Keenan Venuti

Harvard N.E.R.D. Business School Lab

Quantifying Emotional Engagement: Finding the Holy Grail of Headlines – Study Proposal

Abstract:

At its heart, marketing is finding the right audience and delivering the best message to pursue future business or a sale. Digital marketing follows this step by step and iterative process of lead generation, lead funneling followed by customer messaging. In 2011, advertisers spent a total of 1.3 billion dollars on online advertisements (Chen & Stallaert, 2014). In 2014, that doubled to 2.6 billion dollars. Despite the massive budget online advertising garners, there are two distinct camps on how to influence the consumer. Many ad-tech companies focus on targeting first and foremost; this is for good reason, as behavioral targeting alone can increase click through rates by 670% (Chen et al, 2014). Other companies create advertisement campaigns focused around the creative, usually some type of picture or headline. Marketing studies have shown that there are certain words that people engage with more than others in a headline (Safran, 2013). While many have focused on the individual message and others have invested in reaching the right audience, there exists little market analysis comparisons of the content different groups engage with. My goal is to find significant trends within and between consumer groups to better understand online consumer behavior by demographic and the type of advertisement headlines they engage in. I predict there are significantly different words used in headlines leading to higher engagement between groups of consumers.

Background/Literature Review:

As hard copy content is rapidly replaced with online content, advertisers have found that adapting their campaign strategies is not a luxury, it is a necessity. The days where an advertiser could expect to generate decent customer interest in sales from a newspaper are gone; in the web driven world that encompasses openly accessible content, the best way to monetize is to market products to consumers engaging with that content. The total online advertising budget has doubled within recent years, and for good reason. Online advertising is the most cost effective way for a brand to distribute its message, generate consumer interest, and create sales. With techniques like targeting, online advertisers often know exactly who will read their message. Some companies take this opportunity to deliver a message referencing why the consumer was targeted. This demographic centric specificity can dissuade people from engaging in the advertisement (Goldfarb & Tucker, 2011). In the online advertising context engagement is a general metric considering how many people have clicked on an advertisement. Findings regarding targeted headlines that cause less engagement should not discourage marketers from targeting their advertisements as precisely as they can; instead it should encourage digital marketers to look for the optimized message to show that consumer subgroup.

Targeting

While there are many ways to target within online advertising, they all follow the same general format. There exist N amount of total available advertisement placements in the digital world. When a brand, through an advertisement platform, chooses to pay to place an advertisement in any given position, it is their hope that the consumer will see that advertisement and engage by clicking on it. Targeting is the methodology of reducing the space of the N total available advertisement places, based on presumptions as to who will click on the advertisement, who will buy the product, and who the brand wants to reach with their message.

Researchers have empirically shown that targeting serves as a benefit to an advertising platform when one of three conditions are met (Athey & Gans, 2010). It is beneficial when there are limited advertisement placement spaces or each of those placements is costly to fill or if there are diverse advertising brands competing for an individual advertising space. Finally, targeting is optimal when there is a highly competitive market of advertising brands (Bergemann & Bonatti, 2011). Yet, there is a tradeoff of targeting from an advertisement platform's point of view. Targeted advertisements generally generate more revenue because more relevant consumers see the advertisement leading to higher engagement (Chen et al, 2014). But as targeting gets more specific, the total number of advertising brands trying to purchase that slot is reduced. This creates less overall competition and drives down the price of the advertising spot. However, the advertising brands prefer higher engagement rates in an advertisement as an increase in rates usually means the brand has to spend less for the individual engagement (Goldfarb & Tucker, 2011). Instead of paying for their advertisement to appear on a consumer's computer, who has no intention of engaging, that impression would not be wasted with proper targeting. Targeting the consumer appropriately and delivering any message does not guarantee an online advertisement's success.

The Headline

Most advertising brands have a creative team dedicated to writing and editing headlines that will be run through an advertisement platform. This process is often plagued by assumptions as to what a certain demographic will enjoy or find stimulating, rather than empirically analyzing what has worked in the past. Often, digital marketers believe crafting a specific message within a targeted subgroup should be based on the subgroup itself. This can actually cause less engagement; consumers do not engage with an advertisement when they feel targeted (Goldfarb et al, 2011). In a similar study, researchers found that hyper specific targeting and an advertisement message communicating the targeting process annoyed consumers and decreased total engagement based on the advertisements, “creepiness” (Goldfarb et al, 2011). Again, the problem is not with crafting individual messages after targeting, but delivering the message that fits the consumer subgroup best. For example, one study found that matching an advertisement, displaying on a publisher website, to the content of that website increased purchase intent (Goldfarb et al, 2011). Appropriate consumer messaging based on targeting works, but only when the consumer does not feel their privacy has been violated. Digital marketers should search for engagement trends among similar consumers rather than creating advertisements based on intuitions exposing the targeting process. These trends do exist, and with the proper data analysis and dataset, marketers can better understand how to reach their target audience. Those audience subgroups share similar “web behavior”. In 2009, researchers Yan, Liu, Wang, Zhang, Jiang, and Chen, found that consumers who clicked on the same advertisement exhibited similar web behavior. It follows that trends among demographically created consumer groups will also exist within headline engagement.

While there has been little research dedicated to the type of headlines different consumer subgroups engage in, marketers have found words that lead to higher engagement rates across all consumers. One marketing group found that headlines with numbers, reader-addressing statements (using you within the headline), and how-to headlines shared high rates of engagement (Safran, 2013). A different group found that people tend to engage with headlines when they contain interesting adjectives, negative words, specificity, and numbers (Patel, 2104).

A third group found four categories of words associated with higher engagement rates (Truong, 2015). By using words focused on time, insight, motion, or space, an advertising brand may be able to increase ad engagement. There is clearly content within headlines that lowers engagement and only a few alternate forms of content that increase engagement rates. However, a lot of marketing research regarding words that lead to higher rates of engagement have focused on simultaneous analysis of all consumers rather than subdividing and finding trends specific to each demographic consumer subgroup.

Study Proposal:

My goal is to combine and add to previous research in the digital marketing field focusing on headline-centric online advertisements. I'd like to begin by segmenting consumers based on demographics such as sex, location, age, and any other details provided through an advertising platform. Then, I plan on quantifying the naturally qualitative headline. To do this, I will convert each headline into a vector representing word proportionalities via several different methods. Finally, I plan on running several regression models associating different headlines, now represented as vectors, with engagement rates using a supervised learning algorithm. From there, any model created will yield word sets that drive individual consumer subgroups to engage with an online advertisement. I plan on measuring the accuracy of these models using basic statistical analyses, specifically R-Squared, the Akaike Information Criterion, and the Bayesian Information Criterion.

Method:

This project depends on obtaining and analyzing online advertisement datasets. To obtain these datasets, I plan on contacting ad-tech and marketing group companies that consistently run similar online advertisement campaigns. In exchange for using their data, I will provide them with any insights I find. With datasets that record targeting information, I can extract which consumer groups received which advertisements. I plan on segmenting the entire dataset into broad consumer subgroups, beginning with things like age and gender. From there I will continue to specify smaller and smaller subgroups by eliminating proportions of each general group. I will stop specifying subgroups once the segments are too small to return helpful feedback. With multiple consumer segments, ranging from highly specific to generally broad, I can begin the headline vectorization and analysis process.

Headline Vectorization

Language is naturally qualitative, and advertisement headlines, a form of language communication, inherit the difficulties of analyzing word. However, the process of transforming words into a quantifiable nature, to be used in statistical analyses, carries vast amounts of research and experimentation. Most researches concerned with text analysis choose one of three popular methods.

Linguistic Inquiry Word Count - LIWC

The LIWC is a categorically formatted dictionary program able to sort words into preset categories (Pennebaker, 2011). Professor James Pennebaker created the LIWC to analyze writing samples of people across different demographics. He found that there are subtle yet distinct differences in how different people write; his research included books, scripts, tweets, etc.. His software simply categorizes words and derives their proportionality. Each word, in each advertisement headline, would be counted, sorted, and grouped into one or more of several categories. Then, measuring the proportionality of these categories, a vector representing the proportion of every word type is created. A study aimed at predicting author age using several regressive models on authors' texts used this vectorization technique (Nguyen, Smith, & Rosé, 2011). Researchers found that the LIWC worked well as a means to formatting text quantitatively. Ad-tech startup Sharethrough used the LIWC dictionary as well, when running general analyses on headlines (Truong, 2015). They found significantly higher proportions of four types of words in high engagement headlines.

Entire Headline Text

A more common method is to create each headline vector based on proportionalities of words in the entire headline text (Taddy, 2013). In this method, the entire headline training set is analyzed as a large text file. From there, any repeated words are deleted; the remaining text consists of every unique word, used in every headline. Researchers will then eliminate stop words, or words that carry very little predictive power. These are words like "the", "a", "an", "at", etc. Finally, in an attempt to further reduce the size of each vector to represent each headline, most researchers will stem each word and eliminate repeated stems. Stemming is the process of eliminating suffixes to extract the root of a word. "Jumping", "jumped", and "jumps" all become "jump" by the stemming process. With this set of stemmed unique words, researchers can then look through and count how often each word appears in a headline. Each headline is then run through a similar cleaning process and a vector is created representing the headline by a count of frequency of each word.

Alternate Dictionaries

The final method I will use is less common, but still carries insightful value. This method is similar to the LIWC approach. However, it involves using different language dictionaries to measure proportionalities of different types of words (Nguyen et al, 2011). For example, one dictionary, the Stanford Pos-Neg, categorizes words based on degrees of positivity or negativity (Nguyen et al, 2011). There is no one absolutely correct way to group words together; trying different combinations based on different aspects of word similarities can lead to alternate and possibly valuable trend analyses. As does the LIWC, this system creates a vector for each headline based on proportions of word groups. Similarly, alternate dictionaries exist, making it worthwhile to shop these models, finding which categorization method yields the best predictive model.

Model Creation

With each headline represented by a quantitative vector, I will be able to begin the regressive text analysis. Many forms of text and sentiment analysis are focused around classification; this involves associating certain headlines with a categorical class or binary class (Taddy, 2013). Because my model is meant to predict headline performance based on engagement rates, I will have to use less researched text regression methods. Most research on textual regression uses a form of Linear Regression as a model. Textual analysis can be statistically complicated due to high dimensionality problems. When the number of features, also called predictors, represented in this study by the proportionalities of each word in each headline, is larger than total dataset samples, overtraining can occur. A supervised learning model will learn minutia details to distinguish different headlines in the training dataset based on slight differences between features. Then when testing the model's accuracy on a testing dataset, these same subtle rules of distinguishability no longer apply, creating a model that has poor external applicability. In textual analysis, the number of features is often very large as to accommodate the sorting of very different types of words. For example, the LIWC alone has 70 distinct categories (Pennebaker, 2011). If a segmented consumer subgroup consists of only 30 samples, overtraining will occur. To combat overtraining, I will use one of three regularization techniques previously tested to be effective in textual analysis.

Principle Components Analysis – PCA with Linear Regression

A popular method of dimensionality reduction is PCA. PCA examines every sample and every feature used for prediction and creates a linear model that best represents the most important or effectual features (Taddy, 2013). For example, words like "buy", "have", or "now" may carry more predictive value than words like, "agree", "fun", or "cool". The process of PCA creates a linear equation that values highly effectual words and devalues less important ones. From there, after manually choosing how many components the PCA should contain, its linear outputted equation serves as a reduced feature set for the highly dimensional textual dataset. This reduced dataset is ready for regression. Linear Regression accounts for each feature used in headline engagement rate prediction and maps out a linear predictive equation using a set of features and their associated engagement rate as input (Taddy, 2013). After running PCA, this model finds linear associations of decreases and increases of feature values within each headline vector paired with lower or higher engagement rates. It then uses those associations to estimate an engagement rate when testing headline vectors. PCA Linear Regression was used by researchers estimating ordinal restaurant review star scores based on textual reviews (Taddy, 2013). Linear Regression is the standard for textual regression, but PCA isn't the only way to reduce dimensionality.

Ridge and Lasso Linear Regression

Ridge and Lasso Linear Regression are both performed similarly with slight equational differences. Both methods employ a loss function that creates a penalty when the model uses a large number of predictors (Liu, 2012). This penalty helps mitigate highly dimensional datasets; as the model over-utilizes the features, which can lead to overtraining, the loss function reduces the model's predictive ability. Ridge Regression uses a more strict loss function than Lasso Regression. By using a chosen lambda multiplied by the sum of every feature squared and applying that value to the Linear Regression residual sums squared, the ridge regression reduces the impact of every feature coefficient used in the Linear Regression model. One study seeking to value the scale of positivity in online texts used Ridge Regression found promising results (Liu, 2012). Lasso Regression utilizes a less strict loss function. The Lasso loss function applies a chosen lambda multiplied by the sum of the absolute value of every feature; given the absolute value of a feature will always be less than the square, when that feature is greater than one, this loss function reduces the impact of every feature coefficient used in the Linear Regression model to a lesser degree. A study looking to predict author age based on writing samples successfully implemented a Lasso Linear Regression (Nguyen et al, 2011).

Support Vector Regression -SVR

A less researched textual regression technique showing some process is SVR. This model involves somewhat more complicated statistical modifications and may be out of scope for this project. SVR begins by utilizing a hyperplane, usually drawn to separate classes for classification. That hyperplane is then remodeled in its position and shape to reflect trends within the dataset leading to stronger predictions. Although it may not be used in the final study, two studies found that SVR was an effective method of textual analysis regression. One set of researchers used SVR to predict stock prices from news headlines, while others attempted to find positivity of internet posts based on blogs using SVR (Liu, 2012)(Kirange & Deshmukh, 2016).

Analysis of Output

To measure the accuracy of these models, I will divide each dataset and sub dataset into training and testing datasets. From there I plan on training one of the Linear Regression models on the training dataset headline vectors, associated with an engagement rate. Using the testing dataset headline vectors as input, my model will predict engagement rates based on the learned statistical model. To test for accuracy, I will compare predicted testing dataset engagement rates to actual testing dataset engagement rates using R-Squared, the Akaike Information Criterion, and the Bayesian Information Criterion. Both AIC and BIC are specially designed to account for highly dimensional datasets and will serve as proper measures of accuracy. These measures of accuracy in tandem will show which consumer demographic subgroups engage in similar content and which subgroups exhibit more erratic online engagement behavior. Each model created for each segment will provide feedback as to the types of words that consumer segments tend to engage in. With knowledge of high engagement words by segment and predictive accuracy between subgroups, my model will gain the ability to predict how well a headline will perform when targeted to a certain demographic of consumer.

Results:

Due to the lack of research of online advertisement performance by consumer subgroup segments, predictions regarding future results are at best guesses as to the type of content people tend to engage with. However, headline analysis studies within general datasets considering all demographics have yielded significant trends regarding the types of words that encourage people to click on online advertisements. Given these prior marketing studies, I expect to find certain words to be more popular within each demographic subgroup. I do anticipate commonality of these words across groups; I also predict a general set of words that increase and decrease engagement within all demographic segments. I plan on representing these results within a dictionary of my own creation. One that sorts and categorizes words that lead to higher engagement among consumer segments. Each category will represent a consumer segment and a positive or negative mark, based on those words' ability to increase or decrease headline engagement when presented to that demographic subgroup.

Discussion:

With a well-designed model, a complete dataset, and functional vectorization techniques, research marketers will be able to extrapolate conclusions to more efficiently run textual online targeted advertisements. As studies have shown, there is a fine line and a few words within a headline that will drive engagement or reduce it. Given my model, it is my hope that marketing researchers can assess a proper strategy when pursuing a group of people, separated by demographic. Those marketers will be able to hand craft a message that they know will lead to higher engagement within that specific segment. Marketers will also be able to determine when a similar message or the same message will perform well within other consumer subgroups. By measuring model accuracy, marketers can determine if they want to take an empirical approach to headline creation or would rather try several different forms of creative headline copy. For subgroups with highly erratic web behavior, a marketer may use these insights to determine a need for several different types of headlines, as those subgroups show no significant engagement trends. It is my hope that marketing researchers can take this information and apply it to the already highly effective targeted online campaign. From there, they can avoid advertising faux pas, such as exposing their targeting mechanisms via headline or using certain words that decrease engagement rates.

Conclusion:

As cost decreases and targeting efficiency increases, a growing number of advertising campaigns have become exclusively online. As consumers are exposed to different forms of advertisements, such as banner, pop-up, native, search, and social, finding the best message to send to the right people is more important than ever. The best message to send isn't always intuitive, overly targeted messages can turn consumers off of a product because of how "creepy" the advertisement seems (Goldfarb et al, 2011). Marketers have a responsibility to the brands they are advertising and to those who see the advertisement to not only properly target consumers with products they will want to engage in, but deliver the message that those consumers want to see and click on. When done correctly, targeted advertising and message selection can form a win-win-win for the advertising platform, the advertising brand, and the consumer. I hope to contribute to the developing research done by ad-tech companies, marketing consultants, and companies looking to advertise online by providing an analysis of what type of content different consumers tend to engage with. I also hope to find similarities in advertisement headlines so companies can target demographics more efficiently without wasting impressions or annoying the consumer. If my model does hold predictive value, I plan on creating a tool that can better estimate a headline's performance before it reaches the market.

Future Research

This proposal serves as a general guideline highlighting how researching marketers may conduct analysis of the online advertisements they produce and their effect on segments of the population. Generally speaking, this process can be cross applied to other forms of creative and other types of advertisement efficacy ratings. To being, engagement rate or clicks over impressions, is not the only or best way to measure advertisement efficiency. Marketers are constantly calculating and creating new metrics that better represent consumer online behavior quantitatively. In the future, as computation decreases in expense, I anticipate a similar analysis using online advertisement images with computer image processing to find the types of pictures people tend to engage with. Finally, given the already well established forms of text analysis classification, utilizing a categorical classification learning process instead of a quantitative regression one may yield more statically sound and diverse insights. This paper and the research that follows should serve as a stepping stone and inspiration for similar marketing analysis projects.

References

Athey, S., & Gans, J. S. (2010). The impact of targeting technology on advertising markets and media competition. Available at SSRN 1535325.

Bergemann, D., & Bonatti, A. (2011). Targeting in advertising markets: implications for offline versus online media. The RAND Journal of Economics, 42(3), 417-443.

Chen, J., & Stallaert, J. (2014). An economic analysis of online advertising using behavioral targeting. Mis Quarterly, 38(2), 429-449.

Goldfarb, A., & Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. Marketing Science, 30(3), 389-404.

Goldfarb, A., & Tucker, C. E. (2011). Online advertising, behavioral targeting, and privacy. Communications of the ACM, 54(5), 25-27.

Guha, S., Cheng, B., & Francis, P. (2010, November). Challenges in measuring online advertising systems. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (pp. 81-87). ACM.

Kirange, D.K., Deshmukh, R. R. (2016) Sentiment Analysis of News Headlines for Stock Price Prediction. Compusoft.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Manchanda, P., Dubé, J. P., Goh, K. Y., & Chintagunta, P. K. (2006). The effect of banner advertising on internet purchasing. Journal of Marketing Research, 43(1), 98-108.

Nguyen, D., Smith, N. A., & Rosé, C. P. (2011, June). Author age prediction from text using linear regression. In Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (pp. 115-123). Association for Computational Linguistics.

Patel, N. (2104). "The Formula for a Perfect Headline." Quicksprout. Web. 06 Nov. 2016.

Pennebaker, J. W. (2011). The secret life of pronouns. New Scientist, 211(2828), 42-45.

Safran, N. (2013). "5 Data Insights into the Headlines Readers Click." Moz. Web. 06 Nov. 2016.

Taddy, M. (2013). Multinomial inverse regression for text analysis. Journal of the American Statistical Association, 108(503), 755-770.

Tamir, M. (2016). "A Short Introduction to Using Word2Vec for Text Classification." LinkedIn. Web. 06 Nov. 2016.

Truong, A. (2015). "These Are the Words People Can’t Resist in a Headline." Quartz. Web. 06 Nov. 2016.

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., & Chen, Z. (2009, April). How much can behavioral targeting help online advertising?. In Proceedings of the 18th international conference on World wide web (pp. 261-270). ACM.