Predicting Meme Success with Linguistic Features in a Multilayer Backpropagation Network

Keith T. Shubeck (kshubeck@memphis.edu)
Stephanie Huette (shuette@memphis.edu)
Department of Psychology, 202 Psychology Building
Memphis, TN 38152 USA

Abstract

The challenge of predicting meme success has gained attention from researchers, largely due to the increased availability of social media data. Many models focus on structural features of online social networks as predictors of meme success. The current work takes a different approach, predicting meme success from linguistic features. We propose predictive power is gained by grounding memes in theories of working memory, emotion, memory, and psycholinguistics. The linguistic content of several memes were analyzed with linguistic analysis tools. These features were then trained with a multilayer supervised backpropagation network. A set of new memes was used to test the generalization of the network. Results indicated the network was able to generalize the linguistic features in order to predict success at greater than chance levels (80% accuracy). Linguistic features appear to be enough to predict meme transmission success without any information about social network structure.

Keywords: meme prediction; psycholinguistics; neural networks

Introduction

The term "meme" was originally coined by Richard Dawkins in his book, The Selfish Gene. Dawkins, an evolutionary biologist, describes "meme" as a unit for carrying cultural ideas or behavior, similar to how genes carry genetic information from one generation to the next. Just as genes propagate from organism to organism, memes propagate from mind to mind by way of communication and social learning (Dawkins, 1989). Under this lens, memes are also subject to mutations, where each mutation either strengthens or weakens the meme's fitness. Blackmore (1998) argues for maintaining the original definition of meme, one that emphasizes *imitation* as the means of meme transmission. Blackmore (1998) goes on to explain that a meme is first internalized in the receiver and can then be reproduced. Heintz and Claidière (2014) argue that memes, or replicators, compete with one another for an individual's limited cognitive resources for the chance to replicate again. Thus, some memes will fall into obscurity where others will flourish. With this in mind, successful memes should be those that are easily memorable. Analyzing the properties and features of memes that may influence their fitness has proven to be a challenging endeavor, especially prior to the establishment of various online social networks.

The internet, and more specifically social media, provides researchers interested in the study of information diffusion, meme propagation, and cultural transmission a means to observe these concepts in an ecologically valid setting and on a massive scale. Our understanding of meme propagation runs parallel with our understanding of human culture; the more we understand about memes and their mutations, their origins, and how quickly these are accepted by other individuals, the more we will understand cultural trends that may have been previously considered bewilderingly anomalous. The challenge then becomes for researchers to develop robust and valid methods for detecting memes, tracking their mutations, and predicting their success. The current model attempts to develop a method for predicting meme success by analyzing its linguistic and resultant features. Features such as length, concreteness, and orthographic features such as misspellings may all contribute to cognitive and emotional factors that would predict transmission of a meme to some degree.

The challenge of detecting and tracking memes has been approached in a variety of ways, with varying success. The broad and encompassing nature of the definition for meme has resulted in the term being operationalized differently from study to study. In addition to the changing operational definitions, the domains of meme studies also vary. For example, some studies focus on visual or video content such as YouTube memes (Shifman, 2012; Xie, Nastev, Kender, Hill & Smith, 2011), and others on textual memes, like quoted text in the news cycle (Simmons, Adamic, & Adar, 2011; Leskovec, Backstrom, & Kleinberg, 2009). Other research has focused on microblogging memes in social networks such as Twitter or Yahoo! Meme (Ratkiewicz et al., 2010; Adamic, Lento, Adar & Ng, 2014; Tsur & Rappoport, 2012; Ienco, Bonchi, & Castillo, 2010). For our purposes here, we will focus on popular text-based memes, of which some have visual components that were not included in the model, and others simply contain text.

Another recent study set out with the goal of predicting meme success by observing the meme's early spreading patterns within Twitter (Weng, Menczer &, Ahn, 2014). The authors chose to focus on the structure of the meme's environment because previous research has shown that the structure of underlying networks impacts the spreading process of information (Daley & Kendall 1964; Barrat, Barthelemy, & Vespignani, 2008). Design features of the website itself (i.e., user voting feature on Digg) can also be used to improve meme prediction (Hogg & Lerman, 2012). Weng et al. (2014) operationalize meme success by observing the meme's overall popularity, relative to the other memes in their dataset. They operationalize "meme" as any hashtag observed in their dataset. Hashtags are strings of text following a "#" users insert into their tweets (i.e., short user submitted posts within Twitter) for labeling purposes. Popular hashtags are tracked by Twitter and said to be "trending". Here, the definition of a successful meme is determined by the frequency of usage and overall popularity of that meme.

Weng et al. (2014) found that using topographic, or structural, features of the network enabled their model to accurately predict a meme's popularity up to two months in advance. These topographical features included "community size", where a community is a set of nodes (i.e., individual users) who are followers of one another, and "network surface" (i.e., neighbors of the audience of users). The model used by Weng et al. (2014) is similar to other studies that include user influence in understanding information diffusion (see Romero, Meeder, & Kleinberg, 2011).

Unfortunately, studies that include user influence (i.e., number of followers a given user has, number of those followers' followers, etc.) as a key component of their meme predicting model add little to our understanding of why certain memes are selected and become popular, and why other memes are unsuccessful. We argue that an important question remains unanswered: are there linguistic features and aspects of cognition that can predict the ultimate success of a meme, outside of the characteristics of the social network?

Tsur and Rappoport (2012) attempt to answer that question by taking a closer look at the content of Twitter hashtags in order to predict their popularity. Their study places emphasis on the content features of a meme in determining its popularity, something that prior to their 2012 study, has been largely ignored. Secondly, by stepping away from the costly graph based algorithms, used in the studies mentioned above, Tsur and Rappoport (2012) provide a simple and more global approach for modeling meme acceptance and popularity. The content features that were examined included: hashtag length (number of characters and words), hashtag orthography, emotional content and linguistic cognitive features taken from the Linguistic Inquiry and Word Count Tool, or LIWC. LIWC (http://www.liwc.net/) is a linguistic tool that counts the number of words in various categories that have been built upon relevant communicative dimensions (Tausczik & Pennebaker, 2010). The categories of the program are the essential feature, as they contain a collection of words that fit into 80 validated word categories, ranging from emotion word categories to deception word categories. Using a regression model, with the above mentioned features, they found that the cognitive category of words from LIWC was positively correlated with the hashtag's popularity, when the hashtag's content was also taken into account. For example, the word "think", a cognitive process, would predict increased popularity compared to a non-cognitive word, like "ball". They also found that lengthier hashtags were not as popular as shorter hashtags. They attributed this finding to cognitive load theory and physical constraints for tweets (i.e., 140 character limit per tweet). Cognitive load theory posits that during an instance of complex learning, an individual may be underloaded or overloaded with information, due to the working memory limitations. While these findings are promising, Tsur and Rappoport (2012) point out that future studies using the content of memes to predict success should delve deeper into the psycholinguistic aspects of the content and the cognitive constraints of the receiver of the meme.

These models often posit the relevant connections of meme transmission are between people, but this neglects what happens within an individual's mind when a meme is encountered. Further, language is context sensitive, and at least partially grounded in perceptual-motor features that enrich complex linguistic representations (Huette & Anderson, 2012). The factors contributing to whether the meme is transmitted, or not transmitted, is most likely the product of an interaction of an individual with their environment, thus cognitive factors contribute as well as social factors. However, if the person decides to not transmit the meme further, the number of connections to the user no longer matter and thus are of primary concern to understanding meme transmission. The current work is at the cognitive level of analysis, where connections constitute an information space inside of an individual, and success is determined by whether or not the individual is likely to engage in further transmission of the meme.

The advantage of neural networks over rule-based systems is they are able to solve more complex problems and carve up the solution's space in unanticipated ways. For example, cognitive process words may somewhat predict meme success, but a combination of cognitive process words, emotion words, concreteness, etc. might be interacting in non-intuitive ways that contribute to transmission or non-transmission of the meme. To demonstrate this, we predicted a binary logistic regression would not yield as much predictive power as the neural network model. Neural networks are able to come up with solutions that do not rely on linear or singular relationships or causality, allowing for complex interactions which are well known to be commonplace in thinking, communication, and behavior. Performance of a binary logistic regression will be compared to neural network performance to test this prediction.

Model

Meme Corpus

Memes were collected from the meme wiki-style website, knowyourmeme.com, and were represented as 15 input nodes with binary values. Each element of the input vector represented a linguistic or cognitive variable of the meme that was theoretically and empirically motivated to have an impact on the meme's popularity. The target outputs consisted of two binary winner-takes-all nodes, where one represented "successful" and the other represented "unsuccessful". Meme success was determined by using the number of Google search results of a meme phrase, verbatim. This was similar to the way that hashtag searches were used in the aforementioned Twitter meme studies.

In order to reduce noise in the number of inaccurate result hits, a time range filter was placed on each meme search, based on the month the meme search queries first spiked. This was determined by using Google Trends, which allows users to show how often a particular search term is entered in Google search, over time. If a meme's search queries first began to spike in October of 2009, then the search was limited to October 2009 to the present date. After determining the total number of search results provided for each individual meme, a median split was applied to the data to separate successful memes from unsuccessful memes. For this particular data set, memes that had 37,400 or more search results were considered successful, and any memes below that threshold were considered unsuccessful. Of course all memes were retransmitted to some degree, so this label might be something more akin to "more popular" and "less popular" when discussing memes as a whole. Importantly, the distribution of popularity was exponential, with successful memes being exponentially more popular than unsuccessful memes.

Training set. The dataset used to train the network consisted of 268 established memes collected from knowyourmeme.com, a meme encyclopedia, which uses the wiki web application to collect and categorize various internet memes. The memes included in our corpus contain hashtag memes (e.g., #YOLO), copy-and-paste memes (e.g., Repost this if you're a big black woman who don't need no man), as well as lesser known memes commonly used in smaller online communities (e.g., burst into treats). The average meme word length was roughly four words per meme, with the longest meme having 31 words. Copy-and-paste memes were divided into smaller chunks of text, each chunk having at most one complete sentence. In general, the memes used for the current study are phenotypic memes, meaning their raw text contains the best estimate of the "original" meme. Variants of these phenotypic memes were not included. If it could not be clearly determined which meme came first, then both memes were included separately in the dataset. The linguistic and cognitive properties of the meme text were broken down into 15 binary features that can be categorized as: psycholinguistic features, physical features, orthographical features and meme type. These features were chosen on the basis of sentence processing and memory literature.

Psycholinguistic Features. Eight psycholinguistic features were chosen as meme features. These features were selected based on current cognitive psychology and psycholinguistic theories centered on sentence recall, working memory, and how emotion and arousal affect memory.

Mean word concreteness was determined through the use of Coh-Metrix, (http://cohmetrix.com/) a validated linguistic analysis tool that is able to automatically analyze text for features such as text cohesion, parts of speech, word frequency, lexical diversity, and syntactic complexity (McNamara, Kulikowich, & Graesser, 2011). Concreteness was chosen as a psycholinguistic feature for the current model because previous research has shown that concrete words are easier to recall than abstract words during a short-term serial recall task (Walker & Hulme, 1999). Memes that are easier to recall and more concrete should have a distinct advantage over memes that are more difficult to recall. If a given meme had more concrete terms than abstract terms then it was coded as concrete (1), if it contained no concrete terms, or more abstract terms, then it was coded as abstract (0).

The overall emotional arousal of a meme was determined through the use of the LIWC (Linguistic Analysis and Word Count; Pennebaker, Francis, & Booth, 2001). LIWC's affect dictionaries were based on the emotion rating scales developed by Watson, Clark, and Tellegen (1988). For this feature, if a meme included an emotional word, either positive or negative, it was considered an emotional meme (1), and if the meme contained no emotion words then it was considered a non-emotion meme (0). The emotional arousal feature was included in the current model because previous research has shown emotional arousal, in general, has an impact on long term declarative memory (Cahill & McGaugh, 1998).

Four other finer-grained emotional features were also recorded for each meme. These features were used to determine 1) whether or not positive emotion was present, 2) whether or not negative emotion was present, 3) whether there was more positive emotion than negative emotion and, 4) whether there was more negative emotion than positive emotion. Negative emotion has been found to enhance memory accuracy for specific details during a recall task (Kensinger, 2007). However, the broaden-and-build hypothesis posits that positive moods broaden an individual's scope of attention and thought-action repertoires, whereas negative moods tend to narrow an individual's scope of attention and associations between thoughts and actions (Fredrickson & Branigan, 2005).

In their study, Tsur & Rappoport (2012) chose to include LIWC's "cognitive" categories. They hypothesized that this category should contain words that prompt or encourage specific behaviors (e.g., cause, know, ought). However, overall Tsur & Rappoport found that the more general cognitive category only marginally improved the MSE over the baseline. For the current study we chose to include the more specific "CogMech" LIWC category (i.e., cognitive mechanism) with the hope of improving the overall model.

The last psycholinguistic feature included involves the presence (1) or absence (0) of curse words, or taboo words, in the meme. LIWC was used to determine the presence of curse words in the set memes. LIWC's swear word category includes a set of socially proscribed derogatory or profane words. A slew of previous research has shown that emotionally arousing words, particularly taboo words, are remembered better than neutral or nonarousing words (see Kensinger, 2007 for a review). Memes with curse words should have a distinct advantage over memes without curse words, in terms of the meme's ability to be recalled.

Physical & Orthographical Features. Two physical features of the meme text were also recorded. Intuitively, memory span is inversely related to word length, and words that take longer to read or speak are more difficult to recall in simple recall tasks (Baddeley, Thomson, & Buchanan, 1975). Memes that contained less than four words were considered short (1) and memes that contained four or more words were considered long (0). Additionally, memes that contained words that all had less than three syllables were considered short (1), and memes that contained a word with 3 or more syllables were considered long (0). Shorter and less complex memes should be easier to recall, improving their fitness and overall success.

Two orthographical features were included based on the intuition that slang terms, purposeful word misspellings, or purposeful incorrect grammar usage should set some memes apart from others. Words with incorrect spelling, or novel words and phrases should stand out more than correct word spellings and established words and phrasings. If memes are competing for attention, then memes with novel words or phrases should tend to be more popular or successful than memes using traditional spelling and phrasing.

Meme Type. Finally, three meme type features were coded. The three meme types consist of template memes, copy-and-paste memes, and game memes. These were three different features all mutually exclusive and determined during the search process. Examples of game meme are "The object to your left will be your only weapon during a zombie apocalypse" or "You are now manually breathing". An example of a template meme is provided in Figure 1.

Figure 1: An example of a template meme. The text varies from iteration to iteration, but the image remains static. Text here emphasizes awkward social behaviors.

Network Structure

The current model used a 4-layer backpropagation network that was designed to take linguistic features as inputs and classify them as either successful or unsuccessful. The neural network used to predict meme success consists of four layers: an input layer with 15 nodes encoded in a binary manner, two hidden layers with 20 nodes each, and an output layer with two nodes that represent the probability of success of the meme. The targets for the output nodes were mutually exclusive, however it is possible that the network could generate either high or low probabilities for both successful and unsuccessful nodes. There were a total of 268 memes used to train the network. Network weights were trained on each meme 3000 times in a randomized order, and weights were modified after each learning instance using the delta rule. If the popularity of the meme was high, the "successful" node was set to 1 and "unsuccessful" to 0, and vice versa for unpopular memes. This value was determined by using a median split on the popularity of the meme, where highly transmitted memes were considered successful, and more infrequent memes were less likely to be retransmitted. Learning rate was set to .001, and the momentum term was set to 0.2. These were determined based on the observation the network learned very quickly, and were used to prevent over-fitting. The network reached an average Mean Squared Error of .228. Matlab coding of the network is available from the first author upon request.

Results

In order to test the accuracy of the network, a random subset of 25 coded memes was left out of the training set to test generalization to new items using a fully trained set of connection weights. This is a test of the network's predictive power and generalization to new memes. The resulting output activation values were compared to the expected target values. If the meme's output activation on the "successful" output node was greater than the output activation on the "unsuccessful" output node then the classification was considered accurate. If the meme's output activation on the "unsuccessful" output node was greater than the output activation on the "successful" output node then the classification was considered inaccurate. The network achieved 80% prediction accuracy, or 20% higher than chance. Specifically, the network was able to accurately predict a successful meme to be successful with 73% accuracy, and was able to accurately predict an unsuccessful meme to be unsuccessful with 90% accuracy.

Regression analysis. In addition, a binary logistic regression was performed. The target values (successful or unsuccessful) were considered the dependent variable and each input node was considered an independent variable. Because all data is binary, binary logistic regression is appropriate for analyzing the factors that contribute to predicted success of a meme. The overall logistic regression model was statistically significant, $X^2(14) = 48.893$, $p < .0005$. The model explained 22.3% (Nagelkerke R^2) of the variance in meme success and correctly classified 54.1% of the successful memes as successful and 80.6% of the unsuccessful memes as unsuccessful. Overall the binary logistic regression model had a prediction accuracy of 67.4%. Three predictor variables were statistically significant. First, shorter memes were significant ($p < .005$), and 2.802 times more likely to contribute to success. Memes that contained a swear word were .177 times less likely to be successful than unsuccessful ($p < .05$), a small but significant contribution.

Finally, template memes were 2.223 times more likely to be successful than unsuccessful (p <.05).

Discussion

The results of the current study demonstrate the utility of using linguistic information as a means of predicting successful transmission of a meme. These preliminary results warrant more in depth analyses, particularly a sensitivity analysis that would detail which features contribute most to the outcome. Clearly, linguistic information contributes a rich source of information that could be used in models that incorporate multiple domains of information (user-level, visual feature, social structure, etc.). Some of the features in the network may have contributed more or less to the prediction of success in the network, and as with other neural networks it is difficult to see what is driving these results. However, comparing the network's results with a binary logistic regression helped to provide some insight. Meme length, whether or not a meme is a template meme, and the presence or absence of swear words within the meme contributed significantly to predicting success in the logistic model. However, the logistic model did not have prediction accuracy as high as the neural network model, pointing to the potential contribution of other variables that on their own are not predictive in a regression, but in an interactive context like a neural network, or perhaps other non-linear models, have some predictive power.

The neural network model presented here has several major limitations. The first limitation is the operationalized definition of success. Google search results offer a quick rough grained estimate for overall meme usage, but searching for specific phrases can still sometimes include inaccurate search results. Without extensive and computationally expensive web-crawlers, determining meme context from Google search results may be extremely difficult. Memes that can be used in multiple domains can be considered "flexible memes", a quality that is likely related to overall meme fitness. Another limitation to the current study is the input set and test set are relatively small. Many studies attempting to predict meme success have access to millions of memes, albeit with a broader operational definition. If the success of textual memes is largely dependent on the average person's ability to remember them, then many more cognitive variables can and should be included.

Conclusion

The ability to detect and track memes and predicting their success is essential in order to improve our understanding cultural evolution. Observing textual memes in particular offers unique insights into the evolution of language. Social media provides a petri dish environment for rapid meme generation and mutation. The current study categorized meme content based on 12 features grounded on cognitive theories of memory, emotion, and working memory limitations. This experiment helped support the idea that meme content should be considered when attempting to predict meme success. Future studies on meme prediction should benefit from a more robust operational definition of success. This can likely be achieved by limiting the scope from a global internet search to a specific social network. If a feed-forward backpropagation neural network can achieve relative success in predicting meme popularity, then a more robust network that takes into account working memory limitations should provide more accurate results.

This model demonstrates that it is not only possible to predict overall success of a meme at greater than chance levels, but also argues for there being important parameters at the level of what other models typically neglect: whether or not the node transmits the information further. Other models of meme transmission typically only take into account the change of the meme over time (evolution), the rates of transmission (viral) or the number of connections (small world networks). By incorporating cognitive processes into models that also include information about the network at large, greater levels of prediction could be achieved in future instantiations of meme transmission models.

References

Adamic, L. A., Lento, T. M., Adar, E., & Ng, P. C. (2014). Information Evolution in Social Networks. arXiv:1402.6792 [physics]. Retrieved from http://arxiv.org/abs/1402.6792

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behavior, 14(6), 575–589. doi:10.1016/S0022-5371(75)80045-4

Barrat, A., Barthelemy, M., & Vespignani, A. (2008). Dynamical processes on complex networks. Cambridge, UK; New York: Cambridge University Press.

Blackmore, S. (1998). Imitation and the definition of a meme. Journal of Memetics-Evolutionary Models of Information Transmission, 2(11), 159–170.

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. Trends in Neurosciences, 21(7), 294–299. doi:10.1016/S0166-2236(97)01214-9

Daley, D. J., & Kendall, D. G. (1964). Epidemics and Rumours. Nature, 204(4963), 1118–1118. doi:10.1038/2041118a0

Dawkins, R. (1989). The selfish gene. Oxford; New York: Oxford University Press.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. Cognition & Emotion, 19(3), 313–332. doi:10.1080/02699930441000238

Heintz, C., & Claidière, N. (2015). Current Darwinism in Social Science. In Handbook of Evolutionary Thinking in the Sciences (pp. 781-807). Springer Netherlands.

Hogg, T., & Lerman, K. (2012). Social dynamics of Digg. EPJ Data Science, 1(1). doi:10.1140/epjds5

Huette, S., & Anderson, S. (2012). Negation without symbols: The importance of recurrence and context in linguistic negation. Journal of Integrative Neuroscience, 11, 295-312.

Ienco, D., Bonchi, F., & Castillo, C. (2010). The Meme Ranking Problem: Maximizing Microblogging Virality (pp. 328–335). IEEE. doi:10.1109/ICDMW.2010.127

Kensinger, E. A. (2007). Negative Emotion Enhances Memory Accuracy: Behavioral and Neuroimaging Evidence. Current Directions in Psychological Science, 16(4), 213–218. doi:10.1111/j.1467-8721.2007.00506.x

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009, June). Meme-tracking and the dynamics of the news cycle. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 497-506).

Pennebaker, J. W., & Francis, M. E. (1999). Linguistic inquiry and word count (LIWC). Mahwah, NJ: Erlbaum.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: mapping the spread of astroturf in microblog streams (p. 249). ACM Press. doi:10.1145/1963192.1963301

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In Proceedings of the 20th international conference on World wide web (pp. 695-704). ACM. doi:10.1145/1963405.1963503

Shifman, L. (2012). An anatomy of a YouTube meme. New Media & Society, 14(2), 187–203. doi:10.1177/1461444811412160

Simmons, M. P., Adamic, L. A., & Adar, E. (2011). Memes online: Extracted, subtracted, injected, and recollected. In In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology, 29(1), 24–54. doi:10.1177/0261927X09351676

Tsur, O., & Rappoport, A. (2012). What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities (p. 643). ACM Press. doi:10.1145/2124295.2124320

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5), 1256–1271. doi:10.1037/0278-7393.25.5.1256

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063

Weng, L., Menczer, F., & Ahn, Y.-Y. (2014). Predicting Successful Memes using Network and Community Structure. arXiv:1403.6199 [physics]. Retrieved from http://arxiv.org/abs/1403.6199

Xie, L., Natsev, A., Kender, J. R., Hill, M., & Smith, J. R. (2011). Visual memes in social media: tracking real-world news in YouTube videos (p. 53). In Proceedings of the 19th ACM international conference on Multimedia (pp. 53-62). ACM Press. doi:10.1145/2072298.2072307