Video Audience Analysis using Bayesian Networks and Face Demographics

Ítalo de P. Oliveira, Carlos D. O. Interaminense, Eanes T. Pereira, and Herman M. Gomes
Federal University of Campina Grande (UFCG)
58429-900 Campina Grande, PB, Brazil
Email: {italooliveira, carlosdaniel}@copin.ufcg.edu.br, {eanes, hmg}@dsc.ufcg.edu.br

***Abstract*—In this paper, we propose an approach to study and to model audience attention to videos within digital signage scenarios. An experimental setup was conceived to simultaneously display videos of various categories and to capture videos from the audience and surrounding environment. Face analysis via a deep neural network is performed to estimate gender and age groups. In the proposed approach, a Bayesian Network is built to model possible relationships between the audience's age, gender and face size (which is indicative of the distance to the display) and the video content types. A publicly available video dataset of 152 videos was created for displaying purposes. An experimental evaluation indicated varying degrees of attention to different videos, depending on age and gender. The area under the ROC curve of the built Bayesian Network was 0.82. The proposed approach allows to better understand the possible relationships between audience demographics and video contents, which may, in turn, be useful for displaying the most appropriate content to a particular audience, help with the automatic insertion of ads (based on audience categories), among other applications.**

***Index Terms*—Digital Signage, Computer Vision, Audience Analysis, Bayesian Networks, Face Analysis.**

I. Introduction

Digital Signage exposes dynamic visual information such as urgent news, weather forecast, entertainment, and other subjects with the aim to capture the viewer's attention. Digital signage has been used in different applications, such as in promotion and launching of products and services, conveying information to customers in a certain environment and even on critical occasions. In Japan's earthquake of April 2009, displayed information on large screens was useful for emergency evacuation [1].

To generate better user satisfaction, new technologies have been developed. For example, the Intel's Audience Impression Metric Suite (AIM) ® indirectly acquires information about the viewers, through monitoring cameras. Studies like [2] make use of the information collected by cameras for the automatic selection of ads. Such approaches make digital signage more attractive by dynamically adapting the transmitted contents [3].

Some efforts in the standardization of digital signage have been started, such as the one proposed by the Point of Purchase Advertising International [4], which is a global non-profit trade association dedicated to improving retail environments, having more than two thousand affiliated companies worldwide, such as Cisco® and LG®. Their standardization defines technologies to be employed as a way to organize Digital Signage broadcasts and suggests quality standards to be followed (e.g. audio and video coding). The ultimate goal of these efforts is to improve the Return on Investment (ROI) of companies with studies that optimize the audience.

In this sense, monitoring cameras can be used to viewers categorization and thus, use information about the demographic profile for customization of the exhibited content. Viewers' analysis may also serve as input for billing strategies for advertising space, so that companies willing to allocate a time slot for the transmission of their ads would have access to accurate reports on the target audience. Finally, systems with that feature could be used to automatically adapt to the stream of ads. By learning the preferences of demographic groups and by detecting the presence of a prevailing group, the goal is to broadcast the ad which holds the highest interest of those viewers or to broadcast the company's ads which target that particular viewer profile. Several companies have invested in advertising and digital signage, but there is still lack of methods to recognize customers' profiles and/or associate gender, age group and other features to video content.

In this paper, a methodology for acquiring knowledge and making inferences about viewers during a digital signage experience is proposed. The main objective was to develop a method to allow the analysis of possible relations between viewers and exhibited videos, while taking into consideration the video genre, viewer's gender and age group. The proposed method aims to improve the interactivity and effectiveness of digital signage systems by analyzing the demographic profile of viewers to provide insights and valuable data to advertising companies. The proposed approach was fully implemented and validated using data acquired in a real-world case study.

Viewer's information such as genre and age are estimated via a facial analysis component. Viewers' attention was analyzed as a result of different content types stimulus. Age groups have been defined based on sociological studies. As a result of Bayesian Network inference, it was verified that the viewers attention to different content types displayed varied depending on the age group and gender. A database of 152 public videos, available on the internet, have been manually categorized and made available.

II. Related work

The research of [2] presented a framework for selecting ads using the AIM Suite, which was installed in five different locations at Intel® campuses. The collected data was analyzed according to two different approaches for predicting viewers' preferences: (i) using demographic information (i.e., gender and age); and (ii) using the context (i.e., video content, date, device location, and other information). Finally, decision trees were used to predict the videos based on demographic and contextual information. The prediction model based on demographic information showed to be up to 4% more efficient than the context-based model.

Another relevant research was conducted by [5], who developed a system for real-time audience analysis. The application consists of capturing videos from digital signage viewers, using a camera positioned below the TV that broadcasts the ads. A case study, inside a clothing shop, indicated that the gender of the viewer had a significant impact on time of stay and attention, with men being more receptive than women. Results also showed that the age group had an influence on the attention time. Finally, the relation with the content type transmitted was analyzed and it was found that video ads retained 43% more attention time of the viewers when compared to image ads.

In the research of [3], a monocular camera and a Kinect[1] device were used to capture data from viewers. For distances greater than four meters, the monocular camera obtained lower Root Mean Square Error (RMSE) in face detection.

The research carried out by [6] showed the gender had a low correlation with the age group of viewers and with type of exhibited content, while gender and age group highly correlated with attention time of the viewers. The evidence obtained in these works was used to support the work of [7], which proposed an ad distribution model which adapted the ads selection in an interactive way, depending on the local audience.

In the research of [8] a camera positioned above a TV set was used to capture a video from the viewers, from which faces, eyes and mouth areas were detected. The facial poses were then estimated in order to find out which faces were facing the TV. In the experiment, different types of ads were exhibited, such as: international news, Taiwan news, climate news and others. The types of the ads were correlated to the number of viewers and to the average visualization time per viewer. Different exhibited content attracted viewers' attention in different ways. For instance, Taiwan news attracted a larger number of viewers when compared to International news, but the viewers stayed for a shorter period of time. Differently from [3], in the research of [8], demographic viewers analysis (eg, gender and age considerations) was not performed.

An approach for analyzing the effects of adds on people, based on reported opinions, was presented by [9]. In that study, students' preferences from two different colleges were analyzed. The ads were placed with different variations in background color, text positioning, and font type. Experimental results showed ad color was not a relevant factor for women in choosing the ad of their preference, while men preferred gray ads. Thus, implicit factors in ads may draw viewers attention in a differentiated way, as also discussed in [10], who studied how men and women have brain areas stimulated differently.

[1]Motion sensor developed by Microsoft that captures color and depth information.

The research conducted by [11] used questionnaires to analyze viewer's satisfaction during train travels, focusing on the excellence of the passengers experience of a public transport system in Holland. In the experiment, different content types were transmitted, such as: travel information, news, curiosities about the route and trip destination, climatic conditions, sports highlights and others. The 835 passengers were divided into two groups, those receiving and those not receiving advertisements, in order to test the effects of digital signage on emotional experience, based on a questionnaire which was subsequently applied. Trip satisfaction was higher for those who received the ads. Moreover, when the content of the transmitted ad was related to the trip, a more favorable opinion was observed. In conclusion, the transmitted content influenced the viewers and, in such a closed system, the application of an automatic demographic recognition system could be used to select the best ads based on the opinions contained in the forms.

In the research of [12], a system for book recommendation was developed. The system estimates the gender and age group of an individual based on their silhouette. Even with low accuracy, 77.5% for the gender classifier and 37.5% for the age group using a Support Vector Machine (SVM) as classifier, experimental results showed 25% of people who walked through the shop appeared to be interested in acquiring the item which was being advertised. Those results demonstrated the effectiveness of digital signage systems and the importance of demographic recognition.

In the sales context, the research of [13] evidenced digital signage impact on retail store sales. In that research, after the digital signage insertion, the sales had a direct relationship with the size of the store. Sales increased in the large commercial stores, such as hypermarkets. In supermarkets, there was no change in sales when including digital signage, while in smaller stores (such as convenience stores), there was a negative impact on sales. This fact reinforces the need for experimental studies that analyze viewers' attention, especially in small spaces so that digital signage favors sales.

III. Materials

The following datasets were adopted or produced: FDDB [14], used to train and validate the face detector; MORPH [15], used to train and validate the gender and age classifier; a catalog of videos used for displaying purposes; a dataset of videos captured from the viewers, synchronized with each of the video exhibitions.

The videos shown to viewers came from open air TV channels or from public internet channels such as Youtube®. The video dataset consists of 152 videos, which were displayed interchangeably, on alternate days and at the same time of day. Shown videos have been manually tagged in one of the following categories: police (14 videos), economy (19 videos), health (20 videos), politics (25 videos), entertainment (34 videos) and news about technology (40 videos). A Microsoft Excel spreadsheet containing the URLs and genres of this dataset is available for download at https://github.com/italoPontes/Sibgrapi2019.

A LG 42-inch Full HD TV set was used for video display. A Logitech® C920 model camera with full HD resolution, H.264 encoding and capture rate of 30 frames per second (fps) was used to capture videos from the viewers. A computer was used to send videos to the TV HDMI interface and to capture videos from the audience using the USB webcamera. TV set was placed at about 1.8m from the ground using a metalic tower stand, which also housed the computer and webcamera.

Pixel Intensity Comparisons Organized in Decision Trees (PICO) method proposed by [16] was used for face detection. Open Source Computer Vision library (OpenCV) was also employed for providing image processing and computer vision functionalities for video capture and analysis. We choose PICO for face detector for three main reasons: (1) it is an open-source solution; (2) it was written in C/C++, and all other modules of our systems were developed in C/C++ too, so providing a more efficient integration between the parts; and (3) it presents good results in the Face Detection Data Set and Benchmark (FDDB) [14].

The CAFFE Deep Learning Framework, developed by [17], was used for training the age estimation and gender classification neural networks. GoogLeNet, developed by [18], was the base architecture, with small adaptations, adopted to perform the training tasks. The age and gender classifiers have been defined according to the results of a previous investigation [19]. The WEKA environment, developed by [20] provided the algorithms for the construction of Bayesian Networks.

IV. Methods

Figure 1 presents a diagram of the proposed method. The main goal was to discover possible relationships between the variables of the digital signage problem (e.g. the gender and age of the viewers, the distance to the TV and the content type displayed). The method receives two inputs: (i) the videos to be displayed, and (ii) the faces found in the audience. Videos captured from the audience are processed through the following steps: (i) face detection; (ii) gender classification and (iii) age classification. In this article, a face detector already trained [16] was adopted. Upon detection, an increase in the face region is applied so that the face received by the classifier has proportions compatible with the input of the age and gender neural network models.

After face detection, the gender and age classifiers, previously trained, perform the gender and age estimation in parallel. All information extracted from the processed video is stored: the genre and age estimated by the classifier, the displayed video content, as well as secondary informations, such as face coordinates and the processed frame number, which was used to make considerations about the viewing moment of the viewer. Next step is to train a Bayesian Network that provides a Knowledge Base on the *Capture Video Base*. Finally, using the *Knowledge Base* it is possible to perform the analyzes on the viewers behavior with the *Audience Report* generation.

An approach commonly used by sociologists [21], [22] and [23] was employed to discriminate age groups. The approach use the term *generations* to specify time periods which have some historical aspect in common, thereby allowing individuals born in that period to identify with the collective, i.e. other individuals who were also born in the same period. Additionally, the resolution of the National Health Agency (ANS) [24] was adopted, which defines age ranges for pricing purposes in private health insurance. The following generations and birth year intervals, respectively, are considered in this paper: **Z**, $\geq$ 1987; **Y**, 1986-1977; **X**, 1976-1967; and Baby Boomers - **BB**, $\leq$ 1966. The age and gender classifier used in this study were originally proposed by [19] and further trained with additional samples in this research. The obtained F1-Score was 72% when considering the above four age groups, and 92% for gender classification using 5 fold cross-validation.

The proposed method receives acquired videos from the audience for off-line processing, no real-time processing is required. During data acquisition experiment, there is a synchronous display of video contents and video capture from the audience. Captured video is stored with displayed video identification (ID) for association purposes in the off-line analysis. Once the video dataset from the viewers is captured, the demographic analysis process begins. The process consists in applying the face detector (PICO) to all frames of all the captured videos. After that, the age and gender classifiers are applied. Figure 2 illustrates the described method.

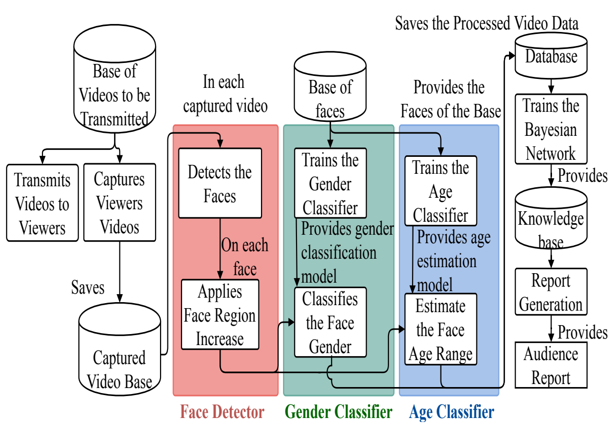


Figure 1: Diagram of the method proposed in this article.

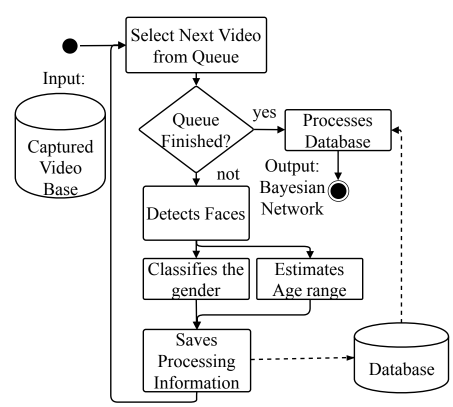


Figure 2: State diagram for face demographics analysis.

A. Statistical analysis

The statistical analysis used Bayesian Networks to study the dependence between the variables collected in the experiment, which are: the type of content displayed, the genre and the age of the viewer, as well their estimated distance to the TV display. The Bayes' Theorem is the mathematical foundation of the Bayesian Network. The structure of the Bayesian Network was automatically infered via the K2 [25] learning algorithm.

1) Bayes Theorem: Terminology and mathematical notations related to Bayes Theorem are presented next. Sample space Ω is defined as a possible result set of a random experiment, that is, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, where, ω_i are the elements of this set with size n. An A event in Ω is an Ω subset, that is, $A \in \Omega$. These possible outcomes are events, where the probability distribution P is a function that calculates the probability of an ocurring event from the sample space, providing a real number between 0 and 1, $P : P(\Omega) \to [0, 1]$ [26].

The sample space for the content type of the video displayed in this study can be defined as: Ω = [*health, politics, technology news, entertainment, economy, police*]. The a priori probability that a captured frame comes from a $health$ video type is 20.2%, so P($health$) = 20,2%. Considering that the probability of two events occurring, A and B for example, is specified as $P(A \cap B) = P(B|A)P(A)$, then the conditioned probability of event A occurs, since event B has already occurred, is defined by: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. Obviously, for this equation be valid, $P(B)$ must be strictly positive. Using the aforementioned concepts and based on the variables acquired from the captured videos, it is possible to compute the conditioned probabilities to perform the attention analysis of the viewers. A Bayesian Network was created to both represent the structure of the data collected in this case study and to make inferences, which will be discussed in the following subsection.

2) Bayesian Networks: Bayesian networks are diagrams that represent the dependence between variables and organize the data into a graph oriented between *cause* and *effect*, where the variable that knows the probabilities a priori is called *cause* and the *effect* is the variable that calculates the conditional probability. According to [27], a Bayesian Network is composed of a Directed Acyclic Graph (DAG) and a set of probability tables for discrete valued variables. The DAG contains the topology (also called structure) of the Bayesian Network, consisting of a set of nodes and arcs, where nodes correspond to variables and arcs represent the links between nodes. An arc connecting A to B indicates that there is a dependency of A in B; Then, the A node is the parent of B. For being an acyclic graph, if there is an arc from A to B and another arc from B to C, then there can not be an arc from C to A. A probability table represents the probability distribution conditioned to the parents of A. For a root node, that is, a parentless node, the probability distribution of this node represents its a priori probability distribution.

Arrow direction is always from the *cause* variable to the *effect* variable, or from the parent variable to the child. Thus, it is possible to structure the factors involved in an experiment, such as the content of the video exhibited, the genre and the age of the viewer, to perform an objective analysis of these variables.

3) K2 Learning Algorithm: The K2 learning algorithm was used to generate the structure of the Bayesian Network. This algorithm generates different Bayesian network topology possibilities and calculates the probability of generating the input data set [25]. The algorithm K2 is formally defined as: CD, a set of input data with m observations (or cases); ER, the network structure; and $P(ER|CD)$, the probability of the conditioned network structure of the data set. The best structure of the network is sought based on the highest value obtained.

In the Bayesian Network construction, the K2 algorithm initially admits that nodes do not have parents, so the variables are added as parents of the variable a priori. At each step the value of $P(ER|CD)$ is calculated and checked the increase when a new variable is added as parent, seeking to maximize that value. When it is not possible to increase the value of $P(ER|CD)$, then the next variable is selected and the cycle is repeated until all variables are interconnected in the network.

Knowing that the problem has n variables, then there are $2^{\frac{n(n-1)}{2}}$ possible structures [28], which are determined based on the value defined by Equation 1, as discussed in [29].

$$Q_{K2}(ER, CD) = P(ER) \prod_{i=0}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(r_i - 1 + N_{ij})!} \prod_{k=1}^{r_i} N_{ijk}! \tag{1}$$

in which: $P(ER)$ is the a priori probability of the variable; X_i are the possible states of the variables contained in the CD; N_{ijk} is the total number of cases in CD; N_{ij} is the number of occurrences that have the parents set of X_i; n is the number of variables in X_i; q_i is the number of possible combinations for the parents of X_i; and n is the total number of possible values that the variable $X_i = (i = 1, \ldots, n)$ can assume.

Other metrics were tested and compared, such as K2 [25], Simulated annealing [29], Tree Augmented Naive Bayes (TAN) [30], and Naïve Bayes [31]. A cross-validation was performed on the data, with 10 folds and the K2 algorithm obtained the best performance, with Receiver Operating Characteristic (ROC) of 0.82. More details on these statistical measures can be seen in [32] and [33].

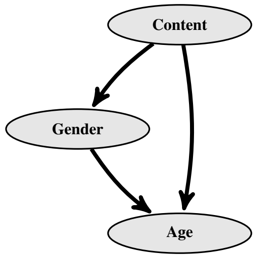
V. Experimental evaluations

Three evaluations performed in this paper are discussed in the following subsections: relation between exhibited content, gender and age; relation between exhibited content and age; and relation between age, gender and distance to display.

A. Exhibited video, gender and age

We started analysing the relationship between the exhibited content and the genre of the spectators. In Figure 3, it is possible to observe the Bayesian network diagram generated using the algorithm K2 [25] with its respective probability tables for each variable. Based on the structure of the diagram, we can notice the effect of the *cause* variable (in this case, the

Content	Probability a priori (%)
Health	20.2
Politics	19.6
Technology	17.9
Entertainment	15.2
Economy	14.3
Police	12.8



Conditional probability (%)	Gender F	M
Health	52.1	47.9
Politics	48.9	51.1
Technology	61.7	38.3
Entertainment	57.6	42.4
Economy	60.6	39.4
Police	68.4	31.6

Conditional Probability (%)	Gender: Female, Age groups Z	Y	X	BB	Male, Age groups Z	Y	X	BB
Health	37.5	25.2	33.7	3.5	45.0	26.6	24.1	4.3
Politics	34.1	25.9	37.7	2.3	42.1	27.6	27.1	3.2
Technology	29.1	33.4	35.4	2.2	42.1	23.1	31.4	3.4
Entertainment	30.1	32.3	35.2	2.5	47.7	30.3	18.4	3.6
Economy	26.9	35.6	36.1	1.5	50.2	21.2	27.2	1.5
Police	21.4	49.3	28.0	1.3	33.5	43.6	21.3	1.6

Figure 3: Bayesian network and probability tables generated based on the processing of all videos captured from viewers.

content of the video displayed) on the *effect* variables (gender and age of the viewer). Thus, it is observed that the viewers genre is influenced only by the content of the transmitted video, whereas the viewer age depends on both the genre and the content.

Based on the probability tables shown in Figure 3, the effect of the content displayed on viewers attention was analyzed. However, the profile of the viewer, represented, in this article, by the genre and age group which they belong to, may have a high a priori (or low) probability and affect the conditioned probabilities. That is, if there were many individuals of a certain gender (or age group), the probabilities conditioned to that profile will be greater (or less), depending on the a priori probability. Thus, to avoid erroneous deductions[2] an intra-class analysis was performed, the conditioned probabilities were compared to the same viewer profile by varying the exhibited content.

Viewers may vary their attention depending on the content transmitted. For example, when considering only the genre of the viewer, a great contrast is perceived in the attention given by the spectators to the police videos. These videos were preferred by women, accompanied by videos about technology and economy. On the other hand, the political videos were those of less interest. Since conditional probabilities are complementary, that is, the sum of men and women corresponds to 100%, the inverse logic is applicable to men. That is, the male audience's favorite theme was political, while the less interesting categories were: economy, technology and police, respectively.

Based on the probability tables shown in Figure 3, the following subsection is reserved to present an analysis on the probability tables, when considering only the age categories of the viewers.

B. Exhibited video and age categories

In this section, the relation between the types of exhibited videos and the age groups of the viewers is analysed. Figures 4 and 5 were generated to make the visualization of these relationships clearer. Figure 4 shows viewer representativeness for each video category. For example, half of the female viewers who watched the police videos are from generation Y and half of the male viewers who watched the videos on economy are from generation Z, being this generation the predominant of male audience. That chart makes it possible to demographically estimate the audience by age group. A more detailed analysis is presented next.

In Figure 5, the attention distribution is organized by age group. Thus, it is possible to better analyze the preferences and aversions which each generation presented. For example, women of the generation Y had a greater interest in police videos and less interest in health and political videos. There was also a low expressiveness of the spectators of the BB generation and the imbalance in the male group with a tendency to fall in the generation advance. In other words, the higher the age group, the lower the expressiveness. In addition, with the exception of police content, the other videos were more represented by the individuals of generation Z in the male group.

Table I contains the preference order of each generation per video category. The table served as a basis for argumentation about spectator attention analyzes, which are discussed next.

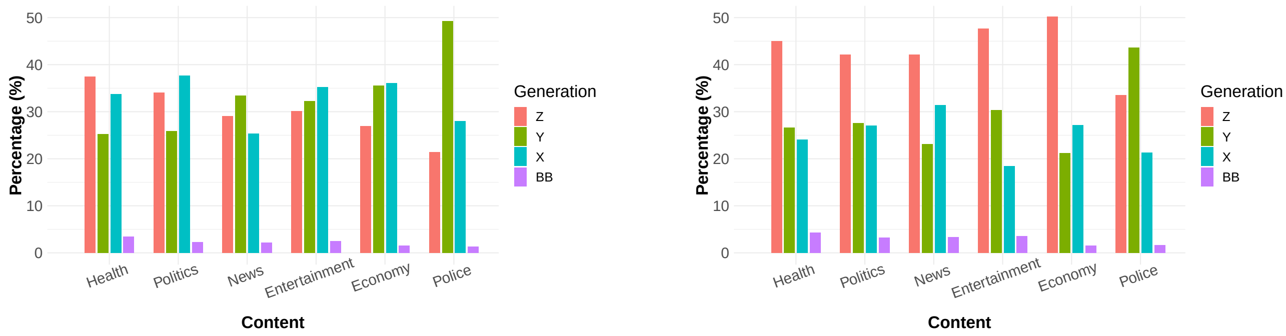
Table I: Order of viewers preferences based on each generation and its gender.

Video content	Gender: Female, Age Groups Z	Y	X	BB	Male, Age Groups Z	Y	X	BB
Health	1^{st}	6^{th}	5^{th}	1^{st}	3^{rd}	4^{th}	4^{th}	1^{st}
Politics	2^{nd}	5^{th}	1^{st}	3^{rd}	4^{th}	3^{rd}	3^{rd}	4^{th}
Technology	4^{th}	3^{rd}	3^{rd}	4^{th}	4^{th}	5^{th}	1^{st}	3^{rd}
Entertainment	3^{rd}	4^{th}	4^{th}	2^{nd}	2^{nd}	2^{nd}	6^{nd}	2^{nd}
Economy	5^{th}	2^{nd}	2^{nd}	5^{th}	1^{st}	6^{th}	2^{nd}	6^{th}
Police	6^{th}	1^{st}	6^{th}	6^{th}	5^{th}	1^{st}	5^{th}	5^{th}

Health was the preferred content for viewers of the generation BB (independent of the gender) and women of generation Z. While the women of the intermediate generations (X and Y) aroused little interest.

For politics videos, there is a contrast between genres in generation Z. Politics was among the most attractive content to women of generation Z, while the men of that same age group paid little attention. The political type was also attractive to women of generation X, while the women of generation Y contrast with the other generations because they paid little attention to this type of video.

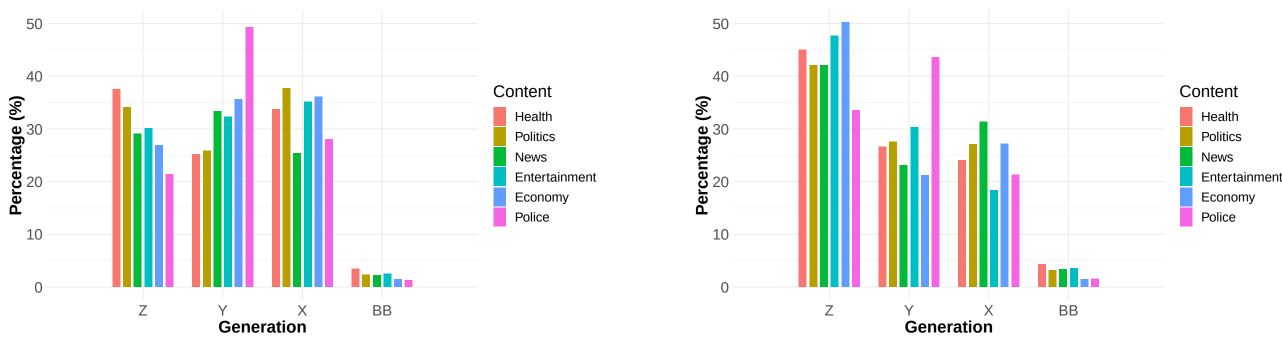
[2]For example, to conclude that individuals of the BB generation pay little attention because their conditional probabilities are low, when in fact there were few individuals with this profile and therefore their conditioned probabilities were lower than the other generations.



(a) Viewers of the female gender.

(b) Viewers of the male gender.

Figure 4: Relation between the exhibited content and the age group of the viewer, grouped by content.



(a) Viewers of the female gender.

(b) Viewers of the male gender.

Figure 5: Relation between content transmitted with the age group of the viewer grouped by generation.

Technology theme was indifferent to women. However, men of the younger generation (Z and Y) paid little attention, whereas this theme was one of the favorites for men of generation X.

For entertainment videos, the prevailing public was men, except those who are of generation X, since they paid little attention to the theme. In addition, this theme was one of those that drew more attention to women of the BB generation.

Economy videos caught the attention of viewers in a diversified way. In the younger generations (Z and Y), men paid attention in inverse proportion to women. That is, this theme was one of the preferences of the men of generation Z and of less interest for generation Y. O the other hand, women of generation Z aroused little interest, while women of generation Y had great interest. With the more advanced ages (X and BB), it is noticed that the men and women begin to converge in their preferences. That is, generation X viewers of both genres had little interest, while viewers of the BB generation paid more attention.

Police videos drew attention of both genders, with a high correlation. Regardless of the gender of the viewer, the variation of their attention was the same for the different age generations. The police theme was among the less interesting themes of all (as it was said above, it was invariant to the genre). However, a divergent behavior in generation Y was observed, police videos were of greater interest to this age group.

C. *Age, gender and distance to display*

In order to evaluate the relationship between the genre and the age of the viewer with the proximity to the screen, a new Bayesian network was built with one additional variable included - the distance from the screen. An heuristic rule was used to estimate the distance from the face to the monitor, which was based on the width of the viewer's face, i.e., the wider (in pixels) the face is, the closer the viewer is to the TV display. In this way, the face widths were grouped into four sets: (A) 201-270; (B) 151-200; (C) 101-150; and (D) 50-100 pixels wide. The faces of the group "A" are the closest ones, and "D", the further ones.

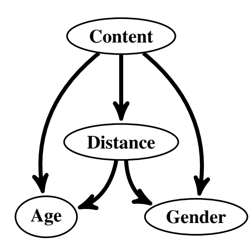


Figure 6: Structure of the Bayesian network.

Figure 6 represents the structure of the Bayesian network automatically constructed using the algorithm K2 [25] and implemented in the tool Weka [20], when considering the addition of the distance to the TV display. It can be observed that the gender and age range of the viewer depend on their distance from the monitor. The Table II shows the probabilities for all transmitted video content, in which the highest ones are highlighted in bold.

Table II: Relation between content and viewers proximity to the screen, considering gender and age.

Conditional Probability (%)	Age Groups				Genre	
	Z	Y	X	BB	F	M
P(Health ∣ d = A)	**51.9**	**40.6**	6.6	0.9	47.1	52.9
P(Health ∣ d = B)	37.3	38.6	22.9	1.2	**54.5**	45.5
P(Health ∣ d = C)	38.3	27.6	28.6	**5.6**	46.1	**53.9**
P(Health ∣ d = D)	42.1	25.1	**29.4**	3.4	54.2	45.8
P(Politics ∣ d = A)	**45.4**	17.7	34.6	2.3	**68.0**	32.0
P(Politics ∣ d = B)	31.4	25.2	**40.5**	2.8	51.0	49.0
P(Politics ∣ d = C)	36.6	**29.6**	30.2	**3.6**	48.8	**51.2**
P(Politics ∣ d = D)	38.8	26.0	32.7	2.5	48.9	51.1
P(Technology ∣ d = A)	8.49	**61.0**	30.2	0.3	**91.1**	8.9
P(Technology ∣ d = B)	15.9	53.3	29.5	1.3	83.1	16.9
P(Technology ∣ d = C)	30.4	33.9	32.3	**3.4**	57.8	**42.2**
P(Technology ∣ d = D)	**37.3**	25.4	**34.6**	2.7	59.3	40.7
P(Entertainment ∣ d = A)	9.2	**60.4**	30.3	0.1	**83.9**	16.1
P(Entertainment ∣ d = B)	12.7	54.8	**30.9**	1.6	83.3	16.7
P(Entertainment ∣ d = C)	**42.0**	29.2	25.1	**3.7**	52.9	**47.1**
P(Entertainment ∣ d = D)	38.5	30.1	28.5	2.9	56.7	43.3
P(Economy ∣ d = A)	8.6	**61.5**	29.8	0.1	**93.0**	7.0
P(Economy ∣ d = B)	10.8	58.5	29.5	1.1	86.6	13.4
P(Economy ∣ d = C)	39.4	25.8	32.9	**1.9**	57.1	42.9
P(Economy ∣ d = D)	**40.7**	24.6	**33.1**	1.5	55.6	**44.4**
P(Police ∣ d = A)	7.0	**62.9**	**30.1**	0.1	**92.5**	7.5
P(Police ∣ d = B)	9.2	60.2	29.6	1.1	88.1	11.9
P(Police ∣ d = C)	29.1	51.8	16.7	**2.4**	58.2	**41.8**
P(Police ∣ d = D)	**29.9**	41.8	27.0	1.4	63.9	36.1

Results analyzes indicated, by unanimity, women were closer to the TV than men. Moreover, Generation Z viewers appeared further away from the monitor, except when the videos on health and politics were broadcast, which were the women's favorite videos. Generation Y viewers positioned themselves closer to the TV, with the exception of the political and police videos transmission. In this case, sitting away from the monitor does not mean an aversion to the content transmitted, since, although spectators of this generation sat more distant from the TV during the political and police videos transmission, those themes were among the favorites of men of this age group. However, a certain aversion is perceived for the women of this age group in the political videos transmission. Generation X viewers positioned themselves in front of the monitor in a balanced way, that is, without predominant preference. Finally, generation BB viewers always sat away from the TV.

D. Relation to other researches

The global survey realized by the Nielsen about Lifestyles took place between February 23 and March 13, 2015, with the participation of more than 30,000 consumers interviewed in 60 countries across the globe. The survey is considered representative for Internet consumers, because it was made online and its sample has quotas based on age and sex for each country, based on their Internet users [34]. Respondents were asked about what their personal aspirations are, and they could choose to: Get in shape and healthy, Make money and others. Figure 7 shows the result of that research. When considering the age of the interviewees, it is observed that there is a growing tendency to get in shape and healthy, that is, the higher the age group, the greater the aspiration. The reverse is true of the need to make money, which has declined as the age group advances.

Figure 7: Interests comparison for different age categories [34].

The results presented in Figure 7 are directly related and corroborate with some of the results presented in this paper (shown in Figure 5). When considering two types of displayed videos, health, and economy, in isolation, we notice a growing trend in the viewers interest when considering health videos. This is noted when sorting the preferred videos for each of the generations, as shown in Table I. Thus, men and women paid more attention to health videos as the age groups increase. In addition, female viewers had a declining interest with advancing age. Finally, younger viewers, of the generation Z, were characterized as outliers in these conclusions, since they did not always obey the tendencies of the other age groups. One possible explanation is that the experiment carried out in this paper represented the specific reality of the region in which the data was captured.

Another study that reinforces the results obtained in the market research carried out by Nielsen®, is the research carried out by Ibope Inteligência® developed in partnership with Multivitamínico Centrum. That work carried out 613 interviews with men and women over 50 years old, of all economic classes, with proportional sample to the Brazilian population, by demographic region. The margin of error was 4 percentage points more or less [35]. That survey showed that 77% of Brazilians over 50 considered health as the main concern in life, corresponding to the main topic of interest observed in this article for individuals of the BB generation. Followed by financial stability, which corresponds to only 11% of respondents. The evidence from that previous research corroborates with the conclusions of our method.

VI. Conclusion and Future Work

In this paper, we proposed and experimentally evaluated a methodology for studying the possible relationships between face demographics (age, gender and distance to display inferred via computer vision methods) and different types of displayed video contents. The core of our methodology is a Bayesian Network model that is automatically learned from data generated by a data acquisition experiment.

When considering only genre and age range, the former is mostly influenced by the displayed content, whereas the later depends on both the genre and the content transmitted. In addition, when considering the distance that the viewer is located in relation to the TV display, gender and age are dependent on distance and that the age no longer is dependent on gender. These relations allowed us to generate different inference types about the viewers. For instance, women, in general, positioned themselves closer to the display when compared to men. Age category BB was positioned further away from the display. Women from Generation Z were positioned further away from the monitor in the displays of health and political videos, which were the subjects found less interesting to viewers in this category.

The proposed methodology may be extended to process the following additional information. Time and date may be used by a Recommender System that combine demographic data with collaborative filtering to show a better content in the specific time of day, or day of the week, etc. Similar to what was discussed in the survey of [36], which associated the traffic volume with the time of day. Emotion associated with facial expressions may be automatically recognized and used for customized content display. The impact in viewer emotion when exposed to different videos may also be studied. In addition, other associations may be held, as in [10], who noted that there are differences in the attention of men and women. In content-based recommendation, video information may be associated with the audience. Some examples: presence/absence of faces, amount of motion, among other measurements, in the transmitted video and its relation to viewers' attention. Finally, the use of face grouping techniques can be investigated, which might enable analyzing individual preferences.

Acknowledgment

This study was financed in part by the Brazilian National Council for Scientific and Technological Development (CNPq) and National Council for the Improvement of Higher Education (CAPES).

References

[1] C. Palestini, "Digital heritage and earthquake emergency," in *Digital Heritage International Congress*, 2013.

[2] T. Kinebuchi, H. Arai, I. Miyagawa, S. Ando, K. Kataoka, and H. Koike, "Image processing techniques for measuring advertising effectiveness of digital signage," *R&D on Digital Signage as an Advertising Medium*, 2009.

[3] R. Ravnik and F. Solina, "Interactive and audience adaptive digital signage using real-time computer vision," *International Journal of Advanced Robotic Systems*, 2013.

[4] P. D. S. POPAI, "Aplicações do digital signage," 2014.

[5] R. Ravnik and F. Solina, "Audience measurement of digital signage: Quantitative study in real-world environment using computer vision," *Interacting with Computers*, 2013.

[6] K. Mishima, T. Sakurada, and Y. Hagiwara, "Low-cost managed digital signage system with signage device using small-sized and low-cost information device," in *IEEE Annual Consumer Communications & Networking Conference*, Jan 2017.

[7] R. Ravnik and F. Solina, "Interactive and audience-adaptive information interfaces," *Electronic Lett. on Comp. Vision and Image Analysis*, 2014.

[8] K.-C. Yin, H.-C. Wang, D.-L. Yang, and J. Wu, "A study on the effectiveness of digital signage advertisement," in *International Symposium on Computer, Consumer and Control*, 2012.

[9] J. Grobelny and R. Michalski, "The role of background color, interletter spacing, and font size on preferences in the digital presentation of a product," *Computers in Human Behavior*, 2015.

[10] Q. Feng, Y. Zheng, X. Zhang, Y. Song, Y. jia Luo, Y. Li, and T. Talhelm, "Gender differences in visual reflexive attention shifting: Evidence from an ERP study," *Brain Research*, 2011.

[11] B. Moritz, "Content in the train: Taking a closer look at in-train screens: A study on what types of content, on in-train digital signage, lead to a more positive customer experience," 2016.

[12] M. Mikawa, S. Izumi, and K. Tanaka, "Book recommendation signage system using silhouette-based gait classification," in *Int. Conf. on Machine Learning and Applications and Workshops*, 2011.

[13] A. L. Roggeveen, J. Nordfalt, and D. Grewal, "Do Digital Displays Enhance Sales? Role of Retail Format and Message Content," *Journal of Retailing*, 2016.

[14] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," University of Massachusetts, Amherst, Tech. Rep., 2010.

[15] K. Ricanek and T. Tesafaye, "Morph: a longitudinal image database of normal adult age-progression," in *Int. Conf. on Automatic Face and Gesture Recognition*, 2006.

[16] N. Markus, M. Frljak, I. S. Pandzic, J. Ahlberg, and R. Forchheimer, "A method for object detection based on pixel intensity comparisons," *Computing Research Repository*, 2013.

[17] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *Computing Research Repository*, 2014.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Computing Research Repository*, 2014.

[19] I. de Pontes Oliveira, J. L. P. Medeiros, V. F. de Sousa, A. G. T. Júnior, E. T. Pereira, and H. M. Gomes, "A data augmentation methodology to improve age estimation using convolutional neural networks," in *SIBGRAPI Conf. on Graphics, Patterns and Images*, 2016.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *ACM SIGKDD Explorations Newsletter*, 2009.

[21] N. Howe and W. Strauss, *Millennials Rising: The Next Great Generation*. Knopf Doubleday Publishing Group, 2009.

[22] N. Fitch, *Sylvia Beach and the Lost Generation: A History of Literary Paris in the Twenties and Thirties*. Norton, 1985.

[23] J. D. Miller, "The generation x report: Active, balanced, and happy: These young americans are not bowling alone," *Longitudinal study of American youth*, 2011.

[24] ANS, "Resolução normativa, n° 63," 2003. [Online]. Available: http://www.ans.gov.br/component/legislacao/?view=legislacao&task=TextoLei&format=raw&id=NzQ4

[25] G. F. Cooper and E. Herskovits, "A bayesian method for the induction of probabilistic networks from data," *Machine Learning*, 1992.

[26] R. Daly, Q. Shen, and S. Aitken, "Review: Learning bayesian networks: Approaches and issues," *Knowledge Engineering Review*, 2011.

[27] N. Fenton and M. Neil, *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, 2012.

[28] F. L. Valentim, "Estudo e implementação de algoritmos de inferência e aprendizado em redes bayesianas," 2007.

[29] G. F. Cooper and E. Herskovits, "A bayesian method for constructing bayesian belief networks from databases," in *Conf. on Uncertainty in Artificial Intelligence*, 1991.

[30] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, 1997.

[31] G. H. John and P. Langley, "Estimating continuous distributions in bayesian classifiers," in *Conf. on Uncertainty in Artific. Intell.*, 1995.

[32] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, 2006.

[33] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, 2009.

[34] *Geração Global Estilos de Vida*, Nielsen, 2015. [Online]. Available: https://www.nielsen.com/br/pt/insights/reports/2015/Geracao-global-estilos-de-vida.html

[35] I. Inteligência, "Sinais da nutrição," 2015.

[36] K. Maeda, M. Nishi, T. Yoshida, K. Suzuki, and H. Inoue, "Digital signage with audience detection using tv broadcasting waves," in *Int. Symp. on Applications and the Internet*, 2010.