Predicting Email and Article Clickthroughs with Domain-adaptive Language Models

Kokil Jaidka
University of Pennsylvania
Philadelphia, Pennsylvania
jaidka@sas.upenn.edu

Tanya Goyal
University of Texas
Austin, Texas
tanyagoyal.93@gmail.com

Niyati Chhaya
Adobe Research
Bangalore, India
nchhaya@adobe.com

ABSTRACT

Marketing practices have adopted the use of computational approaches in order to optimize the performance of their promotional emails and site advertisements. In the case of promotional emails, subject lines have been found to offer a reliable signal of whether the recipient will open an email or not. Clickbait headlines are also known to drive reader engagement. In this study, we explore the differences in recipients' preferences for subject lines of marketing emails from different industries, in terms of their clickthrough rates on marketing emails sent by different businesses in Finance, Cosmetics and Television industries. Different stylistic strategies of subject lines characterize high clickthroughs in different commercial verticals. For instance, words providing insight and signaling cognitive processing lead to more clickthroughs for the Finance industry; on the other hand, social words yield more clickthroughs for the Movies and Television industry. Domain adaptation can further improve predictive performance for unseen businesses by an average of 16.52% over generic industry-specific predictive models. We conclude with a discussion on the implications of our findings and suggestions for future work.

KEYWORDS

email marketing, subject lines, linguistic analysis, copy-writing strategies, machine learning, domain adaptation, open rate prediction, clickthroughs, online ads, advertisements

ACM Reference Format:
Kokil Jaidka, Tanya Goyal, and Niyati Chhaya. 2018. Predicting Email and Article Clickthroughs with Domain-adaptive Language Models. In *Proceedings of 10th ACM Conference on Web Science (WebSci'18).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3201064.3201071

1 INTRODUCTION

The language used in a subject line plays an important role as a determinant of relevance. For example, it can lead to recipients choosing to open an email sooner rather than later, or to delete the email without opening it at all; it could lead to them marking the email as spam or even unsubscribing from the mailing list. Email servers also use language models to predict whether or not an email is spam. In the case of news websites, copywriters tailor sponsored articles to get the reader's attention with interesting or provocative headlines that are relevant to their personal interests.

We propose to use click information to predict the performance of future promotional emails. The findings can be used to formulate copy–writing cues for different industries, and for captioning news articles (see, for instance, the subject line performance predictors developed by Persado[1] and Adobe Campaign[2]). Modeling the language of email subject lines is also increasingly relevant for tasks such as inbox management, email stacking, and information retrieval, in order to understand the relevance of an email to its recipient. Intelligent inbox management techniques, such as the stacking of emails under tabs [1] use features based on the subject lines, content and sender information.

Since language-based models do not generalize well to out-of-domain samples, we implement domain–adaptive approaches to tailor a generic predictive model towards making better predictions for different businesses, where the distribution of clicks may differ widely from the industry norm.

In summary, our contributions include: 1) a new NLP task for clickthrough prediction on email subject lines from a variety of businesses and industries (Table 1); 2) an exploration of the effectiveness of language modeling for predicting email clickthroughs and identifying clickbait articles. Our models show promise in predicting email open rates (hereafter referred to as clickthrough rates or *CTR*), with an average mean absolute error (MAE) of 4.5% (Table 2). On the clickbait detection task, the proposed approach provides a performance gain of 7.7% in recall and a 4.8% gain in F1-score over the state of the art (Table 3); and 3) the implementation of domain adaptation approaches to make these language models generalizable to out-of-domain samples, leading to a 16.5% improvement over prior results (Table 6).

2 RELATED WORK

The literature exploring the relationship between the language of emails and recipient behavior has drawn on mixed methods to offer linguistic and predictive insights into the best strategies for garnering email responses. The paper by Miller and Charles [21] provides a qualitative analysis of 150 personal emails from the Enron email dataset[3] and 150 spam emails from the Spamdex dataset[4], to propose 40 rules for improving the impact of marketing emails. However, these qualitative analyses are necessarily small in scope,

WebSci'18, May 27–30, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5563-6/18/05.
https://doi.org/10.1145/3201064.3201071

[1]https://persado.com/wp-content/uploads/2015/06/PersadoGoDatasheet.pdf
[2]https://blogs.adobe.com/digitalmarketing/email/amplifying-email-insights-predictive-subject-lines/
[3]http://cs.cmu.edu/ enron/
[4]http://www.spamdex.co.uk

and do not compare between the best strategies for different types of businesses.

Studies have predicted email clickthroughs [2, 18, 28] using simplistic keywords, syntactic features, and time–based features [9, 19]. The study by Shish et. al has used email subject lines among other features for the detection of email spam [31], while Ferriera et. al have used subject lines to identify phishing emails [10]. Most studies profiling user behavior on emails have been conducted in a relative small scope on a small set of monitored users [14, 18, 26]. The study by Di et al. [9] analyzed a corpus of conversations between known contacts. The problem takes on a different set of challenges when the focus is on targeting marketing emails, sent by a corporate entity. Studies have measured user preferences in terms of ad clickthroughs [12] and consumers' motivations in passing along emails in viral marketing campaigns [23], as well as the cognitive predictions which may lead them to forward or click the links in marketing emails [15, 34]. The authors observed recipients evaluate marketing emails on a number of cognitive factors such as the benefit goals related to the message, trusting beliefs in the message sender, involvement with the message and the cost of future efforts required to follow through with the message.

We have compared our method against the work of Sahni et. al [28]. The authors test the contribution of email subject line personalization in boosting the rate of email clickthroughs. They conducted randomized control trials on 68000 recipients and concluded that adding the name of the message recipient to the email's subject-line increases the probability of the recipient opening it by 20%. Although we have used personalization as one of our features, unexpectedly we did not observe it to play a role in determining email clickthroughs, once linguistic features were also included.

Some studies on ad clickthroughs have modeled the problem using website content features [27], advertising-based features such as hotness, promotion, events and sentiment [5], similarity to surrounding ads [8], user demographics [6, 13] and attractiveness features or attractive words [16]. However, the case of predicting email clickthroughs is quite different because the relevant cues are limited to information about the sender and the email subject line. There are a few online tools and plug-ins such as 'Downworthy' which detects clickbait headlines by following a rule–based approach [11]. The qualitative analysis by Blom and Hansen[3] used a dictionary of forward-references at the discourse level ("This news will blow your mind") and at the phrase level ("This name is hilarious.") to identify that clickbait articles occur mostly in commercial, ad-funded, and tabloid news websites. However, they did not propose an approach for clickbait detection. Evidently, these approaches are not scalable because they operate on a manually curated list of phrases or links. Potthast et al. [24] developing a clickbait classifier for Twitter based on a small annotated dataset.

We compare our approach against the clickbait classifier developed by Chakraborty et. al [4] on a corpus of 15000 article headlines scraped from WikiNews and various tabloid news websites. Their best-performing classifier used sentence structure, topical similarity, lexical patterns and n-grams in a support vector machine (SVM) set-up to achieve an accuracy of 0.93.

3 METHOD

There are three parts to this study:

- **Predictive modeling**: We compare the performance of these features in predicting email opens and clickbait articles. We demonstrate that the best predictive models are data-driven and significantly outperform the state-of-the-art.
- **Language insights**: We highlight the successful copy–writing strategies in different contexts by conducting univariate regressions of these features with email clickthroughs.
- **Domain adaptation**: We demonstrate how domain adaptation improves out-of-sample predictions on unseen domains and businesses and further for diagnosing the differences through insights.

4 DATA AND FEATURE EXTRACTION

The Email dataset: The email dataset was collected from Edatasource[5], an email inbox monitoring organization which tracks over 25 million emails for 90000 distinct businesses per day. The data is categorized into 98 industries. Using their licensed API, we were able to download the email meta-data and aggregate recipient response information for up to 20000 promotional emails each, sent over a one–year period (April 2015 to March 2016) for three categories of businesses: Finance, Cosmetics, and Movies & Television. Statistics about the data sets are provided in Table 1. The distribution of clickthrough rates for Cosmetics (μ = 0.13) and Movies & Television (μ = 0.12) are fairly Gaussian, and much like each other, while Finance (μ = 0.26) has a few outliers and a mean twice that of that the former industries. The final plot demonstrates that there is a wide variance in the clickthrough rates across businesses within the same industry. This highlights the need for intra-industry domain adaptation in predictive models for email clickthrough rates, especially when the prediction is for unseen businesses.

Table 1: Dataset description: email subject lines

	Finance	Cosmetics	Movies & TV
Emails	18941	3394	3165
Businesses	47	57	69
Users	24m	5.7m	3.5m
Training Set	15168	2731	2555
Test Set	3773	663	610
μ Clickthrough Rate (%)	26.0	13.0	12.0
σ S.D.	13.0	9.0	9.0
γ Skewness	0.29	2.1	2.2

The Clickbait dataset: The Clickbait dataset comprises a total of 15000 randomly sampled articles, half of which are labeled positive (clickbait present) or negative (non–clickbait), and were collected from Buzzfeed, Scroll and several news websites. Detailed about data collection and sampling are provided in the original paper by Chakraborty et. al [4].

[5] see http://www.edatasource.com. Edatasource monitors the email inboxes of millions of email users, after obtaining their consent, and saves email contents and user responses in a de–identified form for the purposes of marketing research.

4.1 Pre-processing the Email dataset

We discarded those emails which were received by less than 100 users. This is because the clickthrough rates (as provided by eDatasource) are extrapolated for a larger population, based on the actual responses of a smaller sample of users. Our expectation is that if the actual no. of observations is more, then there is more confidence that the observed clickthrough rate is closer to the actual clickthrough rate. The law of large numbers states that the observed response rate for m recipients converges to the actual response rate as the value of m increases. We conducted a preliminary weighted regression analysis, by assigning a weight to each sample subject line, which dictates its contribution to the regression loss function. Let S denote the n subject lines in the data set. For each sample s_i, the weight is defined as,

$$w_i = log(m_i/m_o) \tag{1}$$

where m_i is the number of recipients of email campaign with subject line s_i, whose responses were recorded.
Thus, in the loss function, we give more weight to subject line samples whose response rates have been calculated based on the reaction of larger number of recipients.
We identified that the best performance was obtained by aggressively down–weighting those subject lines which had an actual recorded volume of under 100 emails, which always had a clickthrough rate of 0%. They also appeared to be mis–categorized as industry–specific emails. After filtering out emails with less than 100 users, we were left with 18941, 3364, and 3165 subject lines respectively for Finance, Cosmetics, and Movies & Television.

4.2 Feature Extraction

This section provides an overview of the feature extraction process used to construct sets of meta-features (lengths and other counts), syntactic features (Parts-of-speech) and linguistic features (lexicon-based and data-driven) from the email and the clickbait datasets.

N-grams(3000 features): We use the bag-of-words representation to reduce the line of text in either dataset (either subject line or article headline) to a normalized frequency distribution over a vocabulary. Due to our sample size, we reduced the dimensionality of our n-gram feature space by retaining only the most frequent 1000 1−, 2−, and 3−grams each, used in at least 10% of the emails.

$$freq_{rel}(line, ng) = \frac{freq(line, ng)}{\sum_{ng' \epsilon ngs} freq_{abs}(line, ng')} \tag{2}$$

(GI) General Inquirer categorization (184 features): The General Inquirer lexicon [32] comprises over 15000 words arranged in 184 thematic categories[6], such as Social, Motion, Food, Power and Money. Each line in either dataset was thus represented in terms of the percentage proportions of the 184 lexicon categories within the General Inquirer.
Word2Vec embeddings(100 features): Data-driven topics are expected to be more representative of the short text in our datasets as compared to the General Inquirer dictionary. We represent subject lines through topic clusters of their neural embeddings trained on the skip-gram Word2Vec Twitter corpus [20], factorized using a word-context PMI matrix [17]. We use the Gensim implementation provided by [25] to generate 100 'topics' of closely related words.
Topic modeling (2000 features): A total 2000 social–media specific topics provided as an open-sourced resource by Schwartz et al. [29] are used for topic modeling. These topics were created from approximately 18 million Facebook updates, by using the Mallet package in Python to implement Latent Dirichlet Allocation (LDA) with the alpha set to 0.30 for the LDA computation in order to favor fewer topics per document. Each line in either dataset is transformed into 2000 features, derived from **a.** its probability of mentioning words, $p(word|line)$ and **b.** the probability of the words being in the given topics, ($p(topic|word)$). The distribution of topics for each line is thus calculated as:

$$p(topic|line) = \sum_{word' \epsilon topic} p(topic|word) \times p(word, line) \tag{3}$$

where $p(topic|line)$ is the normalized word use in a line and $p(topic|word)$, the probability of the topic given the word, is provided by LDA. Furthermore, we use the joint probability, $p(word, topic)$, in order to determine a word's prevalence in a topic.
(POS) Part of Speech tagging (36 features): We extracted part of speech tags for each line using the TweetNLP tagger, which is trained on social media text [22].
Meta-feature extraction (13 features): We also mined the raw counts for character length and word count, number of punctuation marks, number of symbols and the presence or absence of personalization elements in the subject line (for example, the mention of the recipients' names).

5 PREDICTIVE PERFORMANCE

5.1 Clickthrough rate prediction

We used DLATK's implementation of Python's scikitlearn package [30] to model email clickthroughs in terms of the language of the subject lines. We conducted a five–fold cross–validated weighted linear regression on the dataset with Ridge, Elastic–net, and Lasso regularization. The performance is measured on the held–out sample by using the Mean Absolute Error (MAE) and the goodness–of–fit (R^2). To avoid overfitting, we use randomized principal component analysis (PCA) after filtering out any features which were not significantly correlated with the outcome in univariate regressions. We also set a feature occurrence threshold of 10% to discard sparse features. We have reported the results from Elastic Net regularization in the following section.
State-of-the-art model: We use the approach proposed by Balakrishnan and Parekh [2] and the average clickthrough rate over the entire dataset as our baseline models.

All our models improve upon the baselines – the mean clickthrough rate for the industry as well as the state-of-the-art model. Parts-of-speech were not useful for predicting email clickthroughs, which suggests that general structure of a subject line is somewhat similar across all emails. The General Inquirer model had an average goodness-of-fit of 0.31. A remarkable improvement is observed when data-driven features i.e. the topics and the top 1000 n-grams are used with an average goodness-of-fit of 0.45 and 0.58 respectively. Finally, we combined individual feature sets and obtained the best performances using models built on General Inquirer + n-grams, and topics + n-grams, with average goodness-of-fit of 0.52

[6] http://www.wjh.harvard.edu/ inquirer/homecat.htm

Table 2: RMSE, MAE and R^2 results from Elastic Net regression on the held-out sample on different meta-, syntactic and linguistic feature sets. The best-performing predictor is built on topics + n-grams, with an MAE of 0.04.

	Feature Set	Baseline (Mean)	Balakrishnan & Parekh 2014	Meta-features	POS	GI	Word2Vec	Topics	N-grams	GI + N-grams	Topics + N-grams
Finance	RMSE	13.0	13.1	10.3	12.2	11.4	15.1	10.4	7.3	9.2	**7.1**
	MAE	11.1	11.0	10.3	10.5	9.1	12.1	7.0	5.3	6.0	**5.0**
	R^2	–	.09	.09	.15	.33	0.15	.45	.67	.56	**.69**
Cosmetics	RMSE	9.3	9.0	8.1	9.2	8.0	12.4	6.3	6.0	7.1	**6.0**
	MAE	6.4	7.2	6.1	6.0	6.3	9.1	5.5	4.0	5.3	**4.0**
	R^2	–	.06	.06	.11	.30	0.19	.50	.51	.47	**.56**
Movies & Television	RMSE	8.1	9.2	8.4	9.3	7.0	9.4	5.0	5.1	6.2	**5.0**
	MAE	6.0	7.1	6.2	6.4	5.3	7.1	4.3	4.4	4.1	**4.0**
	R^2	–	.06	.06	.14	.32	.17	.58	.63	.57	**.64**

and 0.64 respectively. Our Word2Vec topics did not yield promising results, perhaps because of the mismatch between training and test corpora.

5.2 Clickbait Detection

We compare the performance of an SVM classifier trained on topics + n-grams (our best performing feature set from Table 2) on the clickbait detection task, against the results reported by Chakraborty et. al [4] in Table 3. We report a 7.7% gain in recall, and a 4.8% gain in F1-score over their best performing SVM classifier. This exercise helps establish the validity of our approach across another similar task and a different, standardized dataset.

Table 3: Results on the Clickbait classification task, against the best–performing SVM classifier by Chakraborty et. al.

	Accuracy	Precision	Recall	F1 Score
[4]	0.93	0.95	0.90	0.93
N-grams	0.89	0.85	0.92	0.89
Topics + N-grams	**0.97**	**0.98**	**0.97**	**0.98**

6 LANGUAGE INSIGHTS

We conduct a regression analysis between all features sets, and the percentage user responses per subject line. We use least squares linear regression over standardized independent variables (linguistic and meta–features extracted from subject lines), which produces a standardized coefficient equivalent to Pearson's R correlation coefficients. All results are significant after Benjamini-Hochberg corrections for multiple comparisons.

Table 4 illustrates the textual features among meta-features and parts of speech, General Inquirer categories, n-grams and topics. The effect sizes for individual features ranged from −0.15 to 0.29 across the industries.

6.1.1. N-Grams: Figure 1 depicts the 1-to-3 grams with a) positive and b) negative Pearson correlation with clickthrough rate, represented as a word cloud. All the correlations were Bonferroni-corrected, and are significant at $p < 0.01$. The size of the word reflects a higher Pearson correlation with clickthrough rate, while a darker shade reflects a higher frequency in the dataset. In Cosmetics, words such as 'please' and phrases such as 'surprises!' led to more clickthroughs; on the other hand, phrases mentioning discounts as '% off' were negatively correlated with clickthroughs. In Movies & Television, phrases such as 'might like' in the subject line were more likely to be clicked open, and subject lines mentioning news coverage or livestreaming footage ('is live!') were less likely to be clicked open. Likewise, for Finance we observed that subject lines with words such as 'statement' and 'card' were more likely while 'reward' was less likely to be clicked open.

6.1.2. General Inquirer: Power Gain (words about increasing power, or being powerful such as *emerge, ascend, appoint*) is positively correlated with clickthrough rates in Finance; on the other hand it is negatively correlated with clickthrough rates in Movies & Television. Food (*bacon, breakfast, cereal*) is correlated with higher clickthrough rates in Cosmetics; on the other hand, it is negatively correlated with clickthrough rates in the Movies & Television. Both Cosmetics and Movies & Television demonstrate a negative correlation of clickthrough rates with words depicting body parts (*arms, belly*).

6.1.3. Topics: Data-driven topics are helpful to contextualize the results from the n-gram analysis, as they provide an intermediate level of granularity between the two. Subject lines about saving on grocery shopping (*money, earn, pocket, savings*) are less likely to be clicked open for Finance, and subject lines about specific bathing products (*shampoo, make-up, soap*) are less likely to be clicked open in Cosmetics.

6.1.4. Parts of Speech and meta–features: Short and crisp subject lines devoid of punctuation are evidently preferred in the Finance industry, and are more likely to be clicked open; on the other hand, proper nouns perform well in Cosmetics, and possessive pronouns do well in Movies & Television.

7 DOMAIN ADAPTATION

In the previous section, we showed that the topics + n-grams model has the best in-domain prediction. The challenge arises for predictions for a different industry or a new business. Table 5 shows that predictive performance can drop by 85% on predictions on other domains and by up to 28% on unseen businesses. We diagnose these differences based on the very different regression coefficients for the same General Inquirer categories, in Table 4 and Figure 2. We address this problem by implementing an unsupervised and a supervised domain adaptation method and comparing the performance to the pre-domain adaptation step.

Table 4: **Different copy-writing strategies succeed in different industries. Standardized regression coefficients (βs) between different features of subject lines, and email clickthroughs. All correlations are Bonferroni-corrected and significant at p<.000, two tailed t-test.**

Finance		Cosmetics		Movies	
Feature Set	R^{**}	**Feature Set**	R^{**}	**Feature Set**	R^{**}
Topics					
false, rumors, statement etc.	.25	*wife, sweetheart, hubby* etc.	.28	*flick, horror, scary* etc.	.18
card, credit, sim, visa etc.	.21	*smile, compliment, stares* etc.	.08	*pause, button, snooze* etc.	.18
spree, grocery, shopping etc.	-.11	*sell, discount, item* etc.	-.05	*livestream, lifestyle* etc.	-.12
money, pocket, savings etc.	-.11	*shampoo, make-up, soap* etc.	-.05	*news, report, flash* etc.	-.18
General Inquirer					
Power Gain *(emerge, ascend etc.)*	.06	Food *(bacon, breakfast)*	.12	Commn Object *(content, check)*	.18
Social Relations *(accept, act)*	.05	Causal *(because, chance)*	.10	Communication *(account,address)*	.17
Positive *(create,achieve)*	-.10	Body Parts *(arms, belly)*	-.08	Food *(bacon, breakfast)*	-.09
Understatement *(few, hard)*	-.06	Virtue *(cute, desire)*	-.08	Power Gain *(emerge, ascend etc.)*	-.11
Parts of Speech & Meta-Features					
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

(a)

(b)

(c)

(d)

Figure 1: **The word clouds show the ngrams that are significantly correlated with the clickthrough rate in the Cosmetics (a,b) and Movies (c,d) industry. The word clouds in blue and red represent the positively and negatively correlated words respectively. All the correlations were Bonferroni-corrected, and are significant at $p < 0.01$. The size of the word reflects a higher Pearson correlation with clickthrough rate, while a darker shade reflects a higher frequency in the dataset.**

Table 5: Average out-of-domain prediction performance (RMSE) by the Topics + ngrams model from Table 2 over a five-fold cross-validation. Performance drops by up to 85% on out-of-domain predictions.

Training Set	In-domain RMSE	Out-of-domain RMSE		
I. Prediction on unseen domains				
		Finance	Cosmetics	Movies & TV
Finance	7.0	-	7.0	6.0
Cosmetics	6.0	13.0	-	6.0
Movies & TV	5.0	11.0	7.0	-
II. Prediction on unseen businesses				
Training Set	RMSE on Test Set			
		Business 1	Business 2	Business 3
Finance	7.0	9.0	9.0	10.0
Cosmetics	6.0	9.0	5.0	7.0
Movies & TV	5.0	5.0	5.0	6.0

7.1 Prediction on new businesses

Predictive performance also suffers when a model is used for prediction for unseen businesses within the same industry. Any model trained on an industry-wide dataset normalizes the effect of individual features across individual businesses. In Figure 2, we show the Hinton diagram of β values of a few significantly correlated General Inquirer features for specific businesses within the Finance industry as well as the industry overall. The area occupied by a square is proportional to the magnitude of the coefficient, and the color (green/red) indicates its sign (positive/negative). The variation in the values demonstrates the difference in the influence of features between businesses. For instance, the *Overstate* category has a large positive influence on the clickthrough rate for Business B, however, employing an industry-wide model would predict a decrease in clickthrough rate due to the normalization enforced in generic models. While a business specific model will nearly always perform better than an industry wide model, it is often not possible to train a separate model for every business due to unavailability of sufficient business specific data. Domain adaptive approaches would allow a generic industry-level predictive model to make better predictions for a business.

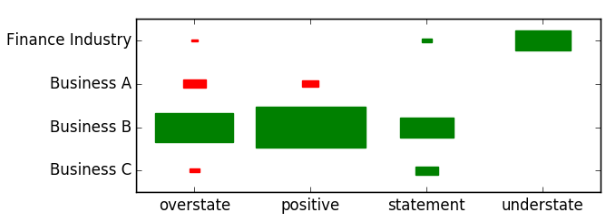

Figure 2: Different businesses in the same industry also pursue different copy-writing strategies. This Hinton diagram illustrates the β values for three different businesses within Finance for a few of the significant General Inquirer categories. The area occupied by a square is proportional to a value's magnitude, and the color (green/red) indicates its sign (positive/negative).

7.2 Domain–adaptation Methods

In the rest of this paper, we test two standard domain adaptation techniques for clickthrough rate prediction for unseen businesses within a particular industry domain. The first approach, Correlation Alignment (CORAL) [33], is an unsupervised approach which was developed for visual domain adaptation. The second approach, Easy Adapt [7], is a supervised approach that uses labeled data from both the source and the target to retrain a model.

7.2.1 CORAL. CORAL minimizes the distance (using the Frobenius norm) between the covariance matrices of the source and target features after the source domain is transformed [33].

$$\begin{aligned} & min_A ||C_{\hat{S}} - C_T||_F^2 \\ & = min_A ||A^T C_S A - C_T||_F^2 \end{aligned} \tag{4}$$

where $C_{\hat{S}}$ is covariance of the transformed source features $X_S A$, C_T is covariance of the target features, $X_T A$, and $||.||_F^2$ denotes the squared matrix Frobenius norm. The transformed feature matrices are computed in the following manner:

$$\begin{aligned} C_S &= cov(X_S) + diag(size(X_S; 2)) \\ C_T &= cov(X_T) + diag(size(X_T; 2)) \\ A &= X_S \times C_S^{\frac{-1}{2}} \times C_T^{\frac{1}{2}} \end{aligned} \tag{5}$$

7.2.2 EasyAdapt. EasyAdapt is a supervised approach [7] to transform the source feature space. Let X denote the original feature space, $X = \mathbb{R}^F$. We construct an augmented feature space $\widetilde{X} = \mathbb{R}^{3F}$, by creating an industry-generic, industry\{business} and business-specific version of each feature in X. For this, we define $\Phi_{ind}, \Phi_{bus} : X \longrightarrow \widetilde{X}$ to transform feature vectors corresponding to the industry-wide and business-specific subject lines respectively. The mappings are defined by the following equation:

$$\Phi_{ind}(x) = \langle x, x, \mathbf{0} \rangle, \Phi_{bus}(x) = \langle x, \mathbf{0}, x \rangle \tag{6}$$

Here, $\mathbf{0} = \langle 0, 0, ..., 0 \rangle \in \mathbb{R}^F$ is the zero vector. Intuitively, the above transformation ensures that the model acknowledges that the same feature can have different effect on the clickthrough rate (characterized by the β values in linear regression) for different businesses.

7.3 Predictive Performance after Domain Adaptation

We train a regression model between the augmented feature space described above and the clickthrough rates per subject line. We use a corpora of 9 businesses across the three industries as the target domain set and report results from a 5–fold cross validation, comparing the MAEs and Root Mean Square Errors (RMSEs) obtained after domain adaptation, against the pre-domain adaptation condition: the best-performing predictive model (trained on the General Inquirer + n-grams features). Table 6 provides a summary of the domain adaptation results. CORAL models, being unsupervised, offers modest improvements over the pre-domain adapted models; on the other hand, supervised EasyAdapt models markedly outperform either of the previous models, with an average decrease of 16.52% in the RMSE and 19.2% in the MAE and offer significant improvements over the pre-domain adaptation condition for Finance and Cosmetics businesses. It should be kept in mind that the clickthrough rates of 95% of all emails lie within a 10% margin of error from the mean. This implies that even a 1% drop in MAE after Domain Adaptation is actually an appreciable improvement.

Table 6: Domain Adaptation results per industry, using three businesses each as unseen data. ** indicate that the model significantly improved over the pre-domain adaptation conditions in a two-tailed t-test, p < 0.01.

Training Set		Test Set	Pre-Domain Adaptation		CORAL		EasyAdapt	
Finance								
N_1 #(ind)	N_2 #(bus)	M	RMSE	MAE	RMSE	MAE	RMSE	MAE
5313	432	1728	9.7	6.6	8.0**	4.8	**7.6****	**4.6****
6059	707	707	9.8	7.4	8.9**	6.8**	**8.3****	**6.2****
6881	296	296	10.7	9.2	9.0**	7.0	8.3**	**6.7****
Cosmetics								
N_1 #(ind)	N_2 #(bus)	M	RMSE	MAE	RMSE	MAE	RMSE	MAE
991	360	360	9.7	7.4	9.2	7.0	**8.9****	**6.7****
1514	98	99	5.0	3.7	4.5	3.5	**3.3****	**2.7****
1644	33	34	6.7	5.3	6.5	5.2	**6.2****	**4.7****
Movies and Television								
N_1 #(ind)	N_2 #(bus)	M	RMSE	MAE	RMSE	MAE	RMSE	MAE
2057	227	228	5.6	4.8	5.5	4.5	**4.8****	**3.9****
2412	40	160	3.6	2.9	3.5	2.9	**3.1****	**2.4****
2057	166	389	6.1	5.1	6.0	5.0	**5.4****	**4.3****

We can visualize the implications of these changes in terms of the differences in feature coefficients (β values) before and after the better performing domain adaptation approach – EasyAdapt – is implemented. The Figure 3 demonstrates the change in the β values of a few features for the business-specific model after implementing EasyAdapt, compared to the original industry-specific model for Finance. As before, the area occupied by a square is proportional to a value's magnitude, and the color (green/red) indicates its sign (positive/negative). Consider the feature *Individuals* that has no effect on the clickthrough rate according to the industry specific model, but a large positive impact according to the business-specific model.

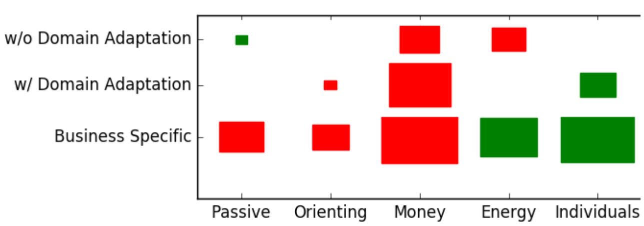

Figure 3: After domain adaptation: Hinton diagram for the β coefficients of business specific model, domain adaptation model and the industry model for a particular business in the Finance industry. Some features have no effect on the clickthrough rate according to the industry-specific model, but a large positive impact according to the business-specific model.

8 DISCUSSION AND CONCLUSION

Our paper presents the first multi-domain study treating language use in subject lines as a factor for inferring the clickthrough rate of promotional emails. We have framed prediction as a linear regression problem. Our results highlight the advantages of data-driven topic modeling in short text corpora. We also use language insights to distinguish the successful copywriting strategies in different domains and highlight the need for domain adaptation in language modeling for this problem. We implement a supervised and an unsupervised domain adaptation approach involving feature space transformations for incorporating out-of-domain data points. Not surprisingly, the supervised domain adaptation approach performs better than the unsupervised approach and considerably improves clickthrough rate prediction for unseen businesses by an average of 19%. The domain adaptation techniques ensure that the model learns a positive β value for the *Individuals* feature, by using the same industry-wide dataset. This explains how a domain adaptive approach allows a predictive model to better emulate a business' data, thereby improving its prediction accuracy for subject line clickthroughs.

In future work, we plan to apply our findings in generative models for email subject line and news headline recommendations. Our results also suggest that language modeling is useful other downstream applications in the area of marketing.

REFERENCES

[1] Douglas Aberdeen, Ondrej Pacovsky, and Andrew Slater. 2010. The learning behind gmail priority inbox. In *LCCC: NIPS 2010 Workshop on Learning on Cores, Clusters and Clouds.*
[2] Raju Balakrishnan and Rajesh Parekh. 2014. Learning to predict subject-line opens for large-scale email marketing. In *Big Data (Big Data), 2014 IEEE International Conference on.* IEEE, 579–584.
[3] Jonas Nygaard Blom and Kenneth Reinecke Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics* 76 (2015), 87–100.
[4] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop Clickbait: Detecting and preventing clickbaits in online news media. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on.* IEEE, 9–16.
[5] Yi-Ting Chen and Hung-Yu Kao. 2013. Constructing Social Intentional Corpora to Predict Click-Through Rate for Search Advertising.. In *ROCLING.*
[6] Haibin Cheng and Erick Cantu-Paz. 2010. Personalized click prediction in sponsored search. In *Proceedings of the third ACM international conference on Web search and data mining.* ACM, 351–360.
[7] Hal Daumé III. 2009. Frustratingly easy domain adaptation. *arXiv preprint arXiv:0907.1815* (2009).
[8] Kushal S Dave and Vasudeva Varma. 2010. Learning the click-through rate for rare/new ads from similar ads. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.* ACM, 897–898.
[9] Dotan Di Castro, Zohar Karnin, Liane Lewin-Eytan, and Yoelle Maarek. 2016. 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, 307–316.
[10] Ana Ferreira and Rui Chilro. 2017. What to Phish in a Subject?. In *International Conference on Financial Cryptography and Data Security.* Springer, 597–609.
[11] Alison Gianotto. 2014. Downworthy: A browser plugin to turn hyperbolic viral headlines into what they really mean. *downworthy. snipe. net* (2014).
[12] Qi Guo, Eugene Agichtein, Charles LA Clarke, and Azin Ashkan. 2008. Understanding abandoned ads: towards personalized commercial intent inference via mouse movement analysis. *Information Retrieval in Advertising* (2008).
[13] Qi Guo, Eugene Agichtein, Charles LA Clarke, and Azin Ashkan. 2009. In the mood to click? Towards inferring receptiveness to search advertising. In *Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*, Vol. 1. IEEE, 319–324.
[14] Simon Jones, Stephen Payne, Ben Hicks, James Gopsill, and Christopher Snider. 2015. Subject lines as sensors: co-word analysis of email to support the management of collaborative engineering work. In *International Conference on Engineering Design 2015 (ICED 2015).* University of Bath.

[15] Rebeca San José-Cabezudo and Carmen Camarero-Izquierdo. 2012. Determinants of opening-forwarding e-mail messages. *Journal of Advertising* 41, 2 (2012), 97–112.
[16] Sungchul Kim, Tao Qin, Tie-Yan Liu, and Hwanjo Yu. 2014. Advertiser-centric approach to understand user click behavior in sponsored search. *Information Sciences* 276 (2014), 242–254.
[17] Omer Levy and Yoav Goldberg. 2014. Dependency-Based Word Embeddings.. In *ACL (2)*. 302–308.
[18] Kwan Hui Lim, Ee-Peng Lim, Binyan Jiang, and Palakorn Achananuparp. 2016. 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, 255–260.
[19] Xiao Luo, Revanth Nadanasabapathy, A Nur Zincir-Heywood, Keith Gallant, and Janith Peduruge. 2015. Predictive Analysis on Tracking Emails for Targeted Marketing. In *International Conference on Discovery Science*. Springer, 116–130.
[20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[21] R Miller and EYA Charles. 2016. A psychological based analysis of marketing email subject lines. In *Advances in ICT for Emerging Regions (ICTer), 2016 Sixteenth International Conference on*. IEEE, 58–65.
[22] Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. Association for Computational Linguistics.
[23] Joseph E Phelps, Regina Lewis, Lynne Mobilio, David Perry, and Niranjan Raman. 2004. Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email. *Journal of advertising research* 44, 4 (2004), 333–348.
[24] Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. 2016. Clickbait detection. In *European Conference on Information Retrieval*. Springer, 810–817.
[25] Daniel Preoţiuc-Pietro, Vasileios Lampos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through Twitter content. The Association for Computational Linguistics.
[26] Rui Filipe Cerqueira Quaresma, Sílvia Paula Rosa da Silva, and Cristina Galamba Marreiros. 2013. E-mail usage practices in an organizational context: a study with Portuguese workers. *JISTEM-Journal of Information Systems and Technology Management* 10, 1 (2013), 5–20.
[27] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 521–530.
[28] Navdeep S Sahni, S Christian Wheeler, and Pradeep K Chintagunta. 2016. Personalization in Email Marketing: The Role of Non-Informative Advertising Content. (2016).
[29] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.
[30] H Andrew Schwartz, Salvatore Giorgi, Maarten Sap, Patrick Crutchley, Lyle Ungar, and Johannes Eichstaedt. 2017. DLATK: Differential Language Analysis ToolKit. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 55–60.
[31] Dong-Her Shih, Hsiu-Sen Chiang, and C David Yen. 2005. Classification methods in the detection of new malicious emails. *Information Sciences* 172, 1-2 (2005), 241–261.
[32] Philip J Stone, Dexter C Dunphy, and Marshall S Smith. 1966. The General Inquirer: A Computer Approach to Content Analysis. (1966).
[33] Baochen Sun, Jiashi Feng, and Kate Saenko. 2015. Return of frustratingly easy domain adaptation. *arXiv preprint arXiv:1511.05547* (2015).
[34] E Vance Wilson, Adrienne Hall-Phillips, and Soussan Djamasbi. 2015. Cognitive predictors of consumers' intention to comply with social marketing email appeals. 52 (2015), 307–314.