

User Study for Generating Personalized Summary Profiles

Lalitha Agnihotri

Philips Research

345 Scarborough Road

Briarcliff Manor, NY

lalitha.agnihotri@philips.com

John R. Kender

Dept. of Computer Science

Columbia University

New York, NY 10027

jrk@cs.columbia.edu

Nevenka Dimitrova

Philips Research

345 Scarborough Road

Briarcliff Manor, NY

nevenka.dimitrova@philips.com

John Zimmerman

Dept. Of Comp Science

Carnegie Mellon

University

Pittsburgh, PA

johnz@cs.cmu.edu

Abstract

The need for personalized summaries of media content has been driven by the recent and anticipated explosive growth in the media world. In this paper we present a methodology and a supporting user study for generating user profiles and content features that can be used to automatically create personalized summaries of broadcast television content. We determined a mapping, from users' personality traits measured by commonly available personality tests, to computable video features that such personality traits appear to prefer. Three common personality profiles (Myers-Briggs, Merrill Reed, and Brain.exe) were elicited from 59 subjects, together with their preferred summary of news, music, and talk show videos. A factor analysis between the personality traits and the features in preferred summaries indicated that only some traits (e.g., gender, extraversion, control orientation, intuitiveness, etc.) and only some features (e.g., faces, reportage, text, chorus, host, etc.) had predictive value. The mapping of personality to feature also differed by genre. However, in general, extraverted users tended to prefer directly experienced content, while introverted users preferred content mediated through analysis. A validation user study is in progress.

1. Introduction

General video summarization will be insufficient when the amount of content grows beyond our ability to search it quickly and easily. A powerful approach for summarization involves the personalization of subject matter (semantics), how it is presented (form), and where and when it is presented (context) [1].

Literature abounds on video summarization [2]. However, little attention has been given to personalized video summarization. Even more, there are no methodologies for generating user profiles at video features level. In order to produce personalized summaries we need an extensive user profile containing preferences to video attributes. We hypothesize that there exists a mapping of personality traits to the preference for inherent video features. In order to establish the mapping between personality traits and computable video features we performed a user study. In this paper, we will detail our methodology and the design of the user study, summarize the results, and provide some initial conclusions.

This paper is organized as follows: In section 2, we introduce personalization. We outline our proposed methodology in section 3. The data analysis is presented in section 4 and the results are presented in section 5. We conclude and present future work in section 6.

2. Personalization

One approach for personalization for the user is to obtain a detailed profile that can be used to filter incoming content. Explicit and implicit profiles have been used in systems built for recommending TV programs [3]. The explicit recommender relies on results from a question-answer session with the viewer, wherein the viewers' explicit likes and dislikes towards particular TV channels, show genres etc. are elicited. The implicit recommenders use a viewer's implicit profile, which is built from the viewing history of a TV viewer. Many users want either minimal or even no interaction with the system in order to make such systems work.

Profiles should require minimal user input. Also, they should accurately represent the user's desires. In order to meet these challenges, we decided to test the hypothesis that the personality traits of the user would serve as an accurate basis for their user profile. We know from commercial media research performed to set advertising ratings, that different TV shows appeal to different demographics of users. We also know that people relate to one another differently based on personality. The media equation states that people react to media the same way they interact with other people [5]. We want to know if there is a relation between personality and inherent video features. Since these different interpersonal strategies make up much of what is called "personality," it is likely that measured personality traits also play an important role in how people interact with the media.

Video has inherent properties called video features: face presence, text presence, anchor segment etc. Our hypothesis can be stated that there is a mapping of personality traits of the user to the preference for features that are inherent in the video. The goal of the present study as depicted in Figure 1 is to explore and to establish a methodology to find this mapping. On one side we have personality attributes and on the other side video features. We are trying to uncover a mapping that possibly exists between the two.

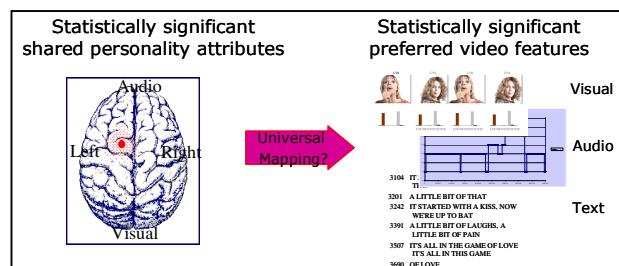


Figure 1 Personality traits to video features mapping

There are many possible personality inventories. One that has been well studied is Myers Briggs Type Indicator (MBTI) [3]. MBTI maps an individual into four characteristics: Extravert vs. Introvert (E/I), Sensation vs. Intuition (S/N), Thinker vs. Feeler (T/F), and finally Judger vs. Perceiver (J/P). In order to minimize the dependence of our results on a specific personality inventory test, we employed two other approaches. In the second personality inventory, Merrill Reid [8], categorizes users into Ask vs. Tell (A/T) and Emote vs. Control (E/C) groups. We chose this one due to the availability of literature and the ability to anticipate certain mapping patterns. For example, sensation people might prefer to get more details (numbers, names etc.). While intuitives might be satisfied with a bigger pictures. As a third test, we decided to use "brain.exe"[7] as an unquantified but popular internet-available personality inventory test. Some of these tests indeed had predictive power. We note that although there are established and rigorous standards for validating personality tests, our interest here is more practical. If our analyses detect a consistent correlation between a measured personality feature and a preferred video feature, we simply exploit that correlation for video summary purposes, independently of those deeper issues of personality testing explored by the psychology community.

Video content analysis community has been working on automatic extraction of audio, visual, and text features from video programs [6]. We annotated a number of these features for our test videos summaries in order to uncover the mapping. Example features include dark vs. bright, text and face presence, who the speaker is, past vs. present vs. future, etc.

3. Methodology

The methodology involves users to take several personality tests and provide their personality traits and then select summaries for a series of videos. For each video, they choose those segments, images, texts, and sounds that summarize the story best for them.

Statistical tools (principal components, factor analysis, histograms, etc.) are then used to discover significant associations from personality to features. We sought a dependable a robust mapping between results of personality test and computable visual, audio, and text features.

3.1 A Case for User Tests

User tests were performed in order to uncover patterns of personality and their mapping to content analysis features. The well-known phrase, "Buyers are Liars!" to realtors who are approached by buyers with a wish list of things they want to have in a house but cannot afford. This maxim is true from the point of view designing this user study. A user is able to determine whether he or she likes particular media content, but is unable to accurately assess the exact features of the media that are responsible for this disposition. Representative real-life use scenarios were constructed and users' preferences were determined through answers to forced choice questions. Thus we opted for complete full media content summaries rather than verbal descriptions of summary content.

3.2 Testing Paradigm and Data Collection

We decided to let the users pick the summary of their choice and then analyzed the video features in the selected segment in order to come up with user preferences. For the data collection task, we designed a web site that the users stepped through. The

users initially gave their personality data and then gave the audio-visual selections for video segments.

3.2.1 Personality Data Collection

Users were asked to enter their name, age, and gender. After this users navigated to the personality information pages. In the first two pages users selected their personality features for Myers Briggs Type Indicator and Merrill Reid based their choice on attributes typical of each personality trait. Figure 2 shows a sample list that the users read through in order to make their choice for extravert vs. introvert trait. For the third personality test, the users answered the twenty questions in the test "brain.exe." At the end of the test, they entered their scores on the third personality test as computed by the program.

Extraverts often:	Introverts often:
• Have high energy	• Have quiet energy
• Talk more than listen	• Listen more than talk
• Think out loud	• Think quietly inside my head
• Act, then think	• Think, then act
• Like to be around people a lot	• Feel comfortable being alone
• Prefer a public role	• Prefer to work "behind-the-scenes"
• Can sometimes be easily distracted	• Have good powers of concentration
• Prefer to do lots of things at once	• Prefer to focus on one thing at a time
• Are outgoing & enthusiastic	• Are self-contained and reserved

● Extravert (E) or ● Introvert (I)

Figure 2 List for Extravert vs. Introvert

3.2.2 Data selection for preferred summary

Figure 3 shows the summary selection page. Subjects first watched the original video in its entirety. On the right the transcript of the video was presented. The users then scrolled down to see two or three pre-selected video only summaries. The users could either choose one of these videos or could specify their own video segment. Similarly, they chose one summary of two or three pre-selected audio only summaries. Finally they selected one of four pre-selected images. In this way subjects selected summaries for eight news stories, four music videos, and two talk shows.

Before the test started, the users were given a brief introduction under five minutes of the task they were expected to do. No mention of relating personality to summary selection was made until after the session was over. For their participation in the user test, the subjects were given \$10 each.



Figure 3 Summary selection page

3.3 Critique of methodology.

In retrospect, we could have structured the data gathering so that it was both more effective and more efficient. Had we collected from a small sample of users the data on the personality tests alone, we would have quickly found through correlations that Brain.exe performed no better than randomly. We then would have eliminated it as a user test, and would have derived simpler user profiles that were easier to understand. A small pilot

sample could have eliminated some of the obviously insignificant video features. Had we attempted to build the automatic feature detection algorithms prior to testing and hand annotation, we would have simplified the analysis in two ways. First, some features, such as the "brightness" detector, proved later to be very difficult to automate due to human subjectivity, and were therefore eliminated anyway. Second, having an automated annotation would have sped the analysis greatly, and would have eliminated annotator error; instead of using a three step process (trait to summary to feature), we could have investigated the relationship of personality to video features directly.

Additionally, we are aware of several potential causes for statistical bias. Our users selected themselves by responding to our advertisement and, possibly, to a promise of a reward. Our sample populations had no controls for education or socio-economic status, and were already somewhat clustered by personality type, by having been drawn predominately from within a research environment (although we did have a number of support staff participating as well). Our personality tests, except for the useless Brain.exe, were much abbreviated from their original forms. And, since we used actual broadcast segments, some of our subjects may have already been familiar with the content beforehand.

3.4 Implementation and Design Issues

The implementation required us to research and resolve some engineering issues. We used QuickTime player embedded in the web page for displaying the original video and also the audio and video summaries. We used HTML and PHP which is a server-side scripting language for creating dynamic Web pages that were used to generate the user test web pages. As the viewers browsed and entered information in the web page the next pages were automatically loaded and the user selection data was stored in text file from each page.

4. Data Analysis

A total of fifty-nine subjects (16 female and 43 male) participated in our user study. They were a mix of researchers working at Philips Research in USA and The Netherlands and students at Columbia University. The subjects spent approximately two hours entering their personality data and their preferred summaries for news, music videos, and talk shows. A concept value matrix was created and was analyzed for generating mapping between personality traits and video features. In the matrix, there was one row for each of the users ($u=59$) who participated in the user test. The initial columns were derived from the personality tests that the user completed. We have $q(=10)$ personality features. V stands for video analysis features. We have w video analysis features which varies by genre. So our concept value matrix is of $ux(q+w)$ dimension.

4.1 First Order Statistics

We first plotted histograms of responses for selection of videos. We wanted to investigate if variability exists in the selection of audio, video, and image segments. If in the histograms, it turned out that everybody consistently picked up the same video and same audio for a given video, then we would not need personalized summarization at all. As an example Figure 4 shows the histograms number of times each audio segment was selected for a specific news story and each video summary for a

specific music video. It can be seen in the figure that there is no clear winner among the summaries. The final bar shows how many people chose their own summary. There is no clear preference for a single summary. An investigation of user preferences for all three genres indicated that there was enough variability in user preferences that further exploration of the underlying correlates was necessary. There is individual variation and we think personality can capture at least part of it.

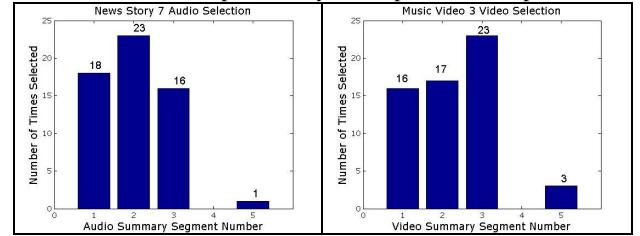


Figure 4 Histogram of Audio Selection of News Video and Video Selection of Music Video

4.2 Second order statistics

In order to find significant patterns in our mapping between personality and content analysis features, we performed extensive factor analysis on our data. Factor analysis is a statistical technique used to reduce a set of variables to a smaller number of variables or factors. Factor analysis examines the pattern of inter-correlations between the variables, and determines whether there are subsets of variables (or factors) that correlate highly with each other but that show low correlations with other subsets (or factors). We used the "factoran" function in MATLAB that computes the maximum likelihood estimate (MLE) of the factor loadings matrix λ in the factor analysis model

$$X = \mu + \lambda f + e$$

Where X is any observed user's vector of dimension $q+w$; μ is a fixed vector of means valid across all users; λ is called the factor loadings matrix, and is of dimension $q+w$ by t , where t is the number of requested factors; f is a vector of independent, standardized common factors; and e is a vector of independent specific factors. By inspecting λ , we were able to determine which personality traits and which video features tended to cluster together, and, conversely, which traits and features were best considered idiosyncratic noise in e . Additionally, by monitoring the significance of the results as t was varied, we were able to determine how much commonality of preferences there was across users.

5. Experimental Results

Using the functions provided by MATLAB we performed extensive factor analysis and eliminating traits and features that did not show much variance, we derived the following results. The results from brain.exe were eliminated earliest, as they did not correspond to any of the trends. One possible explanation for this could be that we had a continuous scale from -1 to 1 for this test, but for others, we were constrained to either -1 or 1 due to the way the "tests" were administered. Thus we could not exploit the richness of this test with responses from fifty-nine users. Another caveat is that the personality traits for the MBTI and MR were not really obtained via user tests but by asking the users to read two lists of features. This was due to unavailability of tools for testing and time constraints.

The final factors for news, talkshows, and music videos are given in Figure 5-Figure 8. The graphs scales from -1 to 1 and depicts the strength of each of the features in the resulting factors. Our results showed genre dependency in the resulting mapping that was obtained. Only some personality traits say things about specific genres. And what they say differs from genre to genre. Only four personality factors out of the original ten showed up in the final factors with a strength > 1.2 . The three different genres responded differently in terms of what people needed to select.

For news, the significant personality features are gender, extravert/introvert, and emotive/control. This factor can be read as suggesting that females, introverts, and people with a control orientation tend to dislike faces, like text, and to prefer to have their news summary include the actual reportage rather than the anchorperson commentary. Additionally, it also says the converse: female, extraverts, and emotive people like faces, dislike text, and prefer being told the news by the anchor. For talk-shows we have two factors which can be summarized as follows: Intuitive people prefer the host saying something from past that is personal in nature. Extravert, Thinkers, prefer when the guest that is present in the video is speaking about his professional life. For music videos the only significant factor shows that Controls like to see text, prefer bright portion, and a section other than chorus of the song.

The primary difference we have observed is that extraverts tend to prefer a pure experience of the video content, whereas introverts tend to prefer to have the content mediated through a host or anchor. To our knowledge, this is a novel result. We were unable to find any reference in the literature to any similar significant dichotomy between a direct versus an indirect interaction with the media.

6. Conclusions and future work

We have presented a methodology for determining how personality traits are correlated with preferred summary content. We have shown that only a small number of traits and only a small number video features appear to influence the subjective properties of summary quality. Further, it appears that these traits and features are genre-dependent, except for the qualitative observation that the extraversion dimension appears to predict the value placed on direct experiences.

Despite our limited population, the results of the factor analysis suggest that a number of these factors are heavily weighted, and are therefore likely to be stable and reproducible phenomena. Nevertheless, we are conducting a second set of user tests to validate the strongest of these factors, and we will run measures of statistical significance on these predictors of user preference.

7. References

- [1] L. Agnihotri et. al, "Study on Requirement specifications for Personalized Multimedia Summarization," ICME 2003.
- [2] M. Barbieri, L. Agnihotri, and N. Dimitrova, "Video Summarization: Methods and Landscape," SPIE ITCOM, 2003.
- [3] J. Zimmerman, et. al, "TV Personalization System: Design of a TV Show Recommender Engine and Interface," Personalized Digital TV, Kluwer 2004.
- [4] D. Keirsey, M. Bates, "Please Understand Me: Character and Temperament Types," 1984, Prometheus Nemesis Book Co.
- [5] B. Reeves, C. Nass, "The Media Equation : How People

Treat Computers, Television, and New Media Like Real People and Places," CSLI Lecture Notes Series.

- [6] G. Wei, L. Agnihotri, N. Dimitrova, TV Program Classification Based on Face and Text Processing, ICME 2000.
- [7] Synergistic Learning Incorporated, <http://ilearn.seneca.on.ca/techwrite/testing.html>

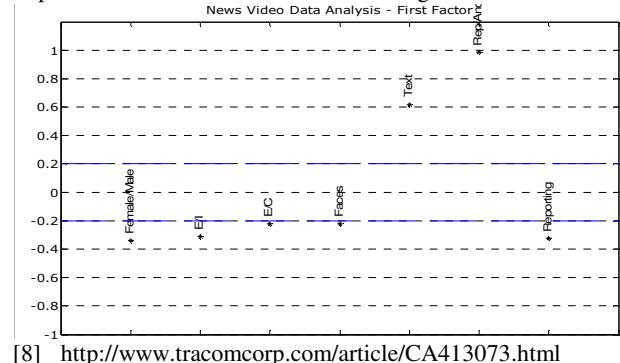


Figure 5 Significant factor for News

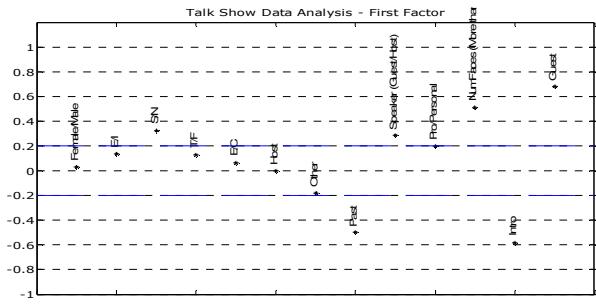


Figure 6 First significant factor for Talkshows

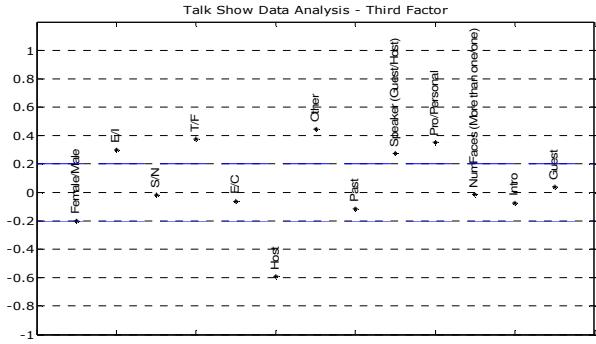


Figure 7 Second significant factor for Talkshows

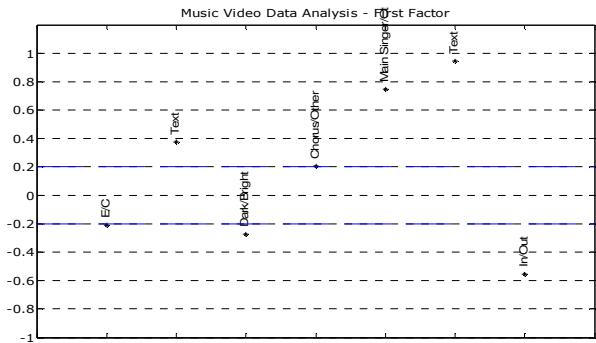


Figure 8 Significant factor for music videos