

Computational Creative Advertisements

Wei Sun

IBM Research
Yorktown Heights, New York
sunw@us.ibm.com

Anshul Sheopuri

IBM
Armonk, New York
sheopuri@us.ibm.com

Ying Li

IBM
Armonk, New York
yingli@us.ibm.com

Thales Teixeira

Harvard Business School
Boston, Massachusetts
thales@hbs.edu

ABSTRACT

Making successful video advertisements has long been considered a combination of art and business acumen. In this work, we propose a system to assist human designers to produce more effective advertisements with predictable outcomes. We formalize this concept with a dynamic Bayesian network (DBN), where we represent the knowledge base with data collected from large-scale field experiments in a novel setting. Face and eye tracking which continuously measures viewers emotional responses and viewing interest on 169 television advertisements for 2334 participants, along with moment-to-moment branding activities in the advertisements are used to estimate the model. The resulting DBN represents relationships across advertisement content, viewers emotional responses, as well as effectiveness metrics such as ad avoidance, sharing and influence on purchase. Conditioned on the specified requirement on the ad, a human designer can draw high scoring samples from the DBN, which represent the optimized sequences of branding activities and entertainment content.

KEYWORDS

Computational creativity; advertisement design; artificial intelligence; dynamic Bayesian network; inference

ACM Reference Format:

Wei Sun, Ying Li, Anshul Sheopuri, and Thales Teixeira. 2018. Computational Creative Advertisements. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3184558.3191549>

1 INTRODUCTION

Advertising is an expensive investment in a brand. On average, a 30-second commercial in broadcast prime time for the 2017-18 season costs \$134,009, a 6% increase from the previous season. The most expensive program in broadcast during the period is NBC's "Sunday Night Football", with an average price for 30 seconds of ad time tagged at \$699,602. CBS and NBC broadcasts of "Thursday

Night Football" are tied for No. 2, averaging about \$550,000 for 30 seconds¹.

The primary objective of this work is to develop a system which assists human designers to produce quantifiably effective advertisements. One of the key challenges is to teach a machine to answer "What constitutes a good ad?" Part of the challenge is that in both theory and practice, there are still ongoing debates on principles behind successful ads. Despite a considerable amount of research has been conducted to demystify the process, it remains inherently complex and error-prone, because of the tremendous variety of creative options and the large number of external factors compounding cognitive challenges for human designers.

To address this challenge, we start by constructing a knowledge base of video advertisements and corresponding viewer responses. For each advertisement, we record second-by-second branding activities, such as placement and size of logo, frequency and duration of logo appearance and the presence of brand in audio track etc. To collect viewer responses, in addition to self-reported metrics, we also utilize eye tracking and facial expressions to provide moment-to-moment measurements on exactly where viewers are focused on screen at any given moment and discrete emotions such as joy and surprise. The tracking data which provides a real-time lens into subconscious consumer reactions is more effective than traditional methods like surveys, as it is a passive, unconscious expression of sentiment and engagement. Utilizing data collected from large-scale field experiments with 2334 participants and 169 television advertisements, we construct a knowledge base on ad designs via a dynamic Bayesian network (DBN), which simultaneously represents a multitude of relationship across advertisement content, viewers' emotional responses, as well as effectiveness metrics such as ad avoidance, sharing, influence on purchase. With this knowledge base, recommendations on ad designs are obtained via inference. Conditioned on the specified requirements on an ad (e.g., product category, target audience, prominent display of brand logo in the end of the ad), a human designer can draw high scoring samples from the DBN, which represent the optimized sequences of branding activities and entertainment content that the system recommends. Figure 1 shows its overall architecture.

Developed in the 1980s, Bayesian networks are a mathematical formalism that can simultaneously represent probabilistic relationships between many variables in a system [24]. They have been

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191549>

¹<http://adage.com/article/news/tv-ad-pricing-chart/310429/>

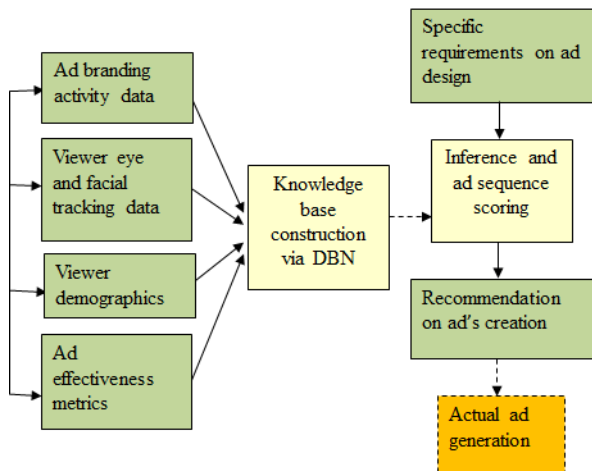


Figure 1: Overall architecture of the proposed system.

used in a wide variety of applications such as computational biology [11, 16], information retrieval [8, 18], computer vision [15, 23] and risk analysis [4, 31]. Bayesian networks allow both predictive and diagnostic inference, which is crucial for reasoning about a complex process such as advertisement designs.

The idea of using a knowledge-based system for advertisement design has been tossed around in the past [5, 12]. The major challenge in building such a system is creating the knowledge base. Prior work solely relies on input from experts to specify key variables and relationships to form a database of rules. As expert knowledge in this domain is often perceived as “soft”, qualitative and subjective, existing systems are unable to provide prescriptive recommendations. Moreover, they also have no mechanism for “creative thinking”, because recommendations are template-like rules based on what has been found to be effective in situations with the same set of characteristics. In this work, we take a different, data-driven approach in extracting knowledge via learning the relationships between content in advertisements and the corresponding viewer responses. Moreover, the system built on top of this knowledge base is able to assist human designers in thinking outside the box and expanding their exploration boundaries beyond what has existed.

There have been some efforts on automatic generation of videos in the multimedia research area. For instance, [33] proposes to automatically generate personalized music sports videos by first detecting semantic events from sports video content then stitching these events together and mixing in music. Similar ideas are also presented in [14], where the authors propose to automatically compose a music video from personal home videos. Nevertheless, prior work mainly focuses on the extraction of relevant snippets from given videos and the composition of these snippets with music, with an objective to represent existing content in a more pleasing and enjoyable way. Producing video ads, which involves many more creative decisions, is far more challenging. Besides generating entertaining content, advertisements also intend to achieve multiple objectives such as influencing sales, building brand awareness and mass communication.

On a high level, the theme of this work is on computational creativity, a subfield of Artificial Intelligence research, which is concerned with machine systems that produce novel and high-quality artifacts and ideas [7]. The study of these generative systems has been explored in a plethora of domains including music [22, 34], jokes [2], stories [25], poems [17], games [19], and culinary recipes [32]. Our work adds onto this body of literature in the domain of advertising, where human and computer creativity forms a two-way relationship, where hypothesis and rules learnt from human creativity is used to realize computational creativity, which in turn generates new hypothesis and rules to empower creativity in humans.

The paper is outlined as follows: In Section 2 we describe the experiments and the data. We give a brief overview of Bayesian network and describe the learning of our network in Section 3. In Section 4 we discuss how to use the network for recommendations. Finally, Section 5 concludes the paper and discusses ongoing research.

2 EXPERIMENTS AND DATA COLLECTION

2.1 Data Overview

The data for this research were collected from four large-scale controlled experiments on ad-viewer responses [26–29]. A sample of 169 television commercials were selected. They consist of a wide variety of product categories (e.g., beverages, consumer package goods, telecom), featuring both well-known brands (e.g., Coca-cola, T-Mobile) and obscure names (e.g., Sobe, Radio 538). 2334 consumers who are ad viewers were paid for participation. Each participant was exposed to a subset of commercials in a randomized order. On average, each advertisement was watched by 66 participants.

A summary of data collected from the experiments is shown in Table 1. It includes viewer demographics data (e.g., gender, age, familiarity with the brand and the ad) and static ad information such as product category and ad length. For each ad, we also collected its dynamic advertising content, which includes branding activity data and entertainment content data. Branding activity which refers to the appearances of brand identity (name, logo, trademark) in the advertisements is used to promote awareness and persuasiveness. Ads are analyzed to identify the brand’s presence, size, position, duration across frames and visual complexity of the scene.

In order to effectively and systematically measure viewers’ responses, we collect both facial tracking and eye tracking data. We use facial tracking and emotion analytics to measure viewer’s moment-to-moment emotional states in terms of joy and surprise. These measurements on emotions are used as proxies for entertainment content in ads, based on the notion that entertainment induces the emotional gratifications that viewers experience [27]. To collect this data, facial expression footage from each participant is collected via a webcam. This approach unobtrusively captures expressed feelings without significantly impacting viewer’s decisions to stop viewing. The video images serve as input to the emotion analytics software, which works by fitting a virtual facemask to the video image of the face (see Figure 2). This facemask adjusts to the form of the face (eyes, nose, face, and mouth) so as to capture line segments that form the basis of the facial coding. If a participant smiles or laughs, for example, a few deviations in line segment will increase,

Table 1: Data description.

Feature name	Description	Mean	Range
Gender	Participant's gender (binary, 1 = male)	0.475	0 - 1
Age	Participant's age in years (discrete)	36.50	17 - 62
Ad Familiarity	Participant's prior familiarity with the ad (binary, 1 = familiar)	0.336	0 - 1
Brand Familiarity	Participant's prior familiarity with the brand (binary, 1 = familiar)	0.726	0 - 1
Entertain	Participant's perception of brand as entertaining (Ordinal, 1 = not at all, 5 = very)	3.117	1 - 5
Extravert	Participant's self-reported degree of extraversion (Ordinal, 1 - 9 point scale)	5.020	1 - 9
Other	Participant's self-reported degree of other-directedness (Ordinal, 1 - 11 point scale)	5.213	1 - 11
Ad Length	Length of ad in seconds (continuous)	39.64	15 - 102
Logo Presence	Whether the brand is currently appearing in the ad (binary, 1 = present)	0.303	0 - 1
Audio Brand	Presence of brand name in audio track of ad (binary, 1 = present)	0.353	0 - 1
Competing	Whether the brand is visually separated from the background (binary, 1 = separated)	0.204	0 - 1
Position	Whether the brand is centrally located in the video (binary, 1 = central)	0.118	0 - 1
Size	Size of the brand as a percentage of the pixels onscreen (continuous)	0.035	0 - 0.615
Cardinality	The number of times the brand has appeared up to this point in time (discrete)	1.220	0 - 39
SC	Complexity of the image onscreen measured by a compression algorithm (continuous)	134.06	1.90 - 662
FTJoy	Face tracking of the emotion joy (continuous)	0.064	0 - 1
FTSurprise	Face tracking of the emotion surprise (continuous)	0.014	0 - 1
FTEyebrow	Face tracking of eyebrow movement (continuous)	0.272	.013 - .919
IAD	The distance between a participant's point of focus compared to the group (continuous)	-0.257	-0.71 - 20.18
Watch time	The percentage of an ad that the participant has finished watching (continuous)	0.759	0 - 1
Zap	Whether the participant skipped the ad at a particular time (binary, 1 = skipped)	0.344	0 - 1
Forward	Whether the participant forwarded the link of the ad (binary, 1 = forwarded)	0.273	0 - 1
NumForward	The number of people to whom the participant forwarded the ad (discrete)	0.663	1 - 6
Seedview	The number of views garnered from the link forwarded by the participant	5.327	0 - 73
Click	Whether the participant clicked on a link on the brand (binary, 1 = clicked)	0.118	0 - 1
Purchase	Participant's purchase intent after viewing the ad (ordinal, 1 = not at all, 10 = high)	4.863	1 - 10

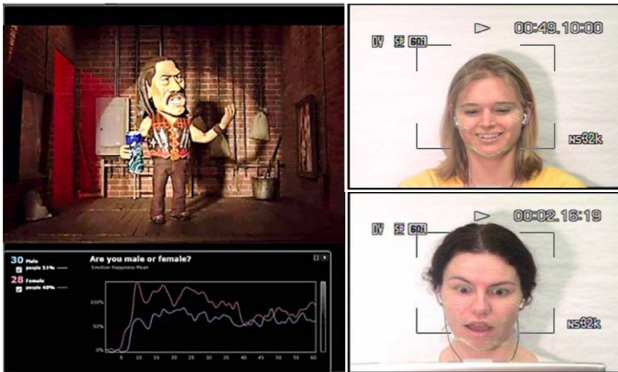


Figure 2: Facial expression footage from each participant is collected via a webcam. The light colored lines over the face represent the virtual mask used to measure facial point deviations in the emotion analytics algorithms. They hone in on points on a face, such as the corners of eyes or eyebrows to detect texture variations that occur when people laugh, frown, or smirk.

such as the one linking both corners of the lips, while others will decrease, such as the one linking corners of the lips to the cheekbone. The output of the system is a probability measure associated

with the intensity of an expression. Studies have shown that the algorithm can identify emotions such as joy with accuracy over 90% [20, 21]. Infrared corneal-reflection eye-tracking technology is used to record the focal positions of the viewers' right eye, in Cartesian X and Y coordinate. The output pinpoints a viewer's focal attention in a commercial, and determines where someone is looking and for how long.

To accurately measure ad effectiveness in a realistic experimental setup, participants are allowed to skip past an advertisement, in contrast to previous studies with forced exposure where viewers cannot avoid the ads [1, 9]. The percentage of an ad that a viewer chooses to watch provides a more accurate measure on ad effectiveness in terms of viewer engagement. We also conduct questionnaires to obtain self-reported metrics such as the viewer's purchase intent, whether he/she clicks on a web link to find out more about the brand after watching an ad, and ad virality based on action of sharing the ad in the social network context. The combination of both tracking and self-reported methods provide us with a more complete and complementary understanding of audience response to advertising.

2.2 Data Processing

As advertisements have different lengths, we vary the sampling rate so as to make them comparable. For each ad, we divide it into K segments with equal time interval, and aggregate the time varying features such as branding activity and emotion tracking

data, for each segment. In the work, we have analyzed the data for $K = 3$ and 10. When $K = 3$, the dynamic features represent the aggregated values in the beginning, the middle and the end of an ad respectively. When $K = 10$, for a 30-second ad, we evaluate it at 3-second intervals.

Our data consists of discrete (e.g., whether a viewer has skipped an ad and whether he has shared the ad with his friends) and continuous features (e.g., joy and surprise intensities), many of which do not exhibit normality. Thus, we focus on constructing a discrete Bayesian network, where both the global and the local distributions are multinomial, and the latter are represented as conditional probability tables. We discretize the continuous variables using Hartemink's method which preserves pairwise dependencies [13], as compared to marginal discretization methods such as using quantiles or fixed-length intervals. We then jointly discretize all the variables into a small number of intervals by iteratively collapsing the intervals defined by their quantiles. The process stops when each variable has 3 levels, that is, low, medium and high intensities.

3 KNOWLEDGE REPRESENTATION FOR AD DESIGNS

3.1 Overview on Bayesian Networks

Bayesian networks are a class of graphical models that allow a concise representation of the probabilistic dependencies among a given set of random variables $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ as a directed acyclic graph G with nodes \mathcal{V} and arcs \mathcal{A} . Each node $V_i \in \mathcal{V}$ corresponds to a random variable X_i . The Markov property of Bayesian networks specifies that every random variable X_i directly depends only on its parents Π_{X_i} . As a result, the factorization of the joint probability distribution $P_{\mathbf{X}}$ can be expressed as the following in case of discrete random variables,

$$P_{\mathbf{X}}(\mathbf{X}) = \prod_{i=1}^n P_{X_i}(X_i | \Pi_{X_i}).$$

Dynamic Bayesian networks (DBNs) extend the fundamental ideas behind static Bayesian networks to model associations arising from the temporal dynamics. Each variable in a DBN is represented by several nodes across time steps. We focus on multivariate time series which are sequences of multivariate random variables measured at successive time steps, and utilize a DBN to describe a discrete-time stochastic process $\mathbf{X} = \{X_i(t); i = 1, \dots, n; t = 1, \dots, T\}$, that takes values in \mathbb{R}^n at every time step t .

3.2 DBN Representation

Various representations have been proposed to handle temporal dependence in DBNs, often with the goal to reduce the complexity introduced by the time dimension. One common approach of modeling DBNs is to consider a uniform structure, where the network contains identical structure at every time step with identical temporal conditional dependencies (arcs) between time steps. Thus, the resulting DBN is built over a static BN, by replicating them for each time slice and adding arcs between states across two consecutive time slices, if they are temporally dependent. Such networks enable easy separation of DBNs in time and efficient inference algorithm when unrolling the network [10]. Unfortunately, the time-invariant nature of this representation imposes an unrealistic assumption

for our application, because there is plenty of empirical evidence which suggests that viewers' responses are highly dependent on time. For example, the tendency to skip an ad is naturally higher in the middle of the ad than in the beginning or at the very end.

Another popular approach of modeling DBNs is to represent the temporal dependence relationships by a vector auto-regressive (VAR) process, where the value at a given time, $X_i(t)$, is given as a linear combination of those at earlier time points. Specifically, for a process of order p , the variables observed at any time $t \geq p$ are assumed to satisfy

$$\mathbf{X}(t) = \mathbf{A}_1 \mathbf{X}(t-1) + \mathbf{A}_2 \mathbf{X}(t-2) + \dots + \mathbf{A}_p \mathbf{X}(t-p) + \mathbf{B} + \epsilon(t),$$

where nonzero coefficients in the matrix \mathbf{A}_k defines the arc set for lag k , i.e., if the element a_{ij} ($i \neq j$) is nonzero, then the network includes an arc from $X_i(t-k)$ to $X_j(t)$. Matrix \mathbf{B} represents the baseline measurement for each variable and $\epsilon(t)$ is a noise vector, with zero mean. However, with this representation, arcs between state variables within a time slice are disallowed. As our data consists of both the ad data and the viewer responses which are highly likely to be dependent at every time step, the VAR representation is inappropriate for our application.

In our work, we do not impose limiting network structures a priori. Each variable at every time step $X_i(t)$ is included into the system as an additional random variable. We explicitly encode the temporal relationships by enforcing constraints during structural learning, which state that dynamic nodes (e.g., $X_i(t)$) could influence all other nodes in current time step as well as those in the later steps (i.e., $X_j(t+k)$ for all j and $k \geq 0$). However, the reverse arcs from the future steps to the earlier time steps are strictly disallowed.

While this method allows a great degree of flexibility in representing the underlying system, one limitation is that the network size and complexity increase as time steps increase. For example, when we increase the time steps (i.e., total segment of each ad) from 3 to 10, the total number of nodes in the DBN rises from 53 to 137.

3.3 Learning

We learn the structure of the DBN via hill-climbing with random restarts [3], which is a greedy algorithm that explores the search space starting from an empty network and adding, deleting, or reversing one arc at a time until the score can no longer be improved. Besides enforcing time dependence as discussed earlier, prior knowledge on the data is integrated into structural learning to prune the existence of equivalent classes of networks that are indistinguishable from a probabilistic point of view. For instance, we set constraints which state that viewer attention and emotion states cannot influence ads, that is, the arcs from eye tracking data and emotion tracking data to ad branding activity are prohibited.

Since the result on learning can be noisy, we use nonparametric bootstrap and learn one network from each bootstrap sample. We repeat structure learning for 500 times and the network structures were averaged to reduce the impact of locally optimal but globally suboptimal networks. Arcs are considered significant if they appear in at least 95% of the networks and in the direction that appears most frequently. The procedure of model averaging is known to result in a better predictive performance than choosing a single, high-scoring network [6]. Once that the network structure is specified,

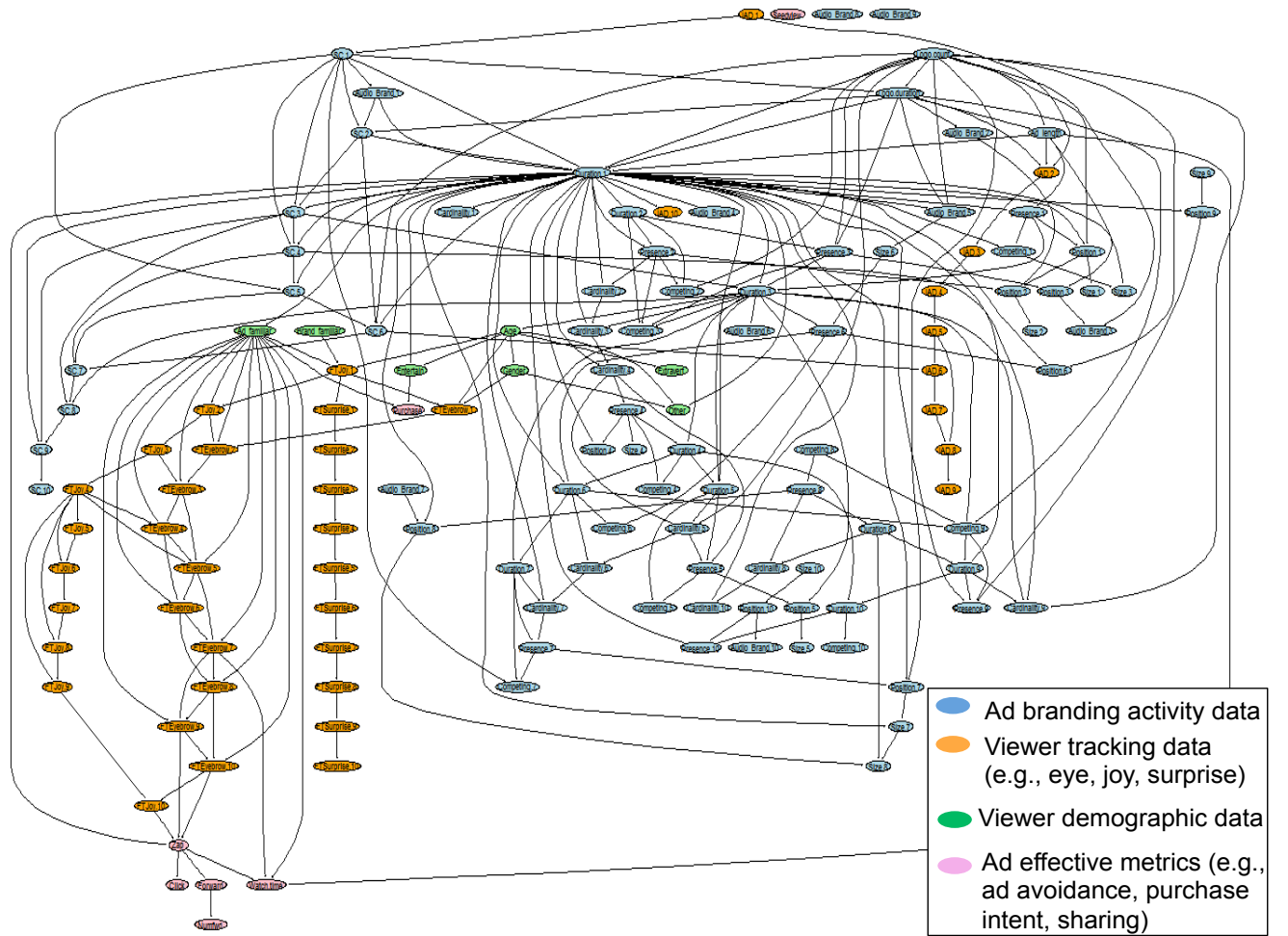


Figure 3: A trained dynamic Bayesian network on ad-viewer response with 137 nodes.

we fit the parameters of the local distributions, which take the form of conditional probability tables, via posterior Bayesian estimation.

The resulting network for $K = 10$ (i.e., each ad consists of 10 equally spaced segments) is shown in Figure 3. The learnt network reveals arcs representing conditional dependence among variables within a same time slice, as well as ones that represent temporal dependence among variables between consecutive time steps and across several time steps. The network consists of with 137 nodes, which are shaded with different colors to indicate the data type. It shows two large clusters of nodes mainly in blue and orange, which represent ad branding activities and viewer tracking data respectively.

There are several interesting observations revealed by the network. For instance, within the cluster of blue nodes on branding activity, nodes corresponding to viewer eye tracking data (orange nodes labeled IAD, see Table 1 for its definition) are closely linked to logo duration and frequency, suggesting the impact of branding activity on viewers' attention. However, the ad effectiveness metrics (pink nodes) are directly linked to the viewer tracking data nodes

in terms of joy and eyebrow movement, which is in agreement with empirical observations on the use of entertainment content on viewer engagement [28, 29]. Among the viewer tracking data, we observe that most nodes representing surprise are isolated from the rest of the network. The use of surprise in advertisements occurs much less frequently compared to the use of joy. In addition, surprise is a more fleeting emotion than joy. It is plausible that at the current level of data aggregation, the effect of surprise has been washed out and its influence would be more accurately captured when we refine the model by increasing the total segments of ads.

The learnt network also reveals that static features from self-reported surveys are useful in deciphering how to make effective advertisements. For instance, purchase intent has two incoming arcs from nodes which indicate whether a viewer is familiar with an ad and whether his or her perception of the brand is entertaining. The relationship between purchase intent and ad familiarity is relatively well studied, in the sense that repetitive advertising exposure reinforces brand loyalty and product preference [30]. The latter relationship is less obvious, implying that having an “entertaining”

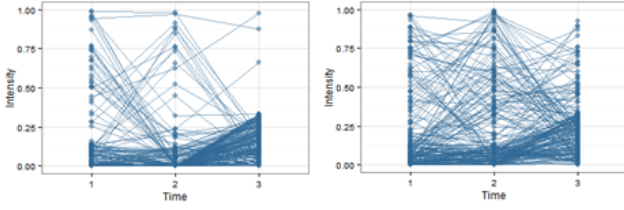


Figure 4: 200 sequences of joyful content sampled from the posterior distributions, conditioned on having low engagement (left) and high engagement (right) respectively.

brand image is related to its ads’ attractiveness and persuasiveness for increasing purchases.

4 RECOMMENDATIONS WITH DBN

Given a Bayesian network with structure G and parameters Θ , the effects of a new piece of evidence E on the distribution of X using the knowledge encoded in the model, can be investigated by two types of queries, namely, maximum a posteriori and conditional probability queries.

Maximum a posteriori queries are concerned with finding the configuration q^* that has the highest posterior probability for a given subset of variables Q , i.e.,

$$q^* = \arg \max_q P(Q = q \mid E, G, \Theta).$$

Thus, conditioned on the specified requirement on the ad (e.g., target audience, product category, last 5 seconds must show logo), a human designer can determine the optimal combinations of ad content based on the existing ad data encoded in the DBN.

Conditional probability queries focus on the distribution of a subset of variables Q given evidence E on another set of variables in X , which is computed as the posterior probability, $P(Q \mid E, G, \Theta)$. Conditioned on the specified requirement on the ad as evidence, the samples drawn from the posterior distribution represent “design proposals” consisting of sequences of branding activities and entertainment content. The versatility in the inference enables the system to be “creative”, as some sequences do not exist in the data.

To illustrate the ad generation process, we condition on one of the viewer engagement metrics, which is measured by the percentage of an ad that a viewer has finished watching. Specifically, conditioned on two scenarios, i.e., having low engagement where viewers skip early in the ad and high engagement, we generate 200 samples from the corresponding posterior distribution on the entertainment content and branding activities. Every sample is a multi-dimensional time series, showing how features evolve with time. Take joyful content as an example, intensities correspond to the beginning, the middle and the end segments of the ads are shown in Figure 4 for both scenarios. By comparing the two sets of trajectories, we observe that the high engagement posterior has generated more “up-and-down” sequences than those with low engagement. This coincides with empirical observation on building an “emotional roller coaster” to improve viewer engagement, by briefly terminating viewers’ positive feelings and then quickly restoring them [29].

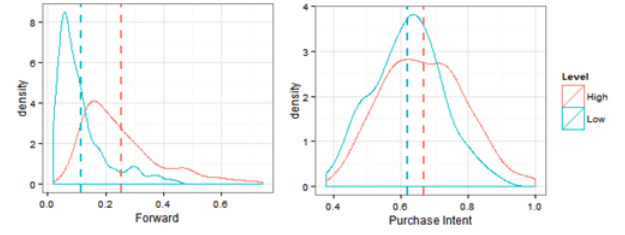


Figure 5: Predicted effectiveness of ads in terms of likelihood of forwarding and purchase intent, corresponding to samples conditioned on having high and low engagement respectively.

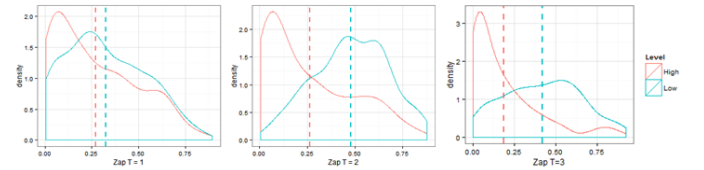


Figure 6: Predicted ad avoidance likelihood in the beginning, the middle and the end of ads, conditioned on having high and low engagement.

Once samples are generated from the above process, we can evaluate their effectiveness by applying predictive models on various engagement metrics. For instance, a random utility framework with time-varying parameters which associates an ad’s content and viewer information to the likelihood of sharing the ad and purchase intent has been constructed in [26, 27]. Figure 5 depicts the distributions of the two metrics for the 200 ad samples generated above. The dashed line represents the median of the respective metrics. It shows that on average, the samples conditioned on high viewer engagement are more likely to lead to sharing and higher purchase intent. We also observe that the improvement in forwarding is more significant than that in purchase intent. The learnt DBN in Figure 3 shows that the node representing purchase intent is directly linked to nodes “Ad Familiarity” and “Entertain”. In this experiment, we are only varying joyful content while keeping all other parameters at their median values as observed in the input data. The improvement in purchase intent is likely to be higher if we restrict to a population of viewers who are familiar with the ads and consider the brand entertaining.

Similarly, we can also predict the likelihood that a viewer will skip an ad. Figure 6 shows the distribution on ad avoidance with respect to different time segments of 200 ad samples. It shows that on average, the zapping probability is lower for samples generated by the high engagement posterior. In particular, we see that the reduction in ad avoidance is relatively small in the beginning of an ad (typically viewers are less likely to skip an ad in the first few seconds regardless of its content) and becomes more prominent in the middle and near the end of an ad. Thus, with such a system, a human designer can generate quantifiably effective advertisements by selecting design proposals with high scoring samples.



Figure 7: Sample output of audiences' emotional responses for a stock footage clip

5 ONGOING WORK AND CONCLUSION

This paper reports our initial work on the cognitive advertisement design system, which focuses on knowledge representation and recommendations. Our ongoing research directions are three folds: 1) construct more refined models so as to offer recommendations at a finer granularity; 2) create ads based on the model recommendations; 3) validate the model via user studies.

To create ads by applying recommendations from this system, we are currently working with nViso (<http://www.nviso.ch/>), a Swiss company specialized in 3D facial imaging and affective computing measurement analytics, to measure viewers' emotional responses to a few hundreds of stock footage clips. Figure 7 shows a screenshot of viewer responses for one of the video clips, where various emotions including sadness, surprise, disgust, anger, fear and happiness, are captured on a second-by-second basis. The change in emotion measurements and the trend line representing the overall involvement for each emotion are shown. For this particular video, it appears to be a highly joyful clip, indicated by the high level of happiness measurements.

Once all clips have been labelled with the moment-to-moment joy and surprise intensities, we will use them to form the video repository for ad creation. More specifically, for a given ad recommendation, appropriate clips from the repository with specified entertainment content will be selected and then stitched together. Next, logo and product appearances will be inserted into the content according to the recommended branding activities. Finally, music, voiceover as well as transition effects, will be mixed into the ad to make it more aurally and visually pleasing.

Lastly, we are planning to conduct user studies to assess the effectiveness of ads generated by this system and will utilize viewers' responses to iteratively enrich the knowledge base and improve the recommendation engine.

REFERENCES

- [1] William E Baker, Heather Honea, and Cristel A Russell. 2004. Do not wait to reveal the brand name: The effect of brand-name placement on television advertising effectiveness. *Journal of Advertising* 33, 3 (2004), 77–85.
- [2] Kim Binsted and Graeme Ritchie. 1997. Computational rules for generating punning riddles. *HUMOR-International Journal of Humor Research* 10, 1 (1997), 25–76.
- [3] Remco Ronaldus Bouckaert. 1995. *Bayesian belief networks: from construction to inference*. Universiteit Utrecht, Faculteit Wiskunde en Informatica.
- [4] Hichem Boudali and Joanne Bechta Dugan. 2005. A discrete-time Bayesian network reliability modeling and analysis framework. *Reliability Engineering & System Safety* 87, 3 (2005), 337–349.
- [5] Raymond R Burke, Arvind Rangaswamy, Jerry Wind, and Jehoshua Eliashberg. 1990. A knowledge-based system for advertising design. *Marketing Science* 9, 3 (1990), 212–229.
- [6] Gerda Claeskens and Nils Lid Hjort. 2008. *Model selection and model averaging*. Vol. 330. Cambridge University Press Cambridge.
- [7] Simon Colton, Geraint A Wiggins, et al. 2012. Computational creativity: The final frontier? In *ECAI*, Vol. 12, 21–26.
- [8] Luis M de Campos, Juan M Fernández-Luna, and Juan F Huete. 2004. Bayesian networks and information retrieval: an introduction to the special issue. *Information processing & management* 40, 5 (2004), 727–733.
- [9] Russell H Fazio, Paul M Herr, and Martha C Powell. 1992. On the development and strength of category-brand associations in memory: The case of mystery ads. *Journal of Consumer Psychology* 1, 1 (1992), 1–13.
- [10] Jeff Forbes, Timothy Huang, Keiji Kanazawa, and Stuart Russell. 1995. The batmobile: Towards a bayesian automated taxi. In *IJCAI*, Vol. 95, 1878–1885.
- [11] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. 2000. Using Bayesian networks to analyze expression data. *Journal of computational biology* 7, 3-4 (2000), 601–620.
- [12] Jacob Goldenberg, David Mazursky, and Sorin Solomon. 1999. The fundamental templates of quality ads. *Marketing science* 18, 3 (1999), 333–351.
- [13] Alexander J Hartemink, David K Gifford, Tommi Jaakkola, Richard A Young, et al. 2001. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks.. In *Pacific symposium on biocomputing*, Vol. 6. World Scientific, 266.
- [14] X. Hua, L. Lu, and H. Zhang. 2004. Automatic music video generation based on temporal pattern analysis. *ACM Multimedia* (2004), 472–475.
- [15] Michael Isard. 2003. PAMPAS: Real-valued graphical models for computer vision. In *Proceedings of Computer Vision and Pattern Recognition. IEEE Computer Society Conference*, Vol. 1. IEEE, I–613.
- [16] Xia Jiang, Richard E Neapolitan, M Michael Barmada, and Shyam Visweswaran. 2011. Learning genetic epistasis using Bayesian network scoring criteria. *BMC bioinformatics* 12, 1 (2011), 89.
- [17] Anna Kantosalo, Jukka M Toivanen, Ping Xiao, and Hannu Toivonen. 2014. From Isolation to Involvement: Adapting Machine Creativity Software to Support Human-Computer Co-Creation.. In *ICCC*, 1–7.
- [18] Christos L Koumenides and Nigel R Shadbolt. 2012. Combining link and content-based information in a Bayesian inference model for entity search. In *Proceedings of the 1st Joint International Workshop on Entity-Oriented and Semantic Search*. ACM, 3.
- [19] Antonios Liapis, Georgios N Yannakakis, and Julian Togelius. 2014. Computational Game Creativity.. In *ICCC*. Citeseer, 46–53.
- [20] Daniel McDuff, Rana El Kaliouby, Thibaud Senechal, Mohammed Amr, Jeffrey F Cohn, and Rosalind Picard. 2013. Affective-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected" In-the-Wild". In *Proceedings of Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 881–888.
- [21] Daniel McDuff, RE Kaliouby, and Rosalind W Picard. 2012. Crowdsourcing facial responses to online videos. *Affective Computing, IEEE Transactions on* 3, 4 (2012), 456–468.
- [22] Kristine P Monteith. 2012. *Automatic Generation of Music for Inducing Emotive and Physiological Responses*. Ph.D. Dissertation. Brigham Young University.
- [23] Vladimir Pavlović, James M Rehg, Tat-Jen Cham, and Kevin P Murphy. 1999. A dynamic bayesian network approach to figure tracking using learned dynamic models. In *Proceedings of the Seventh IEEE International Conference*, Vol. 1. IEEE, 94–101.
- [24] Judea Pearl. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. (1988).
- [25] Federico Peinado and Pablo Gervás. 2006. Evaluation of automatic generation of basic stories. *New Generation Computing* 24, 3 (2006), 289–302.
- [26] Thales Teixeira. 2012. The new science of viral ads. *Harvard Business Review* 90, 3 (2012), 25–27.
- [27] Thales Teixeira, Rosalind Picard, and Rana el Kaliouby. 2014. Why, when, and how much to entertain consumers in advertisements? A web-based facial tracking field study. *Marketing Science* 33, 6 (2014), 809–827.
- [28] Thales Teixeira, Michel Wedel, and Rik Pieters. 2010. Moment-to-moment optimal branding in TV commercials: Preventing avoidance by pulsing. *Marketing Science* 29, 5 (2010), 783–804.
- [29] Thales Teixeira, Michel Wedel, and Rik Pieters. 2012. Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research* 49, 2 (2012), 144–159.

- [30] Gerard J Tellis. 1988. Advertising exposure, loyalty, and brand purchase: A two-stage model of choice. *Journal of marketing research* (1988), 134–144.
- [31] Paolo Trucco, Enrico Cagno, Fabrizio Ruggeri, and Ottavio Grande. 2008. A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation. *Reliability Engineering & System Safety* 93, 6 (2008), 845–856.
- [32] Lav R Varshney, Florian Pinel, Kush R Varshney, Angela Schörgendorfer, and Yi-Min Chee. 2013. Cognition as a part of computational creativity. In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2013 12th IEEE International Conference on*. IEEE, 36–43.
- [33] J. Wang, C. Xu, E. Chng, L. Duan, K. Wan, and Q. Tian. 2005. Automatic generation of personalized music sports video. *ACM Multimedia* (2005), 735–744.
- [34] Geraint Wiggins, George Papadopoulos, Somnuk Phon-Amnuaisuk, and Andrew Tuson. 1998. Evolutionary methods for musical composition. *Dai Research Paper* (1998).