Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference

A Four-Factor Model for Mining Consumer Insights in Social Data

David B. Bracewell
Oculus360
16301 Quorum Rd, Suite 100B
Addison, TX 75001
dbracewell@oculus360.us

Abstract

The voice of the customer is never more loudly heard than through social media. These online comments and reviews provide the insights marketers need to better build, design, and clarify the message around their products and services. Current approaches to mining these insights mainly focus on the volume and trend of sentiment. However, sentiment is not enough to discover actionable insights from these valuable social data. In this paper, we outline a four-factor model (*Attitudinal*, *Sociocultural*, *Personal*, and *Behavioral*) for mining consumer insights from social data that combines research in consumer and social psychology, discourse processing, and sentiment analysis. We present our current efforts in the automatic identification of a subset of the components making up these factors. In particular, we identify beliefs toward and about products and experiences, social actions in the form of recommendations, and intentions in the form of promises.

Introduction

Social media provides a medium for consumers to rant, rave, and recommend products, brands, and companies. For consumers, these data represent extra information that influences their decision on which product or brand to purchase and which company to patron. For companies, social data represents a potential for insights to the when, where, how, and why their products are used, who is buying them, who is using them, and information about those individuals including their associated beliefs, needs, wants, and preferences. These insights facilitate a marketer's understanding of consumer behavior, which allows them to better build, design, and market their products and services to meet consumers' needs and desires.

Traditionally, the volume and trend of positive and negative comments, reviews, tweets, etc (Pang, Lee, and Vaithyanathan 2002; Dini and Mazzini 2002; Smith, Fischer, and Yongjian 2012; Socher et al. 2013) is used as a proxy for brand awareness and placement against competitors, i.e. competitive intelligence. An evolution of this approach is aspect-based sentiment analysis in which sentiment is associated with aspects and higher-level aspect categories for a given target entity, e.g. product (Pavlopoulos and Androutsopoulos 2014b; 2014a).

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

However, aspects and sentiment alone do not fully capture the implicatures found in social data, which inform the attitudes and behavior of the consumers. For example, take the following review excerpts:

1. "I would **totally recommend any other laptop** over **this pile of garbage**."
2. "I know **my children needs (sic) to know computers to be successful**, but I just **can't afford** one."

In the first example the reviewer's negative sentiment for the laptop is strong enough that they recommend others to buy a different laptop. Recommending or suggesting a course of action is a directive speech act (Bunt 2011) and originates from directive modality (Bracewell, Hinote, and Monahan 2014). In this example, sentiment analysis would identify the polarity of the sentence as negative, but would ignore the greater implicature of the negative recommendation. This implicature indicates the loss of a customer and possible loss of potential customers within the reviewer's social network. In the second example the commenter expresses a need, relating to Maslow's (1943) cognitive need, for their children to have knowledge of computers. However, they are unable to meet this need because of the cost. Aspect-based sentiment analysis would identify a negative sentiment associated with "cost." However, the greater insight is the existence of a consumer with need and motivation, but without the ability to fulfill the need.

Capturing deeper insights, such as recommendations, preferences, and needs, requires concepts from dialogue processing and psychology to be combined with sentiment analysis. This synthesis results in a complete model from which all consumer related implicatures can be mined and transformed into actionable insights. In this paper, we present a four-factor model based on current work in consumer and social psychology, dialogue processing, and sentiment analysis, which is made up of: *Attitudinal*, *Sociocultural*, *Personal*, and *Behavioral* factors. These four factors, detailed later in the paper, inform the beliefs, needs, wants and preferences of consumers, the culture, social circle, status, role, and personal factors that define the consumers, and the motivations, intentions, and actions of the consumers as related to a product.

Strengthened by recent work on the discovery of social implicatures (Bracewell et al. 2011; Tomlinson et al. 2012) and personality (Schwartz et al. 2013) we believe that the four-factor model can be fully realized computationally. In this paper, we present our efforts in automatically identifying a subset of the model. In particular, we identify if sentences in online reviews contain linguistic manifestations of beliefs about/toward products and experiences, social actions in the form of recommendations, and intentions in the form of promises.

Related Work

Related research is found in the fields of consumer psychology, affective computing, and dialogue processing. Consumer psychology studies how thoughts, feelings, and perceptions influence the way individuals buy, use, and relate to products, services, and brands. Drawing from other areas in psychology, e.g. social psychology, consumer psychologists formalize the cognitive system of consumers using a categorical representation of products, services, brands and other marketing entities (Loken, Barsalou, and Joiner 2008). Supported by work on prototype theory, Loken and Ward (1990) find a link between the prototypicality of a product and consumers' affect toward it. Importantly, the categories making up of the cognitive system go beyond just product and brand to encompass goal-directed, cultural, and service categories among others.

A large volume of research exists in the area of affective computing of which sentiment analysis may be considered a component. Early work including that of Turney (2002) and Bracewell et al. (2006) relies heavily on techniques from information retrieval and information extraction. Machine learning techniques result in close to human-level ability for determining sentiment. In their seminal work on sentiment classification, Pang et al. (2002) examine the use of machine learning techniques to classify documents as positive or negative polarity. Common system errors in sentiment classification often are caused due to a lack of discourse level understanding (Pang and Lee 2004). More recent work starts to partially address the discourse gap. Yan et al. (2006) examine the use of semantic dependency analysis as a methodology for decomposing the sentiment in a sentence. Socher et al. (2013) examine the compositionality of sentiment by representing and determining the sentiment of sentences based on its structure.

A recent area of active research that is examining sentiment at a finer level is aspect-based sentiment analysis (Pontiki et al. 2014). The goal of which is to determine the sentiment toward aspects of a target entity, e.g. the screen of a TV or the food at a restaurant. The 2014 SemEval task (Pontiki et al. 2014) breaks down aspect-based sentiment analysis into four subtasks: aspect term extraction, aspect term polarity, aspect category detection, and aspect category polarity. The culmination of these four tasks is a system that can identify the aspects of a product as well as the more general category of the aspect (e.g. "too expensive" belongs to a "price" category) and discover the polarity (positive, negative, neutral, or conflict) toward the aspects and categories.

More general than aspect-based sentiment analysis and closely related to the current research is sentic computing (Cambria and Hussain 2012). Sentic computing synthesizes common-sense computing, linguistics, and psychology to infer both affective and semantic information about concepts. Cambria, Olsher, and Rajagopal (2014) show how SenticNet, a semantic and affective resource, can detect topics and determine polarity in patient opinions.

Extensive research has also been done for the construction of affective resources including corpora, dictionaries, and ontologies. Online reviews represent a rich source of emotion as most consumers have strong opinions about the products they use. Because of this, there a number of review related corpora, including movie reviews (Pang, Lee, and Vaithyanathan 2002; Maas et al. 2011), and Amazon product reviews (Uribe 2010). Corpora exist for other genres of textual data and for other languages as well. The MPQA corpus (Wiebe, Wilson, and Cardie 2005) is made up of news articles and includes annotations for beliefs, emotions, sentiments, and speculations. Minato et al. (2006; 2008) describe a Japanese-English bilingual affect corpus made up of text for English as a Second Language learners.

Lexical and semantic resources for sentiment analysis are often tied to Wordnets. WordNet-Affect provides A-Labels for a number of synsets in WordNet (Valitutti 2004). SentiWordNet (Esuli and Sebastiani 2006) assigns a score for positive, negative, and objective to all synsets in WordNet. Bracewell (2008; 2010) assign polarity and affect to a portion of WordNet according to Parrot's (2001) grouping of emotions.

A final area of related research is discourse parsing. In particular, we draw upon the work done in the automated identification of dialogue acts. Dialogue acts are specialized speech acts which include the internal structure, such as grounding and adjacency pairs, of a dialogue. Dialogue act schemes, such as DIT++ (Bunt 2011), define acts for questions, agreement/disagreement, promises, and directives among others.

The Four-Factor Model

Consumers need and desire products[1] and marketers target their products to consumers. In order to create products to fulfill consumer needs marketers must have insight into the consumer. We propose a four-factor model for gaining these valuable consumer insights. Our model, shown in Figure 1, draws heavily from research in consumer psychology, discourse processing, and sentiment analysis. The four factors inform to consumers' *attitudes*, *behavior*, *sociocultural* self, and *personal* qualities.

Attitudinal factors define consumers' beliefs, needs, wants, and preferences. Sociocultural Factors refer to the influence in decision making arising from the consumers' culture and group identity and their role and status in it. Personal factors include psychographics (e.g. personality) and demographics (e.g. age and gender). Behavioral factors inform to the motivations, intentions, actions and abil-

[1]We use the term product to refer to both a consumer good, e.g. laptop computer, and a commercial business, e.g. a restaurant

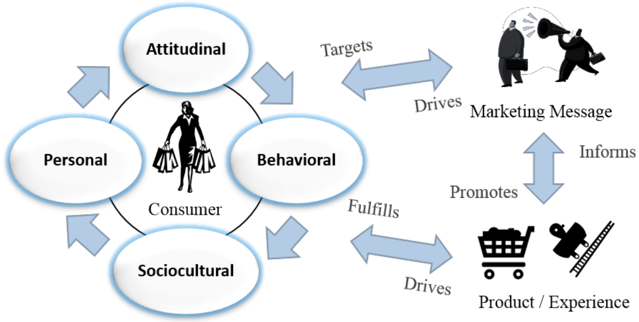

Figure 1: The Four-Factor model and its interaction with products and marketing messages.

ity to perform those actions, e.g. buy a new car. They blend together with each factor interacting and influencing the others (Triandis 1989; Chartrand and Bargh 1999; Ajzen 2005) and ultimately the beliefs, needs, and wants of a consumer (Bailey 2005).

Attitudinal

Attitudes represent a consumer's evaluation of a product or brand. They direct consumer behavior and are strong indicators of a brand or product's health and market activity (Hanssens et al. 2014). We define four attitudinal components, The first is **beliefs**, which are feelings held by a consumer about a product or brand. Beliefs may be positive (The screen is bright), negative (The price is too high), neutral (The TV is new), or contradictory (The TV has great features, but is built poorly). Beliefs can be modeled using aspect-based sentiment analysis and sentic computing techniques. The second component is **needs**, which are desires for a specific benefit, functional or emotional, from a product or service. Needs, as defined by Maslow (1943), are universal across cultures but the propensity of various needs may be culture dependent (Bracewell 2014). The third component is **preferences**, which define the likes and dislikes, i.e. tastes, of a consumer. Consumers' preferences drive their measure of a product's or service's utility. The final component is **wants**, which are the desires for products or services that are not necessary, but for which consumers wish.

Sociocultural

Sociocultural factors relate to influences of a consumer's culture and social circle on their personality, attitudes, lifestyle, and behavior. Social influences directly impact the behavior and attitudes of consumers (Bailey 2005). Marketers routinely use culture, or group, specific words to better relate their products to groups of consumers (Granot, Alejandro, and La Toya 2014).

We breakdown sociocultural factors into five components. The first is **cultural** relating to the geographical, historical, and familial influences on the consumer decision making process. Second is **acceptability**, which is the degree to which an action or product adheres to the norms of the consumer's social group, e.g. "eating meat" is unacceptable to "vegans." Third is **social status**, which is the relative status of the consumer within their social circle and in relation to the product or brand, e.g. Is a CEO of a fortune 500 more likely to buy a Nissan Versa or a Porsche? Fourth, is **social role**, which defines a consumer's role within their circle, e.g. "Trendsetter" or "Influencer." The role of the individual informs to their ability to activate their social network (Bracewell and Tomlinson 2013) and motivate or influence others. The final component is **social action**, which are actions by an individual to persuade, command, or call-to-action others in their social circle.

Personal

Personal factors represent the unique combination of personality, values, and morals that define an individual. Personality highly influences the behavior of consumers manifesting in among other things purchasing behavior and product choice (Kassarjian 1971). We breakdown personal factors into two main components (each of which can be further sub-categorized). The first is **psychographics**, which is the study of personality, values, opinions, attitudes, interests, and lifestyles. Personality traits defined in the Big Five model (McCrae and Costa 1987) are correlated with buying behavior (Myszkowski and Storme 2012). The second component is **demographics**, which define the characteristics of individuals such as their age, sex, sociocultural identity, organic systems, capabilities, etc.

Behavioral

Behavior is roughly defined as the actions taken in response to a stimuli. The stimuli in the case of consumer behavior can be from their needs and desires (internal) which may be influenced through marketing messages (external). The unconscious needs and desires of the consumer give rise to their motivations. These motivations evolve into intentions once the consumer desires to act. Consumers with the ability to fulfill their needs and desires will then act on their intentions.

We define four components for behavioral factors. The first is **motivation**, which is what drives consumers to identify and buy products or services that fulfill their conscious and unconscious needs and wants. The second is **intention**, which is a determination by the consumer to act in a certain way, e.g. switch products or remain loyal. Third is **ability**, which is the possession of the necessary skills or means to carry through with an intention, e.g. a teenager may intend to buy a Tesla, but most do not have the ability. The final component is **action**, which defines the actions performed by the consumer in regards to a company / product, e.g. purchasing.

Identifying Beliefs, Actions, and Intentions

We have just begun to computationally realize the four-factor model. Here we present our current effort for the automatic identification of the following components of the model: (1) Beliefs about/toward a product, which are a component of the *Attitudinal* factor; (2) Beliefs about/toward an experience, which are a component of the *Attitudinal* factor;

(3) Social actions in the form of recommendations and suggestions, which are a component of the *Sociocultural* factor; and (4) Intentions in the form of promises to purchase or not purchase a product again, which are a component of the *Behavorial* factor.

The SemEval 2014 task on aspect-based sentiment analysis (Pavlopoulos and Androutsopoulos 2014b; 2014a) provides polarity and aspect category annotations for roughly 3,841 sentences contained in online restaurant reviews. We use these annotations as a starting point for our current work by automatically mapping their aspect categories into *product* and *experience* focused beliefs. Product focused beliefs are beliefs toward/about attributes of a product, e.g. screen quality, cost, etc. We map the "price" and "food" categories into product focused beliefs. Experience focused beliefs relate to the procurement and consumption of products as well as the services, ambiance, and interactions related to a product. we map the "ambiance" and "service" categories into experience focused beliefs. We ignore the SemEval category of "anecdotes/miscellaneous." Additionally, we restrict our conversion to only categories ignoring the aspect terms.

We further extend the corpus with annotations of social actions in the form of recommendations and suggestions and intentions in the form of promises. We have also begun this annotation on the laptop reviews in the dataset, but in this paper limit our analysis to only the restaurant reviews. An example of a recommendation is: *I recommend the garlic shrimp , okra (bindi) , and anything with lamb.* An example of a promise is: *If this computer ever breaks down on me i will most definatly get the same one again.* In total there are 1,336 reviews containing a belief about an experience, 2,052 containing a belief about a product, 23 containing an intention, and 169 containing a social action. We believe that the smaller number of intention and social action annotations is a result of how the original data was generated and that with the complete review text we would find larger quantities. Besides the conversion and extension of annotations, no preprocessing or other modification to the SemEval corpus is done.

Each review can have zero or more of the components manifested, i.e. they may for example contain an intention and a belief. We construct a multi-label classifier made up of multiple one-versus-the-rest logistic regression classifiers. Individual classifiers use L2 regression with the default cost parameter of 1.0 via the LibLinear library (Fan et al. 2008). L2 regression was chosen as it results in near but not zero weights in its shrinkage whereas L1 regression tends to shrink most weights to 0 with only a few other non-zero features. We find that the weights from the L2 model are more easily interruptible and believed by end-users. We examine the following set of features based on results from the SemEval competition and our intuition:

- **N-Grams (1G, 2G, 3G)**: Unigrams (1G), bigrams (2G), and trigrams (3G) normalized to lowercase and weighted using Tf-Idf
- **Learned Vocabulary (V)**: A dictionary of category specific terms learned during training by ranking unigrams using log-likelihood and selecting the top N. Feature values are normalized using Tf-Idf.
- **WordNet Domains (WD)**: Combination of domain and surface form with values normalized by category within a document.
- **Psycholinguistic Categories (PC)**: Combination of general Inquirer (Stone et al. 1962) category and surface form with values normalized by category within a document.
- **Word Embeddings (WE)**: Average word vector for words in sentence using the 6b token 300 dimension trained Glove model (Pennington, Socher, and Manning 2014).
- **Deontic Modality (DM)**: Whether or not the sentence contains a deontic cue word (e.g. should, ought, promise).

Experimentation

We use 10-fold cross-validation for experimentation instead of the train/test split used during SemEval due to the small number of social action and intention annotations. We use standard precision, recall, and F1-measure metrics for multi-label classification to judge the results. Table 1 lists the results using various combinations of features.

Features	P	R	F1
V(N=10)	42.9%	**99.2%**	60.0%
V(N=50)	49.7%	98.7%	66.1%
V(N=100)	54.2%	97.6%	69.7%
V(N=200)	57.4%	96.5%	72.0%
1G	79.1%	90.8%	84.5%
1G, 2G	79.5%	91.4%	85.0%
1G, 2G, 3G	77.8%	91.3%	84.1%
1G, V(N=200)	80.0%	90.4%	84.8%
2G, V(N=200)	79.9%	91.3%	85.2%
1G, 2G, V(N=200)	77.3%	90.4%	83.4%
1G, 2G, V(N=200), WD	**80.2%**	91.2%	**85.3%**
1G, 2G, V(N=200), DM	79.9%	91.3%	85.2%
1G, 2G, V(N=200), PC	79.6%	90.3%	84.7%
1G, 2G, V(N=200), WE	79.8%	91.5%	85.2%
1G, 2G, V(N=200), WD, DM	80.1%	91.1%	**85.3%**
1G, 2G, V(N=200), WD, DM, PC	79.7%	90.3%	84.7%
1G, 2G, V(N=200), WD, WE	79.9%	91.4%	**85.3%**
All (V(N=200))	78.0%	90.6%	83.8%

Table 1: **P**recision, **R**ecall, and **F1**-measure results for 10-fold cross-validation using different combinations of features.

As is seen in Table 1, the best overall F1 score is 85.3% achieved by the combination of unigrams, bigrams, learned vocabulary, and WordNet domains. Unigrams and the learned vocabulary are the most discriminative features for this dataset. However, we believe as the number and type of annotations are expanded that the other features will become more discriminative.

Component	P	R	F1
Belief (Product)	79.0%	96.2%	86.8%
Belief (Experience)	75.0%	86.6%	80.4%
Social Action	86.4%	63.9%	73.5%
Intention	50.0%	39.1%	43.9%

Table 2: **P**recision, **R**ecall, and **F1**-measure results for 10-fold cross-validation using 1G, 2G, V(N=200), and WD

The results in Table 1 present the overall performance, but with a highly unbalanced dataset as is the case of intentions and social actions it is of use to look at individual component results. Table 2 shows the precision, recall, and F1-measure by component. The worst of the four components is, as would be expected, intentions, which had the fewest annotations (23). However, even at 43.9% F1 the classification can still be useful as a part of higher level inference.

When examining the errors made by the system it appears that some are due to insufficient context. For example the review “I complain again ...” was mislabeled as an intention, which is not clear without more context. Mislabeling of social actions often was due to the recommendation/suggestion not being directly related to a product. For example in the following review, “And if you have a reservation you’ll wait for max 5 minutes - so have a drink at the bar.” The suggestion to have a drink at the bar is tangential and not the main thrust of the statement.

Examples of errors for belief (product) include: “Obviously run by folks who know a pie.” and “I really liked this place.” In the case of the first example, pie is a product (food) and one can infer from the sentence that the reviewer thinks the pie at the restaurant was good. In the case of the second example, the system picks up “place” as a product and the fact that is liked. These errors indicate that there are differences between the definition of an aspect category and that of beliefs. Errors for belief (experience)include the following review: “The food is uniformly exceptional, with a very capable kitchen which will proudly whip up whatever you feel like eating, whether it’s on the menu or not.” In this example, there are a number of hints at an experience, e.g. “proudly whip up” and “on the menu or not”, that cause the system to mislabel.

Conclusion

In this paper we introduce a four-factor model composed of *Attitudinal*, *Sociocultural*, *Personal*, and *Behavioral* factors to mine consumer insights from social media. The model is grounded in current research in consumer and social psychology and brings in concepts from discourse processing and sentiment analysis. As components of the model are realized, a marketer’s ability to infer and predict consumer responses to their products and brands will increase.

We demonstrate the feasibility of identifying components of the model using a machine learning methodology. We focus on identifying whether or not a sentence contains a linguistic manifestation of: *Beliefs* toward products and experiences, *Social Actions* in the form of recommendations and suggestions, and *Intentions* in the form of promises, positive or negative, made toward/about a product or experience. The results show that these components can be discovered with an F1-measure of 85.3%.

There is much work to be done in the future as this paper only presents our initial results. In particular, our immediate next step is to enhance the annotated corpus to include sentential context and a diversity of product types. We believe to more accurately capture certain categories it is necessary to more fully model the discourse, including factors present before and after the current sentence. Moreover, to determine certain categories of social and personal factors a model of the consumer must be built where a collection of messages is used to classify demographics and their social networks are used to aid in discovering their social role and status.

References

Ajzen, I. 2005. *Attitudes, personality, and behavior*. McGraw-Hill International.

Bailey, A. A. 2005. Consumer awareness and use of product review websites. *Journal of Interactive Advertising* 6(1).

Bracewell, D. B., and Tomlinson, M. T. 2013. In you we follow: determining the group leader in dialogue. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer Berlin Heidelberg. 240–248.

Bracewell, D. B.; Minato, J.; Ren, F.; and Kuroiwa, S. 2006. Determining the emotion of news articles. In *Computational Intelligence*. Springer Berlin Heidelberg. 918–923.

Bracewell, D. B.; Tomlinson, M.; Shi, Y.; Bensley, J.; and Draper, M. 2011. Who’s playing well with others: Determining collegiality in text. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on*, 21–26. IEEE.

Bracewell, D. B.; Hinote, D.; and Monahan, S. 2014. The author perspective model for classifying deontic modality in events. In *The Twenty-Seventh International Flairs Conference*.

Bracewell, D. B. 2008. Semi-automatic creation of an emotion dictionary using wordnet and its evaluation. In *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, 1385–1389. IEEE.

Bracewell, D. B. 2010. Semi-automatic wordnet based emotion dictionary construction. In *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, 629–634. IEEE.

Bracewell, D. B. 2014. The needs of metaphor. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer International Publishing. 229–236.

Bunt, H. 2011. The semantics of dialogue acts. In *Proceedings of the Ninth International Conference on Computational Semantics*, IWCS ’11, 1–13. Stroudsburg, PA, USA: Association for Computational Linguistics.

Cambria, E., and Hussain, A. 2012. *Sentic computing*. Springer.

Cambria, E.; Olsher, D.; and Rajagopal, D. 2014. Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In *Twenty-eighth AAAI conference on artificial intelligence*.

Chartrand, T. L., and Bargh, J. A. 1999. The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology* 76(6):893.

Dini, L., and Mazzini, G. 2002. Opinion classification through information extraction. In *Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields (Data Mining)*, 299–310.

Esuli, A., and Sebastiani, F. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of Language Resources and Evaluation (LREC)*.

Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; and Lin, C.-J. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research* 9:1871–1874.

Granot, E.; Alejandro, T. B.; and La Toya, M. R. 2014. A socio-marketing analysis of the concept of cute and its consumer culture implications. *Journal of Consumer Culture* 14(1):66–87.

Hanssens, D. M.; Pauwels, K. H.; Srinivasan, S.; Vanhuele, M.; and Yildirim, G. 2014. Consumer attitude metrics for guiding marketing mix decisions. *Marketing Science*.

Kassarjian, H. H. 1971. Personality and consumer behavior: A review. *Journal of Marketing Research* 409–418.

Loken, B., and Ward, J. 1990. Alternative approaches to understanding the determinants of typicality. *Journal of Consumer Research* 111–126.

Loken, B.; Barsalou, L. W.; and Joiner, C. 2008. Categorization theory and research in consumer psychology. *Handbook of consumer psychology* 133–65.

Maas, A. L.; Daly, R. E.; Pham, P. T.; Huang, D.; Ng, A. Y.; and Potts, C. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 142–150. Association for Computational Linguistics.

Maslow, A. H. 1943. A theory of human motivation. *Psychological review* 50(4):370.

McCrae, R. R., and Costa, P. T. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology* 52(1):81.

Minato, J.; Bracewell, D. B.; Ren, F.; and Kuroiwa, S. 2006. Statistical analysis of a japanese emotion corpus for natural language processing. In *Computational Intelligence*. Springer Berlin Heidelberg. 924–929.

Minato, J.; Bracewell, D. B.; Ren, F.; and Kuroiwa, S. 2008. Japanese emotion corpus analysis and its use for automatic emotion word identification. *Engineering Letters* 16(1):172–177.

Myszkowski, N., and Storme, M. 2012. How personality traits predict design-driven consumer choices. *Europe's Journal of Psychology* 8(4):641–650.

Pang, B., and Lee, L. 2004. A sentimental education: Sentiment analysis using subjectivity. In *Proceedings of ACL*, 271–278.

Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 79–86.

Parrott, W. G. 2001. *Emotions in social psychology: Essential readings*. Psychology Press.

Pavlopoulos, J., and Androutsopoulos, I. 2014a. Multigranular aspect aggregation in aspect-based sentiment analysis. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, 78–87. Association for Computational Linguistics.

Pavlopoulos, J., and Androutsopoulos, I. 2014b. Proceedings of the 5th workshop on language analysis for social media (lasm). 44–52. Association for Computational Linguistics.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Proceedings of EMNLP*.

Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; and Manandhar, S. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 27–35. Dublin, Ireland: Association for Computational Linguistics and Dublin City University.

Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Dziurzynski, L.; Ramones, S. M.; Agrawal, M.; Shah, A.; Kosinski, M.; Stillwell, D.; Seligman, M. E.; et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8(9):e73791.

Smith, A. N.; Fischer, E.; and Yongjian, C. 2012. How does brand-related user-generated content differ across youtube, facebook, and twitter? *Journal of Interactive Marketing* 26(2):102–113.

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D.; Ng, A.; and Potts, C. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1631–1642. Seattle, Washington, USA: Association for Computational Linguistics.

Stone, P. J.; Bales, R. F.; Namenwirth, J. Z.; and Ogilvie, D. M. 1962. The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science* 7(4):484–498.

Tomlinson, M. T.; Bracewell, D. B.; Draper, M.; Almissour, Z.; Shi, Y.; and Bensley, J. 2012. Pursing power in arabic on-line discussion forums. In *LREC*, 1359–1364.

Triandis, H. C. 1989. The self and social behavior in differing cultural contexts. *Psychological review* 96(3):506.

Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the Association for Computational Linguistics (ACL)*, 417–424.

Uribe, D. 2010. Domain adaptation in sentiment classification. In *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, 857–860. IEEE.

Valitutti, R. 2004. Wordnet-affect: an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*, 1083–1086.

Wiebe, J.; Wilson, T.; and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 1(2):0.

Yan, J.; Bracewell, D. B.; Ren, F.; and Kuroiwa, S. 2006. A semantic analyzer for aiding emotion recognition in chinese. In *Computational Intelligence*. Springer Berlin Heidelberg. 893–901.