Modeling Causal Impact of Textual Style on a Targeted Goal

Gaurav Verma, Balaji Vasan Srinivasan, Shiv Kumar Saini, Niyati Chhaya
Adobe Research, India
{gaverma,balsrini,shsaini,nchhaya}@adobe.com

ABSTRACT

The consumption characteristics of a textual piece are influenced by both the *core-content* (i.e., *what* is being conveyed) and its stylistic attributes (i.e., *how* it is being conveyed). We present an approach to model stylistic attributes in text and leverage a multi-cause deconfounder model to estimate the *causal* effect of stylistic attributes towards a targeted goal. We show that our approach can identify causally significant attributes along with the ones considered important by conventional supervised approaches. Furthermore, we demonstrate using performance comparison on classification tasks that our approach does not compromise on the modeling capabilities. We believe that such a model can be valuable towards providing statistical feedback to an author to improve on certain style attributes to better achieve a target objective.

ACM Reference Format:
Gaurav Verma, Balaji Vasan Srinivasan, Shiv Kumar Saini, Niyati Chhaya. 2020. Modeling Causal Impact of Textual Style on a Targeted Goal. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3366424.3382696

1 INTRODUCTION

Textual sources on the World Wide Web, like blogs, social media posts, news articles, etc., are the primary sources of information *dissemination* and *consumption*. Before dissemination, authors often aim to position the content in the best way to achieve certain target goals. For e.g., an author may want the text to evoke a certain affect among its readers, or may want it to be popular (in terms of number of clicks or views) on a target platform (Twitter or Facebook), or may want to lure the readers towards certain subscriptions/sign ups; each of these indicate a target objective of the author.

Given such a target objective (henceforth referred as *target*), the author may need to adapt various elements of the text – both *what* is being said (i.e., aspects relating to the core-content in text) and *how* it is being conveyed (i.e., aspects relating to the stylistic attributes of text). Generally, possibility of adapting the core-content (for instance, the main topic of text) is limited since it is pre-determined; whereas, stylistic adaptations are more flexible and can be tuned towards the target. However, the style in text is often entangled with the core-content conveyed in the text because the stylistic attributes in text are not randomly assigned but chosen by an author to suit the core-content. To this end, we present an approach to estimate the causal effects of various stylistic attributes towards a target by deconfounding the influence of the core-content in text as well as other latent variables with stylistic elements. Such a causal estimate can be used to analyze and understand the stylistic aspects that made a text historically successful along the target metrics as well as to present stylistic prescriptions to the authors so that they can compose the text in a manner that better aligns to the targets.

The problem of estimating causal stylistic attributes entails two subproblems: *(a)* identifying stylistic attributes in text, and *(b)* estimating the causal effect of identified stylistic attributes while deconfounding the influence of other latent variables. Accordingly, our contribution in this work is two-fold: (1.) a structured method to enumerate and quantify key stylistic attributes identified at lexical, syntactic, and surface levels in text; (2.) a method that assumes each of the quantified stylistic attributes as a potential cause and estimates the causal effect of these attributes towards a given target.

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3382696

2 STYLISTIC ATTRIBUTES IN TEXT

Although style has been central to several natural language tasks, its modeling is often task-specific and does not cover all aspects of style. Therefore, we begin with quantifying style at three levels:

(a) **Lexical Attributes** are expressed at word-level, e.g., an author's choice of words may be more subjective than objective (*home* vs. *residence*), or more formal than informal (*palatable* vs. *tasty*).

Inspired from Brooke and Hirst (2013), we consider 8 style dimensions to quantify lexical attributes of text: colloquial vs. literary; concrete vs. abstract; subjective vs. objective; and formal vs. informal. We use a list of seed words for each of these dimensions and compute normalized pointwise mutual information index (PMI) [1]; thereby leveraging co-occurrences of words in a large corpus to obtain a raw style score for each dimension. The raw scores are normalized to obtain style vectors for every word, followed by a transformation of style vectors into k-Nearest Neighbor (kNN) graphs, where label propagation is applied. The values $\in [0, 1]$ in the estimated 8-dimensional lexical vector indicate the presence of words belonging to the 8 aforementioned style dimensions.

(b) **Syntactic Attributes** relate to the syntax of the sentence – while some authors construct complex sentences, others construct simple sentences. For e.g., a text on political discourse is likely to have compound-complex sentences, while those from a children's textbook are mostly simple sentences. Following Feng et al. (2012), we categorize syntactic style into 5 different categories – *(i)* simple *(ii)* compound *(iii)* complex *(iv)* complex-compound, *(v)* others. For quantifying these attributes, we compute the fraction of sentences that are categorized into the 5 categories by the algorithm proposed by Feng et al. Since any given sentence will definitely lie in only one of the 5 categories, the 5-dimensional vector averaged across the sentences in a corpus can be thought of as probability distribution over the 5 categories. Additionally, we also quantify the "loose-ness" and "periodic-ness" of sentences by computing the average height

Table 1: Classification performance of *cLR* (proposed) and *LR* (in parenthesis) on the EmoBank dataset.

Metric	Valence (V)	Arousal (A)	Dominance (D)
Precision	0.6105 (0.6113)	0.6503 (0.6521)	0.6922 (0.6946)
Recall	0.4884 (0.4936)	0.5911 (0.5969)	0.5513 (0.5525)
f1-score	0.5428 (0.5461)	0.6219 (0.6253)	0.6101 (0.6112)

and width of syntactic parse trees, resulting in a 7-dimensional representations of syntactic attributes.

(c) **Surface Attributes** relate to aspects like average number of *(i)* commas, *(ii)* semicolons, *(iii)* colons, (iv) periods, (v) question marks, (vi) exclamations, (vii) dashes per sentence, *(viii)* sentences in a paragraph, and *(ix)* number of total words, *(x)* number of unique words in a sentence, yielding a 10 dimensional surface-level vector.

Concatenating the individual style vectors results in a 25 dimensional representation $\mathbf{X}$ (8 lexical, 7 syntactic, and 10 surface-level) of the overall style in text. Each of these elements of the representations are interpretable and can be manifested as stylistic prescriptions to the author, in real-time. However, only a subset of these attributes *causally* influence the target. We discuss our methodology to identify these causal factors in the next section.

3 CAUSAL ESTIMATE OF STYLE IN TEXT

Given the stylistic attributes $\mathbf{X}$ and a target variable $\mathbf{y}$, our aim is to identify and estimate the causal contribution (or lack thereof) of each of the attributes. A standard approach would be to formulate this as a regression task and use the variable-importance (β-coefficients) to identify stylistic attributes that are 'influential' towards the target. However, an author's choice of style is also influenced by the core-content, for example, a topic of "mergers and acquisitions" will induce more objective words. A model that estimates effect of style on the target without any consideration of the content can be confounded by the *core-content* of the text. In other words, it is possible for such a model to attribute a content-imposed element of style to a target – which will be erroneous.

The correct approach, therefore, would be to account for the joint effect of such confounders (the aspects of core-content leading to style) and the style. To this end, we leverage the multi-cause deconfounder model [4] to obtain an estimate of the confounders and account for them to get the correct effect of changes in the style on a given target variable. Following standard procedures, we remove highly correlated features $\mathbf{X}_i \in \mathbf{X}$ and fit a probabilistic PCA factor model[1] on a training subset of standardized $\mathbf{X}$.

To assess the factor model, we hold out a part of $\mathbf{X}$ and replicate this $\mathbf{X}^{\text{rep}}$ from the probabilistic PCA model given the inferred latent variables. If the reconstruction of $\mathbf{X}^{\text{rep}}$ is successful in terms of statistical significance (i.e., p-value > 0.1 for $p(t(\mathbf{X}^{\text{rep}}) < t(\mathbf{X}))$ – where $t(.)$ denotes our test statistics), the substitute confounder $\mathbf{Z}$ which was inferred while fitting the factor model is considered valid, i.e. $\mathbf{Z}$ is a confounder. As previously mentioned, in the context of text, $\mathbf{Z}$ could represent various latent variables (like topics) that impact both style $\mathbf{X}$ and target $\mathbf{y}$. Because of difficulty in observing and/or modeling these latent confounding variables, the inferred correlations cannot be trivially extended to causations. Obtaining a valid substitute confounder allows us to correct for such unobserved latent confounders, and consequentially infer causal relationships. To do this correction, we concatenate the substitute confounder with the features (i.e., $\mathbf{X} \oplus \mathbf{Z}$) and fit a logistic regression model on the obtained vectors to predict the target variable $\mathbf{y}$. The confounder-corrected logistic regression model can then used for identifying causally significant stylistic attributes – i.e., the attributes with the associated p-value of causal significance to be < 0.05 for given $\mathbf{y}$.

[1]Wang and Blei's (2019) code: https://github.com/blei-lab/deconfounder_tutorial

Table 2: Causally significant (✓) and insignificant (✗) stylistic attributes identified by *cLR*. Attributes in **bold** are considered not considered important based on β-coefficients of *LR* but are identified as causally significant by *cLR*. Underlined attributes are causally insignificant but important based on β-coefficients of *LR*.

V	(✓)	**colon**, syntax, concrete words, period, dash
	(✗)	semicolon, syntax, concrete words, period, dash
A	(✓)	**subjective words**, colon, number of chars, exclamation, concrete words, syntax, period, dash, informal words
	(✗)	question mark, colon, number of chars, exclamation, concrete words, syntax, period, dash, informal words
D	(✓)	**abstract**, concrete words, exclamation, question mark, syntax
	(✗)	period, concrete words, exclamation, question mark, syntax

4 EXPERIMENTS

We conduct our experiments on the EmoBank dataset [2] which comprises of over 10,000 pieces of text annotated on affective metrics (Valence (V), Arousal (A), and Dominance (D)). We train two models – a conventional logistic regression (*LR*) model and a deconfounded logistic regression (*cLR*) model as described above to predict the target metrics (i.e., V, A, and D) from the stylistic attributes $\mathbf{X}$. For the ease of experimental modeling, we quantize a given target into 3 bins – indicating low, medium, and high extent of achieved target. Table 1 summarizes the performance of both these models and shows that the proposed model (*cLR*) performs equally well as a conventional logistic regression model. However, as we demonstrate in Table 2, *cLR* allows identification of *causally* significant as well as insignificant stylistic attributes which *LR* does not. Furthermore, the differences in identified elements in Table 2 are well aligned with our hypothesis that attribution of a specific style to a target solely based on learned β-coefficients of *LR* may yield erroneous results. For example, the use of abstract words like patriotism, inspirational, etc., intuitively increases the dominance (D) of the text – something which was not identified by *LR*.

We have thus established the ability of our proposed approach to estimate causal effect of stylistic attributes towards a target goal without compromising on the classification-related modeling aspects. We believe that this causal estimation approach along with structured stylistic modeling described in Section 2 presents a unique method to analyze text and build real-time systems that can assist authors using causal prescriptions towards a target goal.

REFERENCES

[1] Julian Brooke and Graeme Hirst. 2013. A multi-dimensional Bayesian approach to lexical style. In *NAACL*.

[2] Sven Buechel and Udo Hahn. 2017. EMOBANK: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *EACL*.

[3] Song Feng, Ritwik Banerjee, and Yejin Choi. 2012. Characterizing stylistic elements in syntactic structure. In *EMNLP*.

[4] Yixin Wang and David M Blei. 2019. The blessings of multiple causes. *J. Amer. Statist. Assoc.* (2019).