Enhancing Children's Experience with Recommendation Systems

Yashar Deldjoo
Politecnico di Milano
Milan, Italy
yashar.deldjoo@polimi.it

Cristina Frà
Telecom Italia S.p.A. Services
Innovation, Milan, Italy
cristina.fra@telecomitalia.it

Massimo Valla
Telecom Italia S.p.A. Services
Innovation, Milan, Italy
massimo.valla@telecomitalia.it

Antonio Paladini
Politecnico di Milano
Milan, Italy
antonio.paladini@mail.polimi.it

Davide Anghileri
Politecnico di Milano
Milan, Italy
davide.anghileri@mail.polimi.it

Mustafa Anil Tuncel
Politecnico di Milano
Milan, Italy
mustafaanil.tuncel@mail.polimi.it

Franca Garzotto
Politecnico di Milano
Milan, Italy
franca.garzotto@polimi.it

Paolo Cremonesi
Politecnico di Milano
Milan, Italy
paolo.cremonesi@polimi.it

ABSTRACT

Recommender Systems (RSs) offer a personalized support in exploring large amounts of information, assisting users in decision making about products matching their taste and preferences. Most of the research todate on recommender systems have focused on traditional users, *i.e.,* adult individuals who are able to offer explicit feedback, write reviews or purchase items themselves. However, children's patterns of attention and interaction are quite different from those of adults.

This paper presents the first results of a research-in-progress that can be suited to bridge the barrier between children and a recommender system by providing a child-friendly interaction paradigm. Specifically, a web application is developed that employs real-time object recognition on movie thumbnails or DVD cover-photos in a real-time manner. The tangible object can be manipulated by the user and provide input to the system for the purpose of generating movie recommendations. We plan to extend this work to the scenario where the child could ask for a video content showing a related toy (*e.g.,* a car, a plane, the doll of a character that she likes in a cartoon) and the system could generate the videos that matches these implicit preferences expressed by the child.

CCS CONCEPTS

• **Information systems → Recommender systems**;

KEYWORDS

Recommender System, Kid, Child, Movie recommendation, visual recognition, cover photo, movie thumbnails, toy recognition, doll

1 INTRODUCTION

Recommender systems (RSs) are subclass of information filtering techniques which recommend items to users by considering past user-item interactions and inferring knowledge about what can be potentially intriguing for users [17]. Most previous research on recommender systems have focused on traditional users, *i.e.* adult individuals who are able to offer explicit feedback, write reviews, or purchase items themselves. Children's patterns of attention and interaction are quite different from those of adults. Children differ from adult learners in many ways, but there are also commonalities across learners of all ages. It is known today that with carefully designed methods, one could find ways to pose rather complex questions about what infants and young children know and can do [3]. While RS for adults has been well-investigated, RSs for children are in their infancy and just starting to be gaining attention and studied in the community of RSs. The main works on RSs for children however, as is to date, have been limited on education-related environments [13, 15].

This paper presents the first results of a research-in-progress that attempts to bridge the barrier between children and a recommender system by providing a child-friendly interaction paradigm. In our research, we extend traditional RS interfaces (GUIs) with tangible interaction [2, 12] in which the user expresses his/her preferences by grasping some physical items and user profiling information is generated using image processing of such artifacts. The paper exemplifies this approach in the case of movie recommendation. Specifically, we describe a web-application that employs real-time object recognition on movie thumbnails or DVD cover-photos in

a real-time manner. By pointing the DVD cover-photo toward the webcam of the proposed system, the RS is provided with the information which is relevant for elicitation purposes and to build useful recommendations. We designed and implemented the system to be used easily and intuitively during shopping at media stores or at home. We discuss new opportunities for research and applications that may emerge from the integration of tangible interaction for child product recommendation.

2 RELATED WORK

Back in the 19th century, it was commonly thought that infants lack the ability to form complex ideas. For much of this century, most psychologists widely believed the traditional thesis that at birth the human mind is a "blank slate" (known as *tabula rasa*) on which knowledge is recorded gradually from experience. It was further thought that language is the main prerequisite factor for abstract thoughts and in absence of that a baby could not have knowledge [3]. But challenges to this view arose recently where psychologists began to realize about the remarkable abilities that young children possess that stands in stark contrast to the older emphases on what they lacked. A major move away from the *tabula rasa* view of the infant mind was taken by the Swiss psychologist Jean Piaget, a leading figure in analyzing how children's cognitions evolve. According to Piaget's theory , children's development falls into a series of stages:

(1) Sensorimotor (birth to 2 years)
(2) Preoperational (ages 2 to 7)
(3) Concrete Operational (ages 7 to 11)
(4) Formal Operational (ages 11 and up)

In the sensorimotor stage, cognition of the infant is limited to her own actions on the environment as she is heavily dependent on what her senses immediately perceive. All the interactions for infant's at this young age must come in some multimedia form, for example audio, video or animation. In the preoperational stage, little interaction can be expected from the child and all the interaction must come in some multimedia form, for example audio, video or animation. In the concrete operational stage, children's thinking is maturing. At this stage it is reasonable to expect fine control of the mouse input device from children. Also, children's ability to use simple keyboards and to type, grows out through this age group. By the time a child reaches the formal operational stage, designers can assume the child's thinking is generally similar to that of adults. Their interests and tastes, of course, remain different. Designing for this age group is much less challenging, because adult designers can at least partially rely on their own intuitions [1, 18]

The relevant literature consider two main aspects of the problem: (1) Child-Computer Interaction (CCI) papers (2) Algorithmic paper. Despite significant number of papers published related to CCI, only a few works present recommendation solution tailored for Kids considering their needs from multiple perspectives: educational, developmental, and engagement, to name a few [13, 15]. In fact, while RSs for adults have been studied for several years, RSs for children are just in their infancy.

3 METHOD DESCRIPTION

This section presents the development of the first version of a web-based application which provides movie recommendations and includes tangible interaction in the recommendation process. The results of a user study using the web application can be found in [11]. The current work is a new effort to introduce a kid-friendly interaction paradigm into the system in order to make it suitable for kid product recommendation.

3.1 System Architecture

To support the application scenario described in this work the system architecture represented in Figure 1 has been designed. It's a modular solution where the MISRec[1] application exploits a NodeJS cloud component that is connected with several backend modules to offer the proper recommendations to the users.

Specifically, MISRec is a completely web based application running on a screen (*i.e.,* a smart TV or an interactive display) equipped with a camera (*i.e.,* a regular webcam or a more sophisticated sensor like the Microsoft Kinect). MISRec, acting as a smart mirror, detects the presence of a user in front of the camera and, as soon as the video stream is stable enough, it captures a frame and sends it to the TIM (Telecom Italia) Image Matching component. As opposed to the early version of this work presented in [6], where the user had to manually capture a video shot touching a button on the screen, in this work we improved the capturing with automatic object (cover-photo) detection, with a reliable detection algorithm.

The Image Matching component is a general purpose image recognition system developed by Telecom Italia which is beforehand trained by thousands of movies cover photos obtained from IMDB. For each item to be recognized (in our case for each movie in the system catalogue) one or more training image is added to the system; at this stage the system can recognize an input query image based on its visual similarity to the training covers dataset. This module implements the MPEG Compact Descriptors for Visual Search (CDVS) Standard [14] and exposes a set of open RESTful APIs for both training and query features. The response of a query request is a list (in case empty) of unique movie ids representing detected similar movies, ordered according to a similarity score.

The movie catalogue is stored in a backend database where, for each unique id identifying the movie is stored a set of attributes and tags needed to both calculate the recommendation and provide movie metadata to the user. When the MISRec captures the camera frame, it sends a query request to the TIM Image Matching module. The returned id of the similar item with the highest score is used to retrieve from the catalogue the metadata related to that specific movie and to query the Recommender System. The recommendation algorithm integrated in the Recommender System is detailed in Section 3.2.

3.2 Recommendation algorithm

The MISRec is equipped with a recommendations engine that is capable of producing movie recommendation given a query movie. The latter is provided to the system as the result of movie cover photo recognition from the previous step. Each item in the dataset, is associated with its feature vector f_i. For the purpose of empirical

---

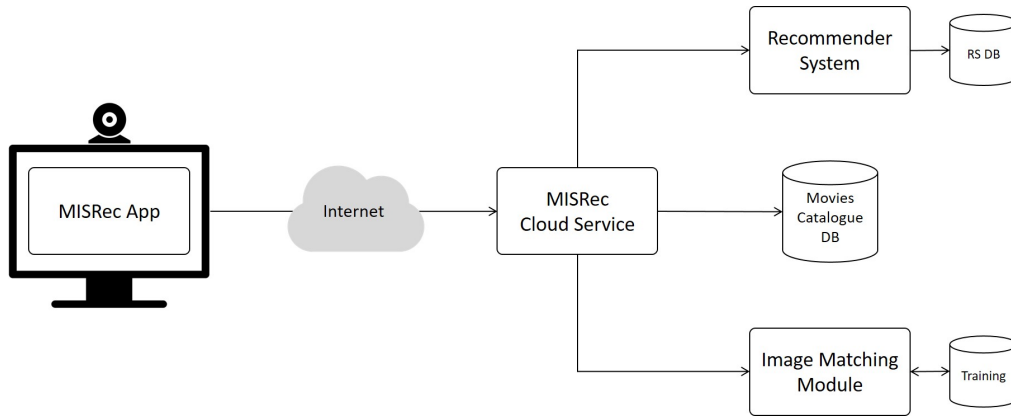[1]short for Mise-en-Scène Movie Recommender

Figure 1: MISRec System Architecture

studies our system is integrated with two classes of movie-related features which can be selected on the choice of the designer. These movie-related features are summarized as follows:

3.2.1 Metadata attributes. The MISRec recommender is integrated with the metadata associated with each movie. The metadata define some high-level attributes of the movies and include: genre, director, and actors. Such metadata are human-generated, either editorially (*e.g.* title, cast) or by leveraging the wisdom of the crowd (*e.g.*, tags). these information are collected from Movielens and IMDB API.

3.2.2 Mise-en-Scène features. Recently, low-level stylistic features – such as color, motion and lighting – have been proposed for content-based video recommendations (mise-en-scène features) [8]. Mise-en-Scène features can be extracted automatically by processing the video files. We have selected five categories of mise-en-scène low-level features described in [5, 7, 8, 16] as they are explainable, easy to extract from video files, and show promising results in offline experiments: *shot duration*, *color variance*, *lighting key*, *average motion*, and *motion variation*. For each pair of items i and j, the similarity score s_{ij} is computed using *cosine similarity*. In order to generate recommendations, we adopted a classical content-based similarity search based on "k-nearest neighbor".

4 APPLICATION SCENARIOS

Figure 2 shows user's interaction with the system. The user is invited to show the cover of a movie that she is currently interested in, to the system (*e.g.,* if in a movie store, the user can select a DVD on a shelf and show it to the system webcam) as shown in Figure 2a. On the the system backend works a general purpose image recognition system which implements the MPEG7 Compact Descriptors for Visual Search (CDVS) standard [10, 14] and can be trained to recognize any kind of items with a defined shape (*e.g.,* 2D images like magazine covers or brand logos and 3D ones like common retail products exposed in shops, monuments, buildings profile, *etc.,*). As opposed to the earlier version of this work [6], the detection is done automatically without the need for the user to manually press a capture button as shown in Figure 2b and Figure 2c. Once detected, the detection results are used to retrieve similar movies between the

just detected movie and the movies in our dataset based on different notions of similarities, *e.g.* similarity based on (1) movie genres or (2) plots as this information accompanies movies as metadata (and are stored in our dataset) and, (3) mise-en-scène characteristic of the videos. An example of this recommendation is shown in Figure 2d.

(a)   (b)
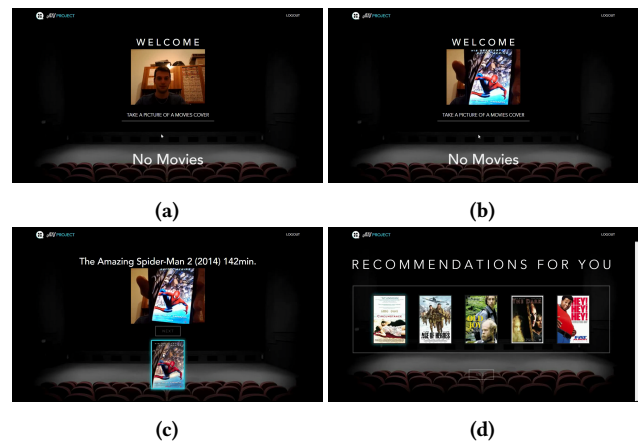

(c)   (d)

Figure 2: Screenshots of the first version of the proposed movie recommender web application coupled with tangible interaction for a user-friendly interaction engagement

The Interaction with a RS by grasping objects would be appealing and intuitive particularly in multimedia recommender for young children. A scenario could be the following as illustrated in Figure 3. A child could ask for a video content showing a related toy (*e.g.* a car, a plane, the doll of a character that she likes in a cartoon) as shown in Figure 3a (all the recommended movies are about cars). The system could generate video recommendation that matches these implicit preferences expressed by the child as shown in Figure 3b. This would require manual/automatic annotation of such symbolic objects present in the video beforehand. The recommendation list are updated in a subsequent phase by taking into account the age group of the kid (detected automatically or provided to the system

beforehand) or by mise-en-scène characteristic of the video. In the latter case for example, the fact children at low age prefer to watch cartoons is a phenomenon that can be captured by mise-en-scène features. The would results in generation of a new and more matching list of recommendation for the kid (considering her age and style preference) as shown in Figure 3c . Note that in the final list, all the recommended movies are about cars but tailored better for kids (*e.g.* they are all cartoons).

In principle, this approach could facilitate the communication with the system, offering also the benefit of embodied cognition in developing cognitive skills, and supporting the development of fine motor skills. This is another arena for empirical research that has never been explored in the Interaction Design and Children community. Because of the physical dimension of this paradigm, tangible interaction would be motivating for the users and could engage them in an easier, more active and more entertaining manner with the RS.

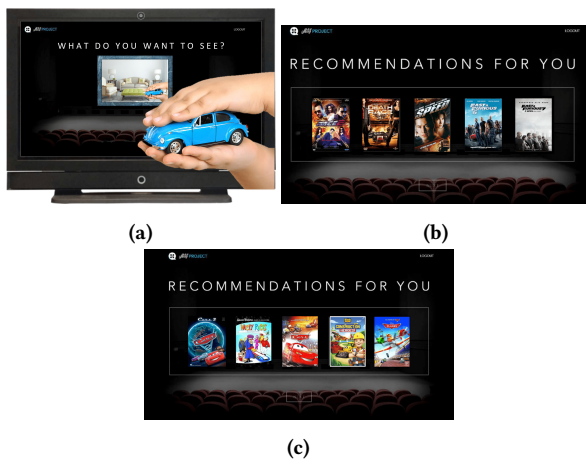
(a)   **(b)**

(c)

Figure 3: Illustration of a hypothetical scenario where the proposed interaction paradigm can be used for kid-friendly communication with a RS in the future work.

5 CONCLUSION

Traditionally, most previous research on RSs have focused on adult users who are able to offer explicit feedback, write reviews, or purchase items themselves. However, children's patterns of attention and interaction are quite different from those of adults. In this paper, we proposed a web-based application namely MISRec for prototyping of tangible interaction integrated with a recommender system. In our approach, the user expresses his/her preferences by grasping some physical items and user profiling information is generated using image analysis of such artifacts.

The Interaction with a RS by grasping objects would be appealing and intuitive particularly in multimedia recommender for young children. A scenario could be the following. A child could ask for a video content showing a related toy (*e.g.* the doll of a character that she likes in a cartoon) and the system could generate the videos that matches these implicit preferences expressed by the child.

Advances in visual recognition technology enables the adoption of a number of embodied interaction paradigms to interact with recommender systems, that could include, beside tangible interaction, also gesture & motion based interaction [4, 9] and emotion-based interaction (exploiting face expressions recognition) to make the UX with the system more engaging and pleasing. As for the future plan, we intend to include design of additional TI scenarios together with users studies and field evaluation involving real users belonging to the properly selected targets.

ACKNOWLEDGMENTS

REFERENCES
[1] David F Bjorklund and Kayla B Causey. 2017. *Children's thinking: Cognitive development and individual differences.* SAGE Publications.
[2] Guy André Boy. 2016. *Tangible Interactive Systems: Grasping the Real World with Computers.* Springer.
[3] National Research Council et al. 2000. *How people learn: Brain, mind, experience, and school: Expanded edition.* National Academies Press.
[4] Yashar Deldjoo and Reza Ebrahimi Atani. 2016. A low-cost infrared-optical head tracking solution for virtual 3D audio environment using the Nintendo Wii-remote. *Entertainment Computing* 12 (2016), 9–27.
[5] Yashar Deldjoo, Paolo Cremonesi, Markus Schedl, and Massimo Quadrana. 2017. The effect of different video summarization models on the quality of video recommendation based on low-level visual. In *Content-Based Multimedia Indexing (CBMI), 2017 15th International Workshop on.* ACM.
[6] Yashar Deldjoo, Fra Crisitina, Valla Massimo, and Paolo Cremonesi. 2017. Letting Users Assist What to Watch: An Interactive Query-by-Example Movie Recommendation System. In *Proc. 8th Italian Information Retrieval Workshop (IIR) 2017.* CEUR Workshop Proceedings.
[7] Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi, Franca Garzotto, and Pietro Piazzolla. 2016. Recommending movies based on mise-en-scene design. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, 1540–1547.
[8] Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi, Franca Garzotto, Pietro Piazzolla, and Massimo Quadrana. 2016. Content-based video recommendation system based on stylistic visual features. *Journal on Data Semantics* 5, 2 (2016), 99–113.
[9] Antonella Di Rienzo, Paolo Tagliaferri, Francesco Arenella, Franca Garzotto, Cristina Frà, Paolo Cremonesi, and Massimo Valla. 2016. Bridging Physical Space and Digital Landscape to Drive Retail Innovation. In *Proceedings of the International Working Conference on Advanced Visual Interfaces.* ACM, 356–357.
[10] Ling-Yu Duan, Vijay Chandrasekhar, Jie Chen, Jie Lin, Zhe Wang, Tiejun Huang, Bernd Girod, and Wen Gao. 2016. Overview of the MPEG-CDVS Standard. *IEEE Transactions on Image Processing* 25, 1 (2016), 179–194.
[11] Mehdi Elahi, Deldjoo Yashar, Farshad Bakhshandegan Moghaddam, Cella Leonardo, Cereda Stefano, and Cremonesi Paolo. 2017. Exploring The Semantic Gap for Movie Recommendation. In *Proceedings of 11th ACM Conference of Recommender Systems.* ACM.
[12] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems.* ACM, 234–241.
[13] Nikos Manouselis, Hendrik Drachsler, Riina Vuorikari, Hans Hummel, and Rob Koper. 2011. Recommender systems in technology enhanced learning. In *Recommender systems handbook.* Springer, 387–415.
[14] S. Paschalakis and G. Francini. 2015. Information technology-multimedia content descriptor interface-part 13: Compact descriptors for visual search. *ISO/IEC 15938–13: 2015* (2015).
[15] Maria Soledad Pera and Yiu-Kai Ng. 2014. Automating readers' advisory to make book recommendations for k-12 readers. In *Proceedings of the 8th ACM Conference on Recommender systems.* ACM, 9–16.
[16] Zeeshan Rasheed, Yaser Sheikh, and Mubarak Shah. 2005. On the use of computable features for film classification. *Circuits and Systems for Video Technology, IEEE Transactions on* 15, 1 (2005), 52–64.
[17] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: Introduction and challenges. In *Recommender Systems Handbook.* Springer, 1–34.
[18] Andrew Sears and Julie A Jacko. 2009. *Human-computer interaction: Designing for diverse users and domains.* CRC Press.