Using Twitter Data to Infer Personal Values of Japanese Consumers

Yinjun Hu
Synergy Marketing, Inc.
Dojima-Avanza 21F, 1-6-20 Dojima,
Kita-ku, Osaka 530-0003, Japan.
ko.inshun@syngery101.jp

Yasuo Tanida
Synergy Marketing, Inc.
Dojima-Avanza 21F, 1-6-20 Dojima,
Kita-ku, Osaka 530-0003, Japan.
tanida.yasuo@synergy101.jp

Abstract

Our purpose is to use Twitter data to infer personal values in marketing for Japanese consumers. In this paper, we reintroduce our personal value system and apply the model for inferring personal values with tweets. To adapt the model to the rapid change of wording in tweets, we propose a dynamic model based on time-weighted frequency in this research. We evaluated the prediction results from our previous approach, newly proposed approach (the dynamic model), and other methods with 10-fold cross-validation. Our experiment results show that personal values can be inferred from Twitter data, and our approach based on Bayesian network performs well with skewed training data.

1 Introduction

In marketing science, personal values have been considered the central determinants of consumer behavior. They are widely used for market segmentation and behavioral prediction. VALS (Values And Lifestyles) is a personal value system for market segmentation,[1] and (Wu, 2005) reconstructed it for Chinese consumers as China-Vals. Similarly, we developed a value system for Japanese consumers based on the *AIO (Activities, Interests, and Opinions)* (Plummer, 1974) rating statement during last two years. We gathered about 20,000 Japanese consumers' personal values data via questionnaires, and filled the value system database with this data.

[1]VALS is developed by Mitchell, Arnold (May 1984) in the book *Nine American Lifestyles: Who We Are and Where We're Going*.

Although we can "talk to" consumers directly with a questionnaire approach, subjective biases may appear in answers, such as a response-bias (Peer and Gamliel, 2011) and a choice-supportive bias (Mather et al., 2000). On the other hand, a data mining approach like microblog analysis, which is considered to have a sampling bias problem, as checked by (Mislove et al., 2011), can get objective "answers" for the "questions" without subjective biases. Hence, a microblog mining approach, like mining tweets, may be a complement to the questionnaire approach. Moreover, (Chen et al., 2014) pointed out that word use may be influenced by values, and made an effort to analyze the associations between personal values and their word uses in social media. In our research, we use Twitter for a data mining approach to infer personal values, because we consider that users on Twitter tend to tweet their real intentions with limited impersonation.

In our previous work, we proposed a model for extracting personal values from tweets with text mining technologies. This model demonstrated that we were able to predict consumer behavior with tweets and our value system. As a result, we applied it to marketing consulting and social contribution for trial-and-error. However, the wording on Twitter changes frequently, and the keywords used for inference in this model may be out of date. Hence, in this research, we propose a dynamic model which can update itself automatically. In addition, a detailed methodological comparison between our proposed model and other methods is discussed from the experiment results.

29th Pacific Asia Conference on Language, Information and Computation pages 480 - 487
Shanghai, China, October 30 - November 1, 2015

2 Methodology

2.1 A Value System for Japanese Consumers

To determine the latent benefits and interests of Japanese consumers, we proposed a value system for Japanese people during previous two years. The system contains 61 components *(value component, VC)* covering 8 frames of personal values extracted from 20,000 questionnaires with principal component analysis (PCA). However, an abbreviated and more flexible version of this system with 22 components (mini version) was preferred to use, because the questionnaire of mini version has only 60 questions as opposed to 303 questions for the full version. Table 1 shows details of the 22 components above. The instance of each component can be a binary value (i.e., 0 or 1). In addition, we defined 12 social types named *Societas* from full version with Ward's method, and trained the Societas Model with the Bayesian network (Pearl, 1985). We also take the word *Societas* in the wide sense of the value system we proposed.

Frame Name	**Component Name**
Character	Curiosity, Delicacy, Laxity, Cooperativeness
Positive	Narcissism, Self-realization
Negative	Sensitivity to criticism, Sensitivity to lack of common sense, Sensitivity to disappointment
Human relationship	Stress, Friendship emphasizing
Family relationship	Marriage aspiration, Spousal responsibility as housewife, Family discord
Sense of job	Satisfaction, Stress
Sense of money	Lack of money, Savings, Sufficiency
Sense of time	Priority to family, Sufficiency, Lack of time

Table 1: The detail of 22 value components in the mini version of Societas.

2.2 Societas Inference with Twitter Data

Model Our previous work proposed a model called *TwitterSocietas* for inferring Societas values from Twitter data. We made an effort to construct a TwitterSocietas Model with Bayesian network and the training data is extracted as follows.

1. Obtain Societas values (mini version) and Twitter id via questionnaires. Then, make a unique word set W from tweets of all users, and refine W with a DF(document frequency, we treat each Twitter user's tweets as a document) between 10% and 90%.

2. Calculate importance scores for each <word, VC> paired with the following formula.

$$|P(w \in W_i|V_j = 1) - P(w \in W_i|V_j = 0)|$$

Where, W_i represents the unique word set extracted from Twitter user i, V_j represents the binary set of value component j for all users calculated from the questionnaires, and w is the word in W.

3. For each value component, sort the importance score pairs decreasingly, and extract the top 30 words as the feature keywords *(speech keyword)* of the value component.

4. For each <speech keyword, VC> pair, calculate the relevant coefficients by Algorithm 1.

5. Calculate the binary data *(speech component, SC)* of speech keywords for each value component with the relevant coefficients as:

$$f(SC) = \begin{cases} 1 & \text{if } \vec{kw} \cdot \vec{pca} - \bar{pca} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where $\vec{kw}$ represents the vector of speech keyword appearance in Twitter user's vocabulary W_i, $\vec{pca}$ represents the relevant coefficient(i.e., $pca1$ or $pca2$ in Algorithm 1), and $\bar{pca}$ is the mean for all Twitter users.

6. Merge speech components and values components into the training data of TwitterSocietas Model IN the format of <SC1 for VC1, SC2 for VC1, SC1 for VC2, ..., VC1, VC2>

Notice that only the approach of inferring 22 value components is mentioned here because these components can be seen as the characteristics of 12 social types.

Algorithm 1 Relevant coefficients generation.

Input: speech keywords KW for value component j, word set $\{WC\}$ for all tweets, and the binary set V_j of value component j.

Output: Relevant coefficients data.

1: $WT \Leftarrow$ matrix of $length$(Twitter users) $*$ 30 elements and initialize all elements to 0
2: **for** $i = 1$ to $length$(Twitter users) **do**
3: **for each** kw in KW **do**
4: **if** $KW[k]$ in $WC[i]$ **then**
5: $WT[i][kw] \Leftarrow 1$
6: **end if**
7: **end for**
8: **end for**
9: $pca \Leftarrow PCA(WT, V_j)$
10: **return** 30 pairs of <pca1, pca2>, where pca1 and pca2 represents the weights of the first principle component and the second one from pca for each speech keyword.

Inference We use the same approach of step 5 above to generate the evidence from object Twitter data, and the evidence contains 44 speech components (i.e., each value component has 2 speech components) in binary. To infer the personal values with 44 speech components, we employ the Loopy Belief Propagation algorithm in (Weiss, 2000) . As a result, the inferred data of value components are probabilities so that they can be used more flexibly than binary ones.

2.3 A Dynamic TwitterSocietas Model

As Twitter data is updated frequently, the words used in TwitterSocietas have an *aged problem* as soon as the tendency of wording in tweets changes. Furthermore, we assume that wording changes more frequently than personal values, so that recent tweets may be related to users' personal values more deeply than older ones. Hence, we propose an auto-updating TwitterSocietas Model in which recent tweets are weighted.

- Firstly, we define a weight of the words in each tweet as:

$$W1(w,t) = \sum_{i=1}^{N} \frac{C-i+1}{C} \times f(w,i) \quad (2)$$

$W1(w,t)$ is a function to calculate the weight of word w for Twitter user t. Where i represents the newness of tweet, i.e., $f(w,i)$ in (2) is related to the latest tweet of user t when i equals to 1. $f(w,i)$ is the relative frequency of word w in user t's ith newest tweet, and can be calculated as follows.

$$\frac{\text{frequency of word } w \text{ in user } t \text{ 's } i\text{th newest tweet}}{\text{total frequency of word } w \text{ in user } t \text{ 's all tweets}}$$

In equation (2), C is a constant and N is the amount of tweets for user t. Let C be 2,000 and $N \leq C$, i.e., N will be set to 2,000 when the amount of user t's tweets is more than 2,000. This is because 2,000 tweets are adequate for inferring personal values according to our preliminary experiment.

- Refine the weighting as follows.

$$W2(w,t) = \begin{cases} 1 + w_1 & \text{if } w_1 > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Here, w_1 is an enumeration of $W1(w,t)$ in equation (2).

- Take the weight $W2(w,t)$ calculated by equation (3) instead of word appearance at the formula of Step 2 in §2.2 and "1" at line 5 in Algorithm 1.

- Similarly, we let vector $\vec{kw}$ in equation (1) initialize with the weights calculated by (3) instead of the binary data.

3 Experiment

3.1 Preliminary

Morphology Unlike English, Japanese sentences are written in Chinese characters and kana (Japanese alphabet) without delimiters (i.e., space) between words. We used *MeCab* [2] to tokenize the tweets into words. Moreover, we included *List of all page titles* in the Japanese Wikipedia[3] as an additional dictionary into MeCab, in order to solve new named entities (i.e., book name, software name, etc.).

[2] MeCab is an open-source morphological analyzer software. http://taku910.github.io/mecab/

[3] http://dumps.wikimedia.org/jawiki/

	Curiosity				Cooperativeness				Savings			
Method	**A(%)**	**P**	**R**	**F**	**A(%)**	**P**	**R**	**F**	**A(%)**	**P**	**R**	**F**
SVM	61.5	0.63	0.85	0.72	61.9	0.56	0.36	0.44	82.3	0.00	0.00	0.00
NB	60.0	0.60	0.97	0.74	63.4	0.66	0.24	0.35	82.3	0.00	0.00	0.00
DTS	57.3	0.66	0.58	0.62	55.0	0.46	0.63	0.53	55.9	0.20	0.50	0.28

Table 2: The evaluation. The average of ten 10-fold cross-validation of inferred Societas values.

TF-IDF and LSA To give weight to words without ignoring zero idf terms, we used TF-IDF as:

$$tfidf(w) = tf \times (idf + 1)$$

Where tf represents the term frequency of word w, and idf is the inverse document frequency of word w. In addition, we applied LSA (latent semantic analysis) (Deerwester et al., 1990) to the TF-IDF values from Twitter corpus, and reduce the data to 1,000 dimension.

Baseline We employed a support vector machine (SVM) approach as the baseline in our experiment.[4] Moreover, Naive bayes classifier was employed as an alternative choice for classifying Twitter data.[4] According to (McCallum and Nigam, 1998), multinomial Naive Bayes (NB) is more suitable for context with large vocabulary size, and we considered that our training data (vocabulary size $> 10k$) fits this condition.

Bayesian Network As interpreted in §2, the Bayesian network is employed for our proposed method.[5] We experimented the Dynamic Twitter-Societas Model mentioned in §2.3 (DTS). From the preliminary experiment, it was expected that performance of TwitterSocietas Model would be similar to DTS, however, a DTS method solved the aged problem as mentioned in §2.3. To compare with other methods, we used the means of training data (training means) for discretization. For example, even if a probability of $vc = 1$ (vc is a value component) is close to 0.0 such as 0.02, its binary discretization can also be 1 when the training means of this component is smaller (i.e., 0.01).

[4]We used Machine learning toolkit *scikit-learn* in our experiment. `http://scikit-learn.org/stable/`

[5]We used "Bayonet", a Bayesian network software, to make the Societas and TwitterSocietas models. `https://staff.aist.go.jp/y.motomura/bayonet/`

Component	VC = 0	VC = 1
Curiosity	0.40	0.60
Cooperativeness	0.59	0.41
Savings	0.82	0.18

Table 3: Means of the value components for 1,147 Twitter users.

Dataset The dataset used in our experiment consisted of two subsets: Societas data and tweets, both related to 1,147 Twitter users.

Evaluation Metric We used accuracy (A), precision (P), recall (R), and F-measure (F) with 10-fold cross-validation to evaluate the performance of our proposed method and other methods. We treated the prediction result as the retrieved documents, the training data as the relevant documents, and "1" was the appearance of personal value.

3.2 Evaluation

Results Three representative personal values (*Curiosity*, *Cooperativeness*, *Savings*) were selected for discussion and the evaluation is shown in Table 2. Notice that the results of SVM and NB in Table 2 only involved the words between the DF of 10% and 90% when constructing feature vectors from the words in tweets. This is because keywords used in our proposed method were also extracted in this way as mentioned in §2.2. In addition, we normalized the feature vectors with TF-IDF and LSA for SVM and NB.

Comparison From Table 2 we can see that both SVM and NB performed better in terms of accuracy than DTS. However, DTS had better scores of recall and F-measure in cooperativeness and especially in *Savings*. This is due to the skewed distribution of the value components (i.e., $P(VC = 0) >> P(VC = 1)$) in our training data as shown in Table 3. DTS

Component	VC = 1	SVM: A(%) / F	NB: A(%) / F	DTS: A(%) / F
Delicacy	0.54	55.7 / 0.63	57.9 / 0.67	56.5 / 0.60
Laxity	0.61	64.2 / 0.74	62.8 / 0.76	54.5 / 0.59
Narcissism	0.46	66.3 / 0.61	66.4 / 0.68	62.9 / 0.62
Self-realization	0.56	53.9 / 0.64	55.8 / 0.71	53.9 / 0.58
Sensitivity to criticism	0.39	60.7 / 0.31	60.1 / 0.02	**50.4** / 0.42
Sensitivity to lack of common sense	0.61	58.9 / 0.73	60.0 / 0.75	52.7 / 0.60
Sensitivity to disappointment	0.39	59.8 / 0.22	61.6 / 0.08	**49.3** / 0.42
Stress(Human relationship)	0.50	52.7 / 0.51	53.3 / 0.52	**49.7** / 0.47
Friendship emphasizing	0.56	52.5 / 0.63	54.7 / 0.68	55.5 / 0.61
Marriage aspiration	0.57	65.5 / 0.72	62.1 / 0.72	59.5 / 0.63
Spousal responsibility as housewife	0.41	61.9 / 0.44	60.6 / 0.37	54.2 / 0.48
Family discord	0.41	58.8 / 0.35	58.7 / 0.03	**49.5** / 0.50
Satisfaction	0.47	60.3 / 0.53	57.5 / 0.43	58.0 / 0.58
Stress(Sense of job)	0.47	52.9 / 0.43	52.2 / 0.23	53.6 / 0.51
Lack of money	0.63	62.7 / 0.75	63.0 / 0.77	55.1 / 0.61
Sufficiency(Sense of money)	0.28	71.6 / 0.00	71.8 / 0.00	54.0 / 0.36
Priority to family	0.28	75.2 / 0.24	72.1 / 0.02	60.4 / 0.44
Sufficiency(Sense of time)	0.44	56.8 / 0.37	56.4 / 0.14	**50.8** / 0.45
Lack of time	0.32	68.6 / 0.11	68.2 / 0.00	56.2 / 0.39

Table 4: Evaluation results for other personal values with accuracy(A) and F-measure(F) metrics.

generates a posterior probability of the value component allowing conversion to binary with the most suitable threshold. Table 4 shows the evaluation results for other personal values that were not shown in Table 2, and we find that a Bayesian Network approach may be not appropriate for inferring the personal values corresponding to accuracy scores in bold font because they are as poor as random guessing.

4 Discussion

4.1 Speech keywords

We believe that every morphological element represents an aspect of the user's personal value on Twitter. Hence, dictionary form words without normalization are preferred in our research. For example, to infer personal values, *tsumetai* is different from a variation word use with *kaomoji* (the Japanese emoticon) like *tsumetatt(°∀° ⊂)⊂*, and this is similar to the relation between *cool* and *Coooooooooooooooollllllllllll!!!!!!!!!!!!!!* in (Brody and Diakopoulos, 2011). This is another interesting characteristic of the TwitterSocietas Model, and was also used for sentiment analysis in (Barbosa and Feng, 2010).

4.2 Application of TwitterSocietas Model

In past years, we have succeeded to apply TwitterSocietas Model to infer personal values for marketing researches. For example, we explained TV viewers' characteristics with personal values inferred by TwitterSocietas Model to a TV program which has no demographic distribution data about it but has many Twitter followers (Fujii et al., 2013).

In addition, we succeeded to find the sudden change of personal value "Narcissism" (from a higher probability a lower one) for a blogger, during the period of her mother's death from Alzheimer's Disease (Tanida and Tokumi, 2014). For a text mining interest, we found that TwitterSocietas Model can also be applied to blog data. And we hope that in the future, we can acquire knowledge from online text data with text mining technologies, such as Twitter data or blog data, for giving some mental support to the people with dementia family members.

4.3 Multilinguistic Societas

In our recent work, we started a project on finding out the relations between text data in "Weibo" (A kind of Chinese microblogs) and personal values. Up to now, we have gathered 1,002 pieces of ef-

Component	Weibo N=1,002	Twitter N=1,147
Curiosity	0.76	0.60
Delicacy	0.49	0.54
Laxity	0.38	0.61
Cooperativeness	0.43	0.41
Narcissism	0.69	0.46
Self-realization	0.70	0.56
Sensitivity to criticism	0.63	0.39
Sensitivity to lack of common sense	0.47	0.61
Sensitivity to disappointment	0.28	0.39
Stress(Human)	0.36	0.50
Friendship emphasizing	0.71	0.56
Marriage aspiration	0.63	0.57
Spousal responsibility as housewife	0.64	0.41
Family discord	0.30	0.41
Satisfaction	0.69	0.47
Stress(Job)	0.53	0.47
Priority to family	0.46	0.28
Sufficiency(Time)	0.52	0.44
Lack of time	0.34	0.32
Lack of money	0.56	0.63
Sufficiency(Money)	0.36	0.28
Savings	0.35	0.18

Table 5: Societas personal values (arithmetic mean) of Weibo users and Japanese Twitter users.

fective data for our "Weibo-Societas" Model. However, we found that for Chinese people (especially who use Weibo), some personal values are very different from Japanese Twitter users' ones. Table 5 shows the arithmetic mean of Societas personal values about Weibo users and Japanese Twitter users. As shown in Table 5, Weibo users tend to have the value "Curiosity" than Japanese Twitter users. However, Japanese Twitter users may be more delicate than Weibo users.

5 Related Work

5.1 Demographic Inference for Twitter users

The earlier work by (Zamal et al., 2012) proposed an approach of Twitter users' latent attributes inference including gender, age, and political affiliation with Twitter. Similarly, (Ciot et al., 2013) made an effort to infer gender with non-English-based content and users, and (Bergsma and Durme, 2013) enabled substantial improvements on the task of Twitter gender classification. Moreover, (Beller et al., 2014) proposed a method to predict social roles such as doctor, teacher, etc. These attributes are considered as demographic attributes, and they are very important to market segmentation. In our research, we contributed to the inference of Twitter user's personal values which are also essential factors to marketing science and consumer behavior prediction.

5.2 Personal values Inference for Twitter users

For personal value inference, (Quercia et al., 2011) contributed to personality prediction with Twitter profiles based on Big Five. (Golbeck et al., 2011) provided a method to infer Twitter users' personality of Big Five with the statistics about their accounts and tweets. To infer the personality of Chinese people, (Bai et al., 2013) developed a method to infer the Big Five personality from "Weibo" data with a multivariate regression approach. In our research, we use the words in tweets to calculate the speech components (as mentioned in §2.2 Step 5) for each user, and these speech components can be used for inferring Societas personal values with TwitterSocietas Model. Moreover, we think that personal values are interactive to each other. As a result, we constructed our TwitterSocietas Model with a Bayesian Network approach.

(Sumner et al., 2012) attempted predicting personality traits from the lexicon features extracted tweets. Moreover, (Plank and Hovy, 2015) made an effort on inferring MBTI[6] personality type from tweets, gender and some meta-features (i.e., counts of tweets, followers, etc), and showed that social media can provide sufficient linguistic evidence to reliably predict some dimensions of personality. However, they suggested that it is hard to predict the personality dimensions of Judging/Perceiving with the linguistic evidence from tweets. Similarly, as shown in Table 4, we find that some Societas personal values, i.e., "Stress(Human)", are hard to predict from tweets no matter which classifier is employed.

In (Plank and Hovy, 2015), the words in tweets used as lexicon features were transformed into binary word n-grams. In our previous work, we also used the words in binary (whether the words appeared in users' tweets or not), and counted the co-occurrence between Societas personal values and these words, so that we can select sensitive words for each personal value to construct our TwitterSocietas Model. However, in this paper, we incorporated the

[6]MyersBriggs Type Indicator (MBTI) is a way to measure how people perceive the word and make decisions.

relative frequency of words in each tweet as mentioned in §2.3.

6 Conclusion and Future work

In this paper, we introduced our personal value system and its extension *TwitterSocietas*. Furthermore, we proposed a dynamic way to update the TwitterSocietas Model automatically based on Twitter data. We evaluated the performance of the model and compared it with SVM and NB. As a result, we found that our approach accurately inferred personal values with skewed training data.

In the future work, we will first check whether the Weibo-Societas Model, a derivation of Twitter-Societas, can be applied to predict personal values of Chinese people. Furthermore, we will try to implement the source of Societas personal values inference with other features, such as whether a link or a picture is contained in a tweet, the importance of a mentioned event (i.e., a word after "#" mark).

Acknowledgments

We wish to thank the anonymous reviewers for their valuable comments and suggestions.

References

Shuotian Bai, Bibo Hao, Ang Li, Sha Yuan, Rui Gao, and Tingshao Zhu. 2013. Predicting Big Five personality traits of microblog users. In *Proceedings of Web Intelligence (WI) and Intelligent Agent Technologies(IAT), 2013 IEEE/WIC/ACM International Joint Conferences*, 1:501–508.

Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on Twitter from biased and noisy data. In *Proceedings of the International Conference on Computational Linguistics*, pages 36–44.

Charley Beller, Rebecca Knowles, Craig Harman, Shane Bergsma, Margaret Mitchell, and Benjamin Van Durme. 2014. I'm a belieber: social roles via Self-identification and conceptual attributes. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 181–186.

Shane Bergsma and Benjamin Van Durme. 2013. Using Conceptual Class Attributes to Characterize Social Media Users. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 710–720.

Samuel Brody and Nicholas Diakopoulos. 2011. Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP'11*, pages 562–570.

Jilin Chen, Gary Hsieh, Jalal Mahmud and Jeffrey Nichols. 2014. Understanding individuals' personal values from social media word use. In *Proceedings of the 17th ACM Conference on Computer-Supported Cooperative Work*, pages 405–414.

Morgane Ciot, Morgan Sonderegger, and Derek Ruths. 2013. Gender inference of Twitter users in non-English contexts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145.

Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

Emiko Fujii, Yinjun Hu, Mathieu Bertin, Yasuo Tanida. 2013. Understanding the audience of TV program from microblog with Social type model. *IEICE Transactions on Information and Systems*, 83:25–29.

Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. 2011. Predicting personality from Twitter. In *Proceedings of the 3rd IEEE International Conference on Social Computing*, pages 149–156.

Mara Mather, Eldar Shafir, and Marcia K. Johnson. 2000. Misremembrance of options past: Source monitoring and choice. *Psychological Science*, 11(2):132–138.

Andrew McCallum and Kamal Nigam. 1998. A comparison of event models for Naive Bayes text classification. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, pages 41–48.

Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist. 2011. Understanding the Demographics of Twitter Users. In *Proceedings of the 3rd International Conference on Weblogs and Social Media*.

Judea Pearl. 1985. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the Cognitive Science Society.*

Eyal Peer and Eyal Gamliel. 2011. Too reliable to be true? Response bias as a potential source of inflation in paper-and-pencil questionnaire reliability. *Practical Assessment, Research and Evaluation*, 16(9):1–8.

Barbara Plank and Dirk Hovy. 2015. Personality Traits on Twitter -or- How to Get 1,500 Personality Tests in aWeek. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 92–98.

Joseph T. Plummer. 1974. The concept and application of life style segmentation. *Journal of Marketing*, 38:33–37.

Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcrof. 2011. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In *Proceedings of the 3rd IEEE International Conference on Social Computing*.

Chris Sumner, Alison Byers, Rachel Boochever and Gregory. J. Park. 2012. Predicting Dark Triad Personality Traits from Twitter usage and a linguistic analysis of Tweets. In *IEEE International Conference on Machine Learning and Applications*, pages 386–393.

Yasuo Tanida, Rie Tokumi. 2014. Measuring a change of mind of dementia family caregivers. In *Proceedings of the 28th Annual Conference of the Japanese Society for Artificial Intelligence.*

Yair Weiss. 2000. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 12:1–41.

Yin Wu. 2005. The Research towards Model of China-Vals. *NanKai Business Review*, 8(2):9–15.

Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, pages 387–390.