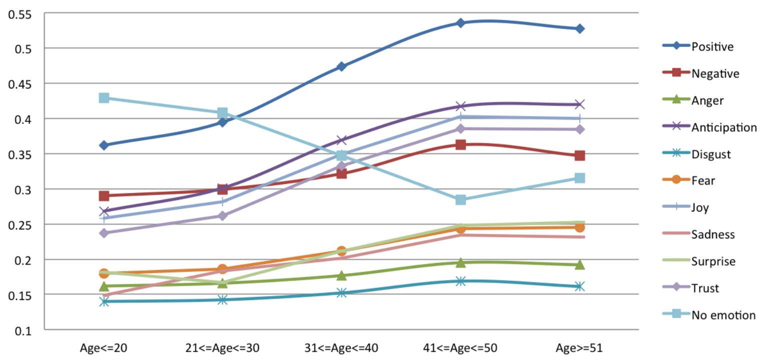

Fig. 3: Probability of occurrences of emotions in status updates from users of different age groups. Users are more likely to post emotions as they get older.

Based on Figure 3, the probability of expression of emotions increases with age. Users post more positive emotions as they get older. We find that older users are more emotional in their posts compared to younger users. Users between 40 to 50 years old have the smallest amount of status updates without emotion expression (less than 30%), which indicates their willingness to share their feelings. On the other hand, more than 40% of young users' posts (users less than 21 years old) are without emotions. This evidence could be caused by their language use and the fact that our dictionary does not contain all possible expressions.

3.3 Emotion and personality

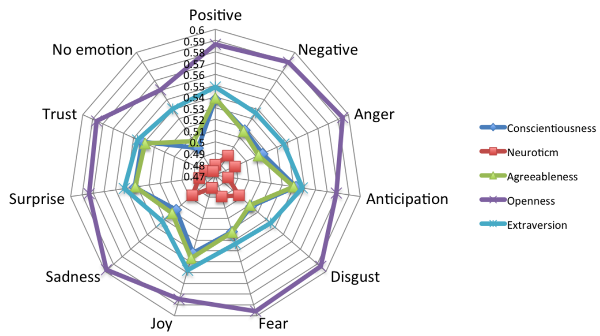

Fig. 4: Probability of occurrences of emotions in status updates from users with different positive personality traits.

Similarly as in the previous sections, for each of the personality traits, we consider the set of status updates written by users who meet the threshold for that personality trait according to Table 1. Using S_p to denote the set of status updates linked in this way to personality trait p, we compute $P(S_p|S_i) = \frac{|S_i \cap S_p|}{|S_i|}$ with S_i (for $i = 0, 1, \ldots, 10$) defined as in the beginning of Section 3. The results are visualized in Figure 4. Similarly, results of $P(\neg S_p|S_i) = \frac{|S_i \cap \neg S_p|}{|S_i|}$ are visulaized in Figure 5.

Neurotic users' posts are less likely to be emotional, while open users' posts convey emotions more frequently than other personalities. After open users, extrovert users express the most emotions in their posts. Interestingly, agreeable users express emotions very similar to conscientious users on Facebook.

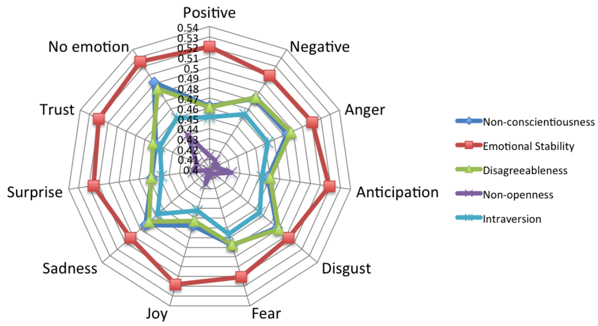

Fig. 5: Probability of occurrences of emotions in status updates from users with different negative personality traits.

Posts containing anticipation are mostly expressed by agreeable, conscientious and extrovert users. Neurotic users use less joy expressions than other personalities and their posts are most likely about disgust, sadness and negative feelings. Sadness appears more than other emotions for neurotic and open users, while joy emotions are expressed most by extrovert, conscientious and agreeable users. Open users also post frequently about their fear and anger.

3.4 Emotion and time

In this section, we investigate the relation between emotion expression and the time stamp of the posts. The graphs in this section depict the conditional probabilities of emotion expression w.r.t. time using $P(S_i|S_t) = \frac{|S_i \cap S_t|}{|S_t|}$, where S_t is the set of status updates posted in a specific time interval. In Figure 6, there are 7 such time intervals, each one corresponding to a day of the week. In Figure 7, the time intervals correspond to the months of the year.

Emotion and day of the week Figure 6 presents that people are more emotional during workdays than weekends. During the weekend (on Saturday and

Sunday), users are less emotional and their posts are more likely without emotion expression. Among other things, the number of posts about trust decreases during the weekend, and a similar observation holds for posts related to fear.

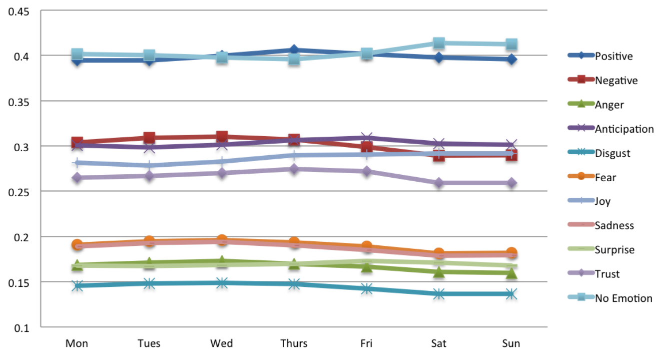

Fig. 6: Probability of occurrences of emotions in status updates depending on the day of the week. Status updates are more likely to contain emotions during workdays than during the weekend.

Facebook status updates are most likely to convey emotions on Thursday. From Friday onwards, the probability of emotion expression decreases. On Saturdays, users are least likely to express any emotions in their posts. Interestingly, the frequency of status updates conveying anger and surprise remains constant from Monday to Thursday. However, on Friday, users express more surprise and become less angry in their posts. In addition, users are more negative during the workdays and less likely express to joy. However, on Saturday and Sunday, users become less negative and more joyful.

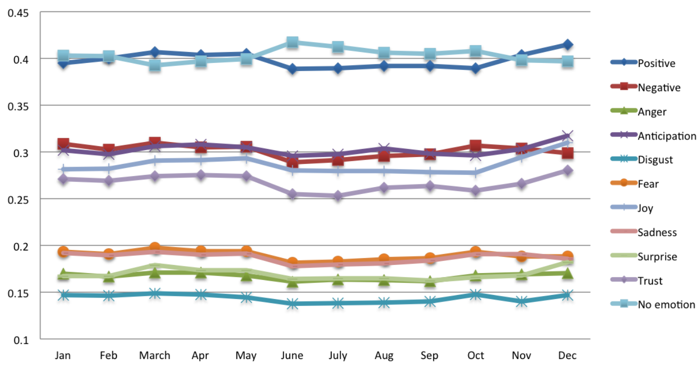

Fig. 7: Probability of occurrences of emotions in status updates depending on the month of the year. Status updates are most likely to contain emotions in December.

Emotion and month of the year Figure 7 presents the probability of emotion expression during different months of the year. Facebook users are more emotional in December; in particular, users are less negative, more joyful, surprised, anticipating and positive compared to other months of the year. This is reflected in posts such as *"Happy holiday", "Happy NYE", "Happy Christmas"* which are very prominent in December and which are tagged as emotion conveying posts by the emotion detection method described in Section 2.2. Although there are no significant changes in emotions during the rest of the year, during the summer months (June, July and August), the amount of positive, fear and trust expressions decreases, and users' posts are least likely to contain any emotion.

3.5 Correlations between features and emotions

Table 2: Pearson Chi Squared test results for on characteristics of users and posts, and emotion categories: *positive (Pos)*, *negative (Neg)*, *anger (Ang)*, *anticipation (Ant)*, *disgust (Dis)*, *fear (Fea)*, *joy (Joy)*, *sadness (Sad)*, *surprise (Sur)*, and *trust (Tru)*.

Features	Pos	Neg	Ang	Ant	Dis	Fea	Joy	Sad	Sur	Tru
Gender	**0**	**0**	**0**	**0**	**0**	0.205	**0**	**0**	**0**	**0**
Age	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Open	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Conscientious	**0**	0.014	0.793	**0**	**0**	**0**	**0**	0.496	**0**	**0**
Extrovert	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Agreeable	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.065	**0**	**0**
Neurotic	0.613	**0**	**0**	0.015	**0**	**0**	**0**	**0**	**0**	0.249
Monday	**0.001**	0.023	0.146	0.050	0.058	0.333	**0**	0.105	0.055	0.081
Tuesday	**0.001**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.029	0.879
Wednesday	0.213	**0**	**0**	0.137	**0**	**0**	**0.001**	**0**	0.220	**0.001**
Thursday	**0**	**0**	**0.008**	**0.002**	**0**	**0**	**0.002**	**0.001**	0.482	**0**
Friday	0.019	0.029	0.517	**0**	0.139	0.566	**0.001**	0.047	**0**	**0**
Saturday	0.437	**0**	**0**	0.891	**0**	**0**	**0**	**0**	0.104	**0**
Sunday	0.029	**0**	**0**	0.170	**0**	**0**	**0**	**0**	0.200	**0**
January	0.039	**0**	0.016	0.704	**0.007**	**0.003**	**0**	**0**	0.066	**0.004**
February	0.432	0.377	0.740	**0.001**	0.035	0.322	**0.006**	0.082	0.094	0.139
March	**0**	**0**	**0.001**	0.019	**0**	**0**	**0.004**	**0**	**0**	**0**
April	**0.003**	0.027	**0.002**	**0.002**	**0.001**	**0.002**	**0.004**	0.025	**0.002**	**0**
May	**0.001**	0.016	0.465	0.269	0.623	**0.004**	**0**	**0.003**	**0.003**	**0**
June	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
July	**0**	**0**	0.011	**0.001**	**0**	**0**	**0**	**0**	**0**	**0**
August	**0**	**0**	**0.001**	0.424	**0**	**0**	**0**	**0**	**0**	**0.001**
September	**0**	0.014	**0**	**0.004**	**0.003**	0.016	**0**	**0.008**	**0**	0.027
October	**0**	**0**	0.614	**0**	**0.001**	**0.003**	**0**	**0.009**	**0.004**	**0**
November	**0.001**	0.130	0.126	0.760	**0.002**	0.410	**0**	**0.003**	0.250	0.513
December	**0**	0.113	**0.004**	**0.004**	0.004	0.211	**0**	0.195	**0**	**0**

To assess the relation between the different features and emotions, we apply the Pearson chi-squared dependence test [15]. Table 2 presents the p-values. The null hypothesis is that features and emotions are independent. The p-values that are lower than the significance level ($p < .01$) denote significant correlations of features with emotions. They are indicated in bold in the table.

Gender is related with all emotion categories except fear. Age is shown to be related to all emotion types. Similarly, Openness and Extroversion are related with all emotion types. Conscientiousness is related to anger, negative and sadness emotions. Agreeableness is not related to sadness. And finally, Neuroticism shows no relation with positive, anticipation and trust emotions.

4 Conclusion and Future Work

In this study, we explored the relation between the emotions of 5,865 Facebook users with their age, gender and personality by using their status updates (almost 1 million posts). We used the NRC hash-tag emotion lexicon to detect emotions from the posts. We also extracted temporal features from the posts' time stamps. Almost 60% of status updates contain at least one type of emotion expression. Positive emotion is expressed with the highest frequency in status updates and disgust is least likely to appear in the status updates of users.

The results confirm a relation between users' characteristics and their emotions. Similar to offline expression, female Facebook users express more emotions in their status updates than male users. Similarly, older users express more emotions in their status updates than younger users. Neurotic users are not very emotional in their status updates, while open users are mostly likely to express their feelings about different subjects. By analyzing the time stamp of the status updates, we examined relations between Facebook posts' time and users' feelings. Interestingly, emotions are more likely to be expressed during the workdays compared to the weekend. The frequency of emotional status updates is lowest during the summer and highest in December.

We found significant correlations between our selected features and users' emotions. In future research, we will develop a model that will predict the most probable upcoming emotion for each user, among other things based on time, demographics and personality. We believe that being able to predict users' emotions and target the end users accordingly would be useful for personalized services.

Aside from the work we have presented in this paper, there is clear potential for more fine grained emotion detectors. Emotion detection in this study has been performed using a lexicon based approach. However, due to the complexity of the status updates, the limited size of the lexicon, and a huge amount of noise in the unnormalized status updates, it is very likely that we have missed many emotion expressions in the status messages. Exploring better techniques to extract emotions not only based on the words, but also based on other features is potentially an open path to explore.

Acknowledgement: This project is part of the SBO Program of the IWT (IWT-SBO-Nr. 110067).

References

1. Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

2. Gediminas Adomavicius and Alexander Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.
3. Johan Bollen, Huina Mao, and Alberto Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM*, 2011.
4. Iván Cantador, Ignacio Fernández-Tobías, Alejandro Bellogín, Michal Kosinski, and David Stillwell. Relating personality types with user preferences in multiple entertainment domains. In *Proceedings of the 1st Workshop on Emotions and Personality in Personalized Services (EMPIRE 2013), at the 21st Conference on User Modeling, Adaptation and Personalization (UMAP 2013)*, 2013.
5. Paul T Costa and Robert R McCrae. The revised NEO personality inventory (NEO-PI-R). *The SAGE Handbook Of Personality Theory And Assessment*, 2:179–198, 2008.
6. Golnoosh Farnadi, Susana Zoghbi, Marie-Francine Moens, and Martine De Cock. Recognising personality traits using facebook status updates. In *Proceedings of the workshop on computational personality recognition (WCPR13) at the 7th international AAAI conference on weblogs and social media (ICWSM13)*, 2013.
7. Jennifer Golbeck, Cristina Robles, and Karen Turner. Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 253–262. ACM, 2011.
8. Lewis R Goldberg, John A Johnson, Herbert W Eber, Robert Hogan, Michael C Ashton, C Robert Cloninger, and Harrison G Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96, 2006.
9. Ai Thanh Ho, Ilusca LL Menezes, and Yousra Tagmouti. E-mrs: Emotion-based movie recommender system. In *Proceedings of IADIS e-Commerce Conference. USA: University of Washington Both-ell*, pages 1–8, 2006.
10. Michal Kosinski, Yoram Bachrach, Pushmeet Kohli, David Stillwell, and Thore Graepel. Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, pages 1–24, 2013.
11. Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.
12. Saif M Mohammad and Svetlana Kiritchenko. Using nuances of emotion to identify personality. *arXiv preprint arXiv:1309.6352*, 2013.
13. Saif M Mohammad and Peter D Turney. Nrc emotion lexicon. 2013.
14. Ante Odic, Marko Tkalcic, Jurij F Tasic, and Andrej Košir. Relevant context in a movie recommender system: Users opinion vs. statistical detection. *ACM RecSys*, 12, 2012.
15. Robin L Plackett. Karl pearson and the chi-squared test. *International statistical review*, 51(1):59–72, 1983.
16. Peter J Rentfrow and Samuel D Gosling. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236, 2003.
17. 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. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791, 2013.