Predicting Email Opens with Domain-Sensitive Affect Detection

Niyati Chhaya[1], Kokil Jaidka[2], and Rahul Wadbude[3]

[1] Big Data Experience Lab, Adobe Research, Bangalore, India
nchhaya@adobe.com
[2] University of Pennsylvania, USA
jaidka@sas.upenn.edu
[3] Indian Institute of Technology, Kanpur, India
rahulwadbude2@gmail.com

Abstract. The content and style of the text constitutes the semantic context of its words—a property that is important for many downstream tasks in natural language processing. We demonstrate the advantages of incorporating domain information for affect analysis, and subsequently for the prediction of user responses to marketing emails. Emails are a primary form of marketing communication, and email subject lines are the only indicators of whether the receiver will open an email especially in the case of bulk communication. We analyze the performance of affective features in predicting email opens, on a dataset of 60,000 unique promotion emails from 3 different industries. Our results show that the use of domain-specific affect words is strongly correlated with email opens and outperforms words from the standard ANEW lexicon and other state of the art affective lexica. Implications of this findings can be incorporated into writing tools to improve the productivity of marketing campaigns.

Key words: affect analysis, email marketing, linear programming, convex optimization

1 Introduction

The email subject line plays a critical role in determining whether the email will be opened. It is the single point of insight regarding the email content for the recipient. This study presents a language-based model to predict the open rate of outbound marketing emails. Our experiments demonstrate the importance of linguistic features for this task. We also show how using domain-specific lexica, as against a standard affect lexicon (e.g.ANEW [1]) are able to tailor a generic predictive model to better predict the open rate for industry-specific emails. Our experiments highlight the word-usage preferences for different businesses and show how they vary across industries. Insights from this study can be applied by content writers to create improved email experiences as well as to better understand the psycholinguistic preferences of their customers.

The work makes the following contributions:

1. It implements a **framework to mine domain-specific affect lexica from any corpora of long or short texts**.
2. It demonstrates the **strong univariate relationships of the new lexica with open rates**, which outperform generic affective lexica.
3. It identifies the **word preferences that characterize high open rate for different industries**. Words providing insight and signaling cognitive processing lead to more opens for the Finance industry; on the other hand, social words yield more opens for the Movies & Television industry.

1.1 Paper Organization

Section 2 presents an overview of prior work in the space of email understanding and explorations with linguistic features. The method describing the regression model as well as the construction of domain-specific affect lexica is discussed in Section 3. An analysis on the linguistics features followed by the experiments is presented in Section 4. We conclude with a note on further explorations in Section 5.

2 Related Work

Early studies of users actions on emails were conducted in a relative small scope on a small set of monitored users [2]. Recent studies have looked at contact interactions [3], the effect of personalization [4] and the role of text-agnostic features [5] for predicting email opens; however, no study has attempted to predict email opens based on the sentiment in the subject line.

Our approach is based on building a custom domain-specific affect lexicon to model linguistic features for the prediction task. General-purpose affective lexica are used to detect emotions and sentiment at the word level, in various natural language tasks [1, 6–8]. However, standard lexica often fail to capture the domain-specific orientation of the content. Several approaches have adapted general-purpose lexica for research problems in specific domains [9, **?**]. We use an approach similar to studies which have explored the use of syntactic structure – such as parsing rules, linguistic patterns [10, 9, 11], and Latent Semantic Analysis [12],[13],[14] and collocations [15] – to identify domain-specific affect words. This is the first paper to apply the use of domain-specific lexica for **predicting email open rates**, since previous work has mostly focused on opinion mining tasks for product reviews.

3 Method

The first objective of this study was to extract different linguistic and meta–features from the labeled corpora, and build three industry-specific predictive models to predict the open rate for individual emails. The second objective is to understand the impact of affect words in open rate prediction.

3.1 Data Collection

Data access was provided by Edatasource[4], an email inbox monitoring organization which tracks over 25 million emails for $90,000$ distinct businesses per day. The data is categorized into 98 industries. Using their licensed API, we were able to download the following information for up to $20,000$ promotional emails each, sent over a one-year period (April 2015 to March 2016) and over fifty businesses each in Finance, Cosmetics, and Movies & Television:

- **Email information**: The subject line, contents, send date, time and time zone, sending email domain, name of the business, and industry category.
- **Recipient responses**: The number of the promotional emails which were received in tracked inboxes, percentage proportion of the recipients who read the email, and a projection of the total number of recipients for the message.

3.2 Domain-sensitive Affect Detection – BATframe

We adapted an optimization-based approach to build domain-specific lexica for subject lines from three industries. This approach was proposed in [16] and outperformed the state of the art in the SemEval 2007 Affect Corpus, with a precision of over 70% as compared to the best performing system at 47%. (Figure 1). According to this approach the affinity between affect words in the neighborhood
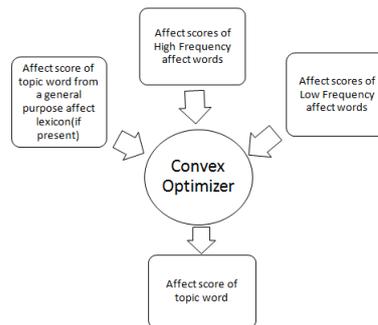

Fig. 1: The BATFrame Framework for Domain-sensitive affect detection [16].

of topic (domain) words is modeled as a optimization function in this approach. First, the subject lines are tokenized using HappierFunTokenizer[5] to produce a total of 0.9 million tokens. Next, the Topic–affect Tuple Extractor word pairs

---

that couple the topic space with standard affect words, using n-grams and dependency rules. Finally, the optimization Framework tunes the domain-specific Pleasure (Valence), Arousal, and Dominance (P,A,D) scores for the topic word on the basis of a set of constraints. The mathematical expression is as follows:

$$\omega_{PAD} = \lambda_1 \sum_{j=1}^{n} I_{w_j}^G ||S_{w_j} - G_{w_j}||_2$$
$$+ \lambda_2 \sum_{j=1}^{n} \sum_{a_k \in HF_j} \alpha_{jk} ||S_{w_j} - G_{a_k}||_2 \qquad (1)$$
$$+ \lambda_3 \sum_{j=1}^{n} \sum_{b_k \in LF_j} \alpha_{jk} ||S_{w_j} - G_{b_k}||_2$$

Now the optimization problem is given by $S_p = \min \omega_{PAD}$, subject to:

$$1 \leq S_{jp} \leq 9 \ ; \ 1 \leq S_{ja} \leq 9 \ ; \ 1 \leq S_{jd} \leq 9 \qquad (2)$$

where λ_1, λ_2 are weighting parameters which should be set to the degree that we trust each source of information, and λ_3 can be set to a small non-zero value such as 0.002. Table 1 illustrates some resulting domain-specific topic words from the three corpora, that were not present in ANEW in any lemma form, which demonstrates how domain-specific lexica capture more affective content as against standard lexica.

Table 1: The top ten highest-weighted domain-specific words in the BATFrame lexica which are not present in ANEW or in Warriner's Lexicon.

Corpus	New Affect words	New N
Finance	app, dividend, anyone, discount, quantitative, +, data, divergence, flight, authentication	198
Cosmetics	men, addition, flirtiest, everyone, off, you, more, 5-star, matte, own	73
Movies & TV	today, prime-time, easy-to-please, nail-art, one-pot, well, loud, while, %, front	100

4 Experiments

The purpose of this evaluation is to test whether domain-specific lexica contributes to predicting email opens. We posit that by producing new features along the three dimensions of Pleasure, Arousal and Dominance, the BAT lexicon will improve on the performance over standard lexica to predict email opens.

We conducted (a) univariate regressions to predict the open rate of emails using (i) meta-features and POS tags and (ii) the three dimensions of affect from ANEW [1], the Warriner's lexicon [6] and the BAT lexica. In order to do so, we generated features representing different kinds of meta-information (subject line length and word count), the percentage use of punctuations and symbols, and part-of-speech tags in each subject line using the TweetNLP tagger, which is trained on social media text [17]. Figure 2 depicts the 1-to-3 grams positively

Table 2: Standardized regression coefficients (βs) between different features of subject lines, and email opens. Subject lines containing the positively correlated features below are significantly more likely to be opened. All correlations are significant at p<.01, two tailed t-test.

Finance		Cosmetics		Movies & Television	
Feature Set	R^{**}	**Feature Set**	R^{**}	**Feature Set**	R^{**}
Meta-features & Parts of Speech					
Word Count	-.11	Brackets	.17	Possessive Pronoun	.20
Verb	-.10	Proper Noun	.09	Verb, third person singular	.16
Currency	-.08	Present Tense	-.08	Numbers	-.12
All Punctuations	-.11	Quantity	-.05	:	-.14
ANEW					
Arousal	.06	Arousal	-.06	Arousal	.05
Valence	-.05	Valence	-.10	Valence	.04
Dominance	-.10	Dominance	-.09	Dominance	-.06
Warriner's Lexicon (Extended ANEW)					
Arousal	.10	Arousal	.05	Arousal	.03
Valence	.09	Valence	-	Valence	.07
Dominance	.12	Dominance	.04	Dominance	.08
BATFrame					
Arousal	.11	Arousal	-.14	Arousal	.10
Valence	.10	Valence	-.14	Valence	.09
Dominance	.09	Dominance	-.15	Dominance	.11

and negatively correlated with open rate. All the correlations were Bonferroni-corrected, and are significant at $p < 0.01$. The size of the word reflects its correlation with open rate, while the shade reflects its frequency in the dataset. In Cosmetics, words such as 'registration' and 'member' and phrases such as 'welcome to' led to more opens; on the other hand, phrases mentioning 'notifications' and discounts ('%') were negatively correlated with opens. In Movies & Television, subject lines with 'you' and phrases such as 'is now' were more likely to be opened, and subject lines mentioning news coverage ('breaking news') were less likely to be opened.

Table 2 illustrates the text features among ANEW features, Warriner's (extended ANEW) features, BAT lexica features, meta-features and parts of speech,

which were most highly correlated with Open Rates for the three industries. The effect sizes for individual features ranged from -0.18 to 0.23 across the industries.

The table enable us to compare the characteristics of subject lines across various industries. We observe that for different industries, different words, phrases, and topics are more likely to yield higher open rates:

- **Meta-features and POS** Short and crisp subject lines devoid of punctuation are evidently preferred in the Finance industry, and are more likely to be opened; on the other hand, proper nouns perform well in Cosmetics, and possessive pronouns do well in Movies & Television.
- **ANEW, Warriner's Lexicon and BATFrame:** BATFrame outperforms ANEW and also Warriner's Lexicon in all three corpora. ANEW has the poorest performance with the weakest coeffiecients. Warriner's lexicon is comparable to BATFrame in the Finance corpora.
- **BATFrame features:** These features highlight that while Valence and Arousal features have similar importance across industries, the importance of Dominance terms is varies.

We also trained three multivariate linear regression models to predict email open rates on a held out test set. The feature set comprised standard tf-idf features [18] calculated from the n-gram distributions complemented with one of the three affective lexica. Because of the large effect size and number of features available from words, there was no significant difference in the effect sizes from either the ANEW, the Warriner's, or the BAT lexica, and all three models yielded an average Mean Absolute Error of 0.07 across the three corpora.

4.1 Qualitative Evaluation

Figure 3 shows that the domain–specific BAT lexica offer more consistent coverage in corpora from all three industries, after excluding stop words. Coverage can be interpreted as how representative any lexicon is of the overall vocabulary of the test set. The BAT lexica achieves around 85% coverage of the vocabulary corpus as against the 10% coverage by ANEW. BAT–Finance lexica outperforms Warriner's Lexicon (approx 13k tokens) as well which is 14 times the size of the BAT–Finance (975 tokens). Note that the BAT lexica (BAT–cosmetic:419, BAT–TV:848) are significantly smaller in size as compared the standard lexica (ANEW:1034).

The results highlight the importance of a domain-dependent lexicon for the accurate affect analysis of short texts. This supports our argument that a domain-specific lexicon would be more representative of text than a general-purpose lexicon.

5 Conclusion

Our study demonstrates that domain-specific affect lexica can improve sentiment and affect detection in different applications and for several predictive problems.

(a)   (b)

(c)   (d)

Fig. 2: The word cloud shows the ngrams that are significantly correlated with the open rate in the Cosmetics (a,b) and Movies & Television (c,d) industry. The word clouds in blue and red represent the positively and negatively correlated words respectively. The size of the ngram is proportional to its correlation with the open rate. The shade of the color signifies the frequency of occurrence; darker shades imply higher frequency as compared to lighter shades.
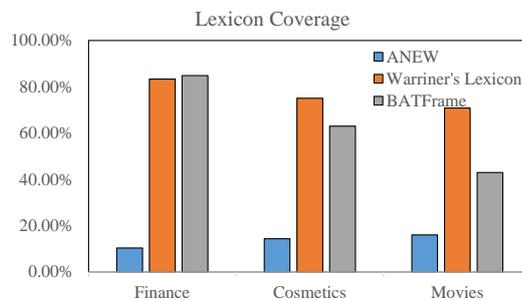
Fig. 3: Coverage Statistics on the Finance, Cosmetics and Movies & Television corpora. This figure shows that domain-specific lexica generated from BAT a significant amount of total vocabulary of a corpus, which is comparable to the Warriner's lexicon, and four times as much as ANEW.

We establish the importance of affect-based linguistic features for email analytics on a real-world dataset and further contrast the language preferences across data three industries: Finance, Cosmetics, and Movies & TV. We are currently extending this work towards generating suggestions for improved email opens. Our results also suggest that such approaches could be useful for word sense disambiguation.

References

1. Bradley, M.M., Lang, P.J.: Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology, University of Florida (1999)
2. Lim, K.H., Lim, E.P., Jiang, B., Achananuparp, P.: Using online controlled experiments to examine authority effects on user behavior in email campaigns. In: Proceedings of the 27th ACM Conference on Hypertext and Social Media, ACM (2016) 255–260
3. Di Castro, D., Karnin, Z., Lewin-Eytan, L., Maarek, Y.: You've got mail, and here is what you could do with it!: Analyzing and predicting actions on email messages. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, ACM (2016) 307–316
4. Sahni, N.S., Wheeler, S.C., Chintagunta, P.K.: Personalization in email marketing: The role of non-informative advertising content. (2016)
5. Luo, X., Nadanasabapathy, R., Zincir-Heywood, A.N., Gallant, K., Peduruge, J.: Predictive analysis on tracking emails for targeted marketing. In: International Conference on Discovery Science, Springer (2015) 116–130
6. Warriner, A.B., Kuperman, V., Brysbaert, M.: Norms of valence, arousal, and dominance for 13,915 english lemmas. Behavior Research Methods **45** (2013) 1191–1207
7. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational Linguistics **37** (2011) 267–307
8. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of LREC. Volume 6., Citeseer (2006) 417–422
9. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. Computational linguistics **37** (2011) 9–27
10. Blitzer, J., Dredze, M., Pereira, F., et al.: Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: ACL. Volume 7. (2007) 440–447
11. Chikersal, P., Poria, S., Cambria, E., Gelbukh, A., Siong, C.E.: Modelling public sentiment in twitter: using linguistic patterns to enhance supervised learning. In: International Conference on Intelligent Text Processing and Computational Linguistics, Springer (2015) 49–65
12. Turney, P.D., Littman, M.L.: Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS) **21** (2003) 315–346
13. Bestgen, Y.: Building affective lexicons from specific corpora for automatic sentiment analysis. In: LREC. (2008)
14. Bestgen, Y., Vincze, N.: Checking and bootstrapping lexical norms by means of word similarity indexes. Behavior Research Methods **44** (2012) 998–1006

15. Wiebe, J., Wilson, T., Bell, M.: Identifying collocations for recognizing opinions. In: Proceedings of the ACL-01 Workshop on Collocation: Computational Extraction, Analysis, and Exploitation. (2001) 24–31
16. Jaidka, K., Chhaya, N., Wadbude, R., Kedia, S., Nallagatla, M.: (Batframe: An unsupervised approach for domain-sensitive affect detection)
17. Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., Smith, N.A.: Improved part-of-speech tagging for online conversational text with word clusters, Association for Computational Linguistics (2013)
18. Ramos, J., et al.: Using tf-idf to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning. Volume 242. (2003) 133–142