ReflectiveSigns: Digital Signs that Adapt to Audience Attention

Jörg Müller, Juliane Exeler, Markus Buzeck, and Antonio Krüger

University of Münster, Münster, Germany

Abstract. This paper presents ReflectiveSigns, i.e. digital signage (public electronic displays) that automatically learns the audience preferences for certain content in different contexts and presents content accordingly. Initially, content (videos, images and news) are presented in a random manner. Using cameras installed on the signs, the system observes the audience and detects if someone is watching the content (via face detection). The anonymous view time duration is then stored in a central database, together with date, time and sign location. When scheduling content, the signs calculate the expected view time for each content type depending on sign location and time using a Naive Bayes classifier. Content is then selected randomly, with the probability for each content weighted by the expected view time. The system has been deployed for two months on four digital signs in a university setting using semi-realistic content & content types. We present a first evaluation of this approach that concentrates on major effects and results from interviews with 15 users.

1 Introduction

As display prices drop and cheaper display technologies are invented, digital signs are beginning to be installed everywhere in public spaces, gradually complementing and replacing paper signs. This leads to a radical change in the urban landscape, as can already be observed in places such as Times Square, New York or Shibuya Crossing, Tokyo. On the positive side, a new generation of information access is enabled, as digital signs have many properties and affordances that differ from that of their traditional paper counterparts (e.g. cheap dynamic updates, context adaptivity and interactivity). On the negative side, such signs may lead to visual clutter and information overload for audiences. In addition there are ecological costs by installation & maintenance, power use and recycling. Signage and its content is known to work differently in different contexts. As Mitchell states: "Literary theorists sometimes speak of text as if it was disembodied, but of course it isn't; it always shows up attached to particular physical objects, in particular spatial contexts, and those contexts—like the contexts of speech—furnish essential components of the meaning." [7], p.9. Traditional signs have been adapted to their context for a long time. However for contexts other than location, this has proven laborious (e.g. manually displaying an "Open" sign when a shop is open). When digital signage is equipped with sensors, this

process of adapting to context can be automated. However its been a difficult task for media planners to estimate how content works in different contexts and manually schedule content accordingly. We propose to automate this process by just using the audience as a laboratory. By simply presenting content to an audience, using appropriate sensors it can be observed how the audience reacts to the content shown in a particular context. Machine learning mechanisms (e.g. the Naive Bayes classifier), can then be employed to automatically learn scheduling strategies from these experiences. Thus, the proposed process consists of two feedback loops: A scheduling loop and a learning loop (see Figure 1).

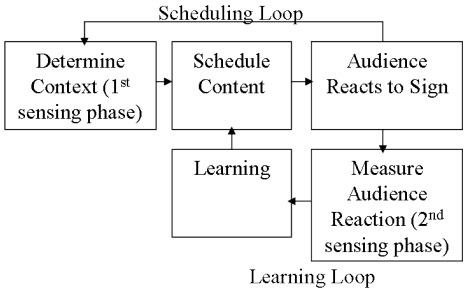

Fig. 1. Information flow in the proposed concept, consisting of the scheduling and learning loops.

2 Related Work

While bodies of research exist for public displays, ambient displays and context-aware systems, less study has been undertaken for context-aware public displays or learning public displays. GroupCast[5]is an example of public displays that identify the audience via a wireless badge and display content according to a pre-stored user profile. The Vision Kiosk[11] observes the audience with a camera and shows an animated face that looks in the audience's direction. The Interactive Public Ambient Display[12] observes the audience via a motion tracking system and adapts content to their distance and posture. BluScreen[10] is a system that identifies the present audience via Bluetooth-enabled mobile phones and employs auctions to select content the audience has not seen before. MobiDiC[8] is a system that distributes coupons to passers-by, measures advertising effectiveness with the coupons and optimizes content shown with auctions. Huang[3] presents an investigation of digital signage deployments *in the wild* and concludes that most digital signs receive relatively little attention. So, while some work has been done for showing different content in different situations, the task of automatically determining which content works best in various contexts has not been approached yet.

3 Adaptive Digital Signs through Sensing, Learning and Content Scheduling

The concept we propose for learning digital signage consists of two feedback loops (see Figure 1). First, the signs are context adaptive, i.e. they can automatically select content that fits the situation (the scheduling loop). Second, the signs are enabled to automatically learn how the audience reacts to different content in different situations (the learning loop). In the scheduling loop, the signs sense their context with any sensors that are available. Interesting context for selecting appropriate content could for example be the location, the time, weather, gender or age of the audience as well as audience profiles. Also, context that can be influenced by the audience, such as where they are looking, their distance from the sign or facial expression can also be used. From this information, the signs then decide which content to show in this situation. The audience hopefully reacts to it in some way (e.g. by watching, smiling, or interacting with it). Then, again, the context can be measured, and the loop begins anew. As long as users do not consciously understand this process, one would speak of incidental interaction[1], where the system reacts to the actions of the audience but no conscious interaction takes place. However, as soon as the audience understands this process, this loop potentially transforms to classical interaction. For example, the audience could notice that whenever they look sad, the signs would present some jokes. As soon as they start to look sad to make the sign present jokes, one would speak of classical interaction. One major difficulty in this scheduling loop is the creation of scheduling rules, e.g. the decision of which content to show in a certain context. The learning loop automates this process. After presenting particular content, the audience reaction to the content shown is measured with sensors. A learning mechanism can then be employed to learn which content provokes which audience reaction in a certain context.

In the presented prototype, the scheduling and learning mechanisms are designed to measure and maximize the time the audience looks at the signs. At the beginning of a content cycle, each sign determines autonomously which content to show. To determine context, instead of a set of active sensors, ReflectiveSigns currently only uses time and location. The sign uses the current time (in the categories night, morning, lunchtime, afternoon and evening) and its location, to retrieve the expected view time (an estimation of how long the audience is expected to look at the content) for each available content category. Then, a category is selected randomly, where the probability of each category to be selected is weighted by its expected view time relative to the other categories. Thus, content that attracts attention (in terms of time spent looking towards the signs) in a certain context is shown more often.

In the learning loop, currently the only sensor is the face detection that measures the audience's view time and then calculates the expected view time. Whenever some content is shown, the number of faces that are oriented towards the sign as well as the duration of time that these faces look at the sign are determined. The sum of these view times is then stored to a database. In order to be able to estimate the view time even with only few data, we used the Naive

Bayes approach to calculate the expected view time. The expected view time e is then calculated as $e = \sum_{i=1..\infty} p(v = i|l, t)i$, where v is the view time, l the location and t the current time. Under the assumption that location and time are conditionally independent, the Naive Bayes rule[6] is used to estimate $p(v = i|l, t) = \frac{p(l|v=i)p(t|v=i)}{p(l)p(t)}p(v = i)$. In practice, these parameters are simply estimated from historical data. As the system starts with no data, there is a problem of many probabilities being zero at the beginning. This problem is circumvented by applying the m-estimate[6] to individual probabilities. The effect of this is to give the system a set of values for a hot-start.

4 Implementation

Fig. 2. A user passing a ReflectiveSign, and exemplary content.

The ReflectiveSigns prototype consists of four digital signs installed at a university department with approximately 60 employees. One sign is located at an entrance (see Figure 2), one in a sofa corner, one in a hallway and one in a coffee kitchen. Before being used for this project, the signs were used for the university information system iDisplays[9]. We measure the audience reaction to content shown via cameras installed on top of the signs. The system uses a face detection algorithm[4] that detects faces when they are oriented towards the signs. For the system, we aimed at providing very different kinds of content. Besides the iDisplays system, which has been designed in a user-centered design process [9], we collected videos as well as text and still images that would be eye-catching, interesting and appeal to different people. Such content was somewhat unusual for a research institute (although many comics can be found attached to walls, and employees have used the displays to show sports channels during olympics). This is reflected in the interviews. Content includes video categories such as animated movies, short films showing people who are cooking, football matches or funny animal videos. There are seven non-video categories including landscape photography, three comic strips, textual news, buzzwords and the iDisplays as a mixed information category. The videos are cut into pieces of 20 seconds, still images and text rest on the display for 20 seconds each. Graphics and photographs are scaled to full screen size. All contents are presented

without audio. The scheduling algorithm does not decide on individual pieces of content but only on categories. Every day there will be a new piece of content for each category. As a consequence, the same items will be displayed multiple times per day. The system consists of four components: face detection software, a MySQL database, a Java-based content scheduler and a Java-based content player. We use the real time face detector from Fraunhofer IIS[4] to analyze the video stream. This software is able to detect multiple faces within the camera image during runtime. The data that is collected by the face detection running on the different machines is stored in the database. The content scheduler decides on a new category to be played every 20 seconds applying the described scheduling mechanism. Based on the category determined by the content scheduler the content player displays one item from this category.

5 Noise

One of the most important problems for ReflectiveSigns is the amount of noise due to the face detection. In order to estimate the error rate, we collected 8 hours of video for two different display locations and hand annotated all view times. In total, 87 views towards the signs occured in this time. The face detection recognized 27 of these views, totaling a recognition rate of 32%. Looking more closely at the nature of the errors however reveals that the face detection only missed views with a duration of under 1 sec. All views with longer duration were correctly recognized. The face detection however also recognized 304 false positives, mostly faces recognized for a single frame in objects like the fire extinguisher. We implemented a filter for false positives by ignoring regions where many faces appeared at exactly the same position. Although methods for coping with people moving to fast or being present but not looking at the sign exist (e.g. high speed cameras and eye detection like Xuuk[1]), the further reduction of error rates is considered future work.

6 Data Collected

The system operated for two months 24 hours per day, seven days a week. The first month served to learn audience attention patterns, the second month to collect data. The data from the second month is analyzed. In total, 38612 views towards the signs were detected. There were obvious effects for different attention towards the signs depending on location, time and content shown. The display installed in the sofa corner received the most attention (mean (μ)= .323s, standard deviation ($sigma$) =1.383s). All times are mean view times when content is shown for 20s. As often nobody is looking the mean values are quite small. However as so much data was collected, most differences are still significant. The sofa corner was followed by the coffee kitchen ($\mu = 0.312s, \sigma = 1.427s$), the hallway

[1] www.xuuk.com

($\mu = 0.229s, \sigma = 1.146$) and the entrance ($\mu = 0.146s, \sigma = 0.920s$). Not surprisingly, attention was highest during lunchtime ($\mu = 0.592s, \sigma = 3.876s$), followed by the afternoon ($\mu = 0.523s, \sigma = 2.728s$), the morning ($\mu = 0.307s, \sigma = 1.424s$) and the evening ($\mu = 0.178s, \sigma = 0.894s$). More interestingly, different content received different degrees of attention. For example, whenever animal videos were shown, they were viewed for 0.287 seconds on average, whereas iDisplays were only viewed for 0.206 seconds. Resulting from this difference, animal videos were shown 28115 times in total, while iDisplays were only presented 21091 times. When we conducted interviews (Section 7), we asked interviewees to rate each content with grades on a scale from 1 (very good) to 6 (bad). Surprisingly, we found no strong correlation between the average viewtime for certain content and these grades (Pearson correlation=-.089, Significance .83). Apparently, user preferences do not significantly influence their attention to display content.

We were interested in how big the influence of the content on audience attention is compared to the influences of location and time. Therefore, we conducted a three-factor analysis of variance on the view times (see Table 1). The influence of all three factors, location, time and content, are all significant (which is no surprise given the large sample of 291,947 content slots of 20s each). More interesting is the relative ordering of the factors (see column for Mean Sq.). This indicates that in our data location has the biggest impact on view times, followed by time. The influence of content on view times is considerably smaller. This is not surprising given that viewing the signs is usually not planned. Nobody will pass a sign only because certain content is being shown. Instead, mere presence of people is obviously only influenced by time and location. Often, people who pass the signs will look at them (or not) regardless of content shown. Presenting the right content only has the opportunity to make users look longer.

Table 1. ANOVA for location, hour, content, regarding view times. df are the degrees of freedom for that variable (e.g. 24 hours -1), Sum Sq is the summed square error for this variable, and mean sq is weighted by the degrees of freedom. These variables indicate how much variance in the view times can be explained by location, hour, and content, respectively. The last column shows that each of the variables has a significant influence on view times.

	df	Sum Sq	Mean Sq	F value	$Pr(>F)$
Location	4	56996	14249	8577.836	$< 2.2e^{-16} * **$
Hour	23	3894	169	101.928	$< 2.2e^{-16} * **$
Content	18	1178	65	39.385	$< 2.2^{e-16} * **$
Residuals	291947	484967	2		

7 Interviews

After running ReflectiveSigns for two months, we conducted semi-structured interviews with 15 employees and regular visitors of the institute (age 23-31,

$\mu = 27$). The interviews were partially transcribed and evaluated using Grounded Theory[2]. The system was understood with mixed feelings. Five users perceived the system as one that would show random videos and comics. While four users liked this content, three experienced an information overload: "[there is] only trash, always changing videos, simply totally crazy, everything colorful and fast. It drives me crazy."(User 9). Three users critized the system as it was apparently not "useful" (as opposed to the iDisplays shown on the same signs before). Asked what they believed the system was for, three users experienced the display as agressively attracting attention: "The display cries: Hello, here I am!"(U 11). Still, five users liked the (static) comics shown ("It's like a noticeboard, its nice to look there and laugh a bit."(U 13)), the iDisplays, and the surfing videos ("There were surfing videos, sport videos. That was an eye-catcher!"(U 11)). Four users considered the content, especially the videos, annoying. One user stated that he considered videos without sound useless: "For most of the videos you need sound. Because there is no sound, it's not interesting. Videos would be better with sound, but—when you don't like to see it, the sound would be horribly annoying."(U 8). Regarding the observation through the cameras, there were mixed feelings. Four of 15 users were heavily annoyed by the cameras, mainly because they did not understand their functionality: "[the cameras] annoy me because I don't know what happens with the videos taken. I don't want others to know the ways I walk [...]"(U 3). Four said the cameras are OK because they know who put them up. Seven did not care at all about the cameras. There was an interesting effect where incidential interaction (i.e. looking at the sign) turned into conscious interaction. Two users said they tried not to look at the sign when they don't like the content: "I think the content is stupid but then I look there and you know that"(U 5). Asked, what other content they would find interesting, four users mentioned news (esp. regarding the university and the city) and three mentioned sports videos (if short and self-contained). Two users said that they prefer useful content to entertainment: "I consider the display to be more for information, less for entertainment."(U 12). The chosen content apparently annoyed some of the audience, and some were annoyed by the cameras. However, most of them found some of the content interesting.

8 Conclusion

In this paper we have presented ReflectiveSigns, a digital signage system that automatically learns the audience attention for certain content (depending on the context), and presents content accordingly. The system was deployed for two months and evaluated through analysis of the logging data and interviews with users. Somewhat to our surprise, the analysis of variance of the view times indicates that the influence of the chosen content categories on view time is relatively small. Apparently, the right choice of sign location bears a much greater potential than the right choice of content. This is an important finding for the use of public displays in Pervasive Computing scenarios. However, the audience was very homogenous for all locations. If signs attract very different audiences at

different locations and times, the impact of the content may be much higher. It was also somewhat surprising to us that there seemed to be no strong correlation between view times and whether users liked the content. It may simply be the case, that users also look at content they don't like. Regarding the cameras, there are three kinds of users. Some disapproved of using cameras at all, some didn't care and for some it seemed OK as long as they trusted those who installed them. The approach presented opens many opportunities for future research. For example, it should be investigated whether a signage system that optimizes for audience attention indeed makes users look more and longer, and if so, how much. It should be further investigated how strongly attention towards different content in various contexts differs. Therefore, the noise in the system needs to be reduced, and more data collected. As such systems appear in urban spaces, visual spam and audience privacy are two major problems that need to be solved to not make them a harmful or annoying experience but beneficial for society.

References

1. A. Dix. Beyond intention - pushing boundaries with incidental interaction. In *Building Bridges: Interdisciplinary Context-Sensitive Computing*, pages 1–6, 2002.
2. B. G. Glaser and A. L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Pub, 2008 edition, 6 1967.
3. E. Huang, A. Koster, and J. Borchers. Overcoming assumptions and uncovering practices: When does the public really look at public displays? In *Proc. of Pervasive 2008*, pages 228–243, 2008.
4. C. Küblbeck and A. Ernst. Face detection and tracking in video sequences using the modified census transformation. *Image and Vision Computing*, 24 (6):564–572, 2006.
5. J. F. Mccarthy, T. J. Costa, and E. S. Liongosari. Unicast, outcast & groupcast: Three steps toward ubiquitous, peripheral displays. *Lecture Notes in Computer Science*, 2201:332–345, January 2001.
6. T. M. Mitchell. *Machine Learning*. McGraw-Hill Science/Engineering/Math, March 1997.
7. W. J. Mitchell. *Placing Words. Symbols, Space, and the City*. MIT Press, 2005.
8. J. Müller and A. Krüger. How much to bid in digital signage advertising auctions? In *Adjunct proceedings of Pervasive 2007*.
9. J. Müller, O. Paczkowski, and A. Krüger. Situated public news and reminder displays. In *Proc. European Conference on Ambient Intelligence*, pages 248–265, 2007.
10. T. Payne, E. David, N. R. Jennings, and M. Sharifi. Auction mechanisms for efficient advertisement selection on public displays. In *Proceedings of European Conference on Artificial Intelligence*, pages 285–289, 2006.
11. J. Rehg, M. Loughlin, and K. Waters. Vision for a smart kiosk. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 690–696, 1997.
12. D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 137–146, New York, NY, USA, 2004. ACM.