Automatically Recommending Multimedia Content for Use in Group Reminiscence Therapy

Adam Bermingham
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Ireland
abermingham@computing.dcu.ie

Julia O'Rourke
Department of Speech & Language Therapy
Adelaide & Meath Hospital
Tallaght
Ireland
Julia.ORourke@amnch.ie

Cathal Gurrin
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Ireland
cgurrin@computing.dcu.ie

Ronan Collins
Age-Related Health Care
Adelaide & Meath Hospital
Tallaght
Ireland
ronan.collins2@amnch.ie

Kate Irving
School of Nursing
Dublin City University
Ireland
kate.irving@dcu.ie

Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Ireland
alan.smeaton@dcu.ie

ABSTRACT

This paper presents and evaluates a novel approach for automatically recommending multimedia content for use in group reminiscence therapy for people with Alzheimer's and other dementias. In recent years recommender systems have seen popularity in providing a personalised experience in information discovery tasks. This personalisation approach is naturally suited to tasks in healthcare, such as reminiscence therapy, where there has been a trend towards an increased emphasis on person-centred care. Building on recent work which has shown benefits to reminiscence therapy in a group setting, we develop and evaluate a system, REMPAD, which profiles people with Alzheimer's and other dementias, and provides multimedia content tailored to a given group context. In this paper we present our system and approach, and report on a user trial in residential care settings. In our evaluation we examine the potential to use early-aggregation and late-aggregation of group member preferences using case-based reasoning combined with a content-based method. We evaluate with respect to accuracy, utility and perceived usefulness. The results overall are positive and we find that our best-performing approach uses early aggregation CBR combined with a content-based method. Also, under different evaluation criteria, we note different performances, with certain configurations of our approach providing better accuracy and others providing better utility.

MIIRH'13, October 22, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2505323.2505333 .

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process

Keywords

recommender systems, reminiscence therapy, search, information retrieval, usability, multimedia, healthcare

1. INTRODUCTION

Reminiscence therapy (RT) has seen success in recent years as a method of therapy for people with Alzheimer's and other dementias. RT refers to the guided recollection of previous life experiences or subjects of interest either in a group or individual context. RT has been proven to have a positive effect in terms of increased life satisfaction, decreased depression, and increased communication skills and patient-caregiver interactions [4].

In a typical RT session, a facilitator (for example a clinician or activity co-ordinator) uses *cues* to stimulate recall of memories. These cues may be objects from a person's past or old photographs, for example. More recently, digital cues have been used in the form of multimedia content.

Identifying relevant content to use in reminiscence therapy can be a time-consuming and resource-intensive task. Traditionally, therapy facilitators have kept either paper or mental records of a person's life history and interests so that they may make an informed decision about which content would likely be beneficial to use in an RT session. RT participants are also encouraged to maintain scrapbooks of their own past, known as *lifebooks*. These methods have significant drawbacks in terms of scalability because of the resources necessary to produce them and the challenges inherent in trying to re-purpose materials or identify generalisable materials for use in a group setting.

Other factors which make identification of reminiscence materials a challenging task include generational and cultural barriers between the facilitator and person with dementia, acquired communication difficulties in dementia and a lack of a collateral history to inform patient biography where such difficulties exist.

During RT sessions, the facilitator often needs to make decisions quickly and monitor participants' reactions, limiting the time they can devote to finding new materials, say in a digital library, during the RT sessions. A common approach is to plan sessions beforehand. However, apart from the extra time required, such a rigid approach limits flexibility in terms of adapting when a pre-planned stimulus has proven ineffective during a session and so sessions need to be dynamic and reactive to the circumstances of how it is unfolding.

Therefore the requirements of a system to support group RT are that it should be efficient, accurate, personalised and provide a high degree of utility to the facilitator, ultimately leading to successful group RT outcome. It should also not distract from other tasks and should seek to be relevant to *all* group members. Our system, REMPAD (**R**eminiscence Therapy **E**nhanced **M**aterial **P**rofiling in **A**lzheimers and other **D**ementias), addresses these requirements using a novel group-recommender approach to multimedia RT.

There are a number of benefits of performing reminiscence therapy in a group context. In particular, therapy sessions enjoy a social component as participants can share experiences and discussion. Whereas it can be challenging to identify suitable content in a one-on-one context, identifying suitable content for a set of individuals is a much more difficult task for facilitators. The facilitators must identify content which optimally benefits the group, while minimising any negative effect. For example, a video which some group members find engaging might be undesirable if this induces a negative effect in other members.

Thus there are motivations and challenges for the application of a recommendation and search approach to supporting group digital reminiscence therapy. Due to the nature of RT, there are a number of task-specific requirements and constraints which make it different from other group recommender systems, which have been traditionally focused on tasks in areas such as e-commerce and entertainment. The approach we take addresses specific challenges related to RT as an application area.

In this paper we present a multimedia system for modelling group preferences and recommendation algorithms and integrating them into an RT system. Our approach uses a combination of case-based reasoning recommendation, content-based recommendation and search to address RT facilitators' content needs. The focus of our evaluation is to assess the efficacy of the recommendation algorithm. Our results are based on a user trial we conducted in residential care homes with 7 user groups. We examine the accuracy and utility of content suggested by the REMPAD system through analysis of system usage logs as well as explicit ratings from users, comparing a number of system configurations. We also report on usability interviews with RT facilitators who participated in the trial.

The rest of this paper is structured as follows. In Section 2 we position our research in the context of related work in the fields of Reminscence Therapy and Recommender Systems. We describe our approach in the REMPAD system in Section 3, followed by Experiments and Results and Discussion in Sections 4 and 5. We conclude in Section 6.

2. BACKGROUND AND RELATED WORK

In this section we discuss related work in the field of recommender systems. First however, we look at reminiscence therapy, and in particular how it is suited to a recommender systems approach.

2.1 Reminiscence Therapy

Reminiscence Therapy (RT) is an intervention that is commonly used to address the psychosocial problems of persons living with dementia [39]. RT involves the discussion of past activities, events and experiences with another person or group of people, usually with the aid of tangible prompts such as photographs, household and other familiar items from the past, music or archive sound recordings [39]. More recently the video sharing website YouTube has been used as a source to facilitate access to digital RT content [30].

Reminiscence groups typically involve structured group meetings in which participants are encouraged to talk about past events at least once a week. A group leader or facilitator assists and guides the group members to recall previous life experiences and facilitates the group's affirmation of the value of these experiences [6]. This activity aims to improve mood, well-being, communication and to stimulate memory and strengthen a sense of personal identity [5], [39]. This treatment is based on the assumption that autobiographical memory remains intact until the later stages of dementia and may be used as a form of communication with the person with dementia [28].

There is evidence to suggest that RT is effective in improving mood in older people without dementia and its effects on mood, cognition and well-being in dementia are present, but less well understood [39]. Improvements in autobiographical memory selectively in RT groups for mild-to-moderate degree dementia have also been described [27], [13]. Despite the limited empirical study of reminiscence undertaken, the vast majority of the results indicate the positive effects of reminiscence [39], [13].

Autobiographical memory is characterized by *multiple types of knowledge*, and refers to a memory system consisting of episodes recollected from an individual's life. This is based on a combination of episodic memories (personal experiences and specific objects, people and events experienced at a particular time and place) and semantic memories (general knowledge and facts about the world) [38], [12]. Flashbulb memories are a particular type of autobiographical memory of vivid mile marker events with associated personal meaningful experiences [34]. They rely on elements of personal importance, consequentiality, emotion, and surprise [11]. They may include collectively shared public events marked by their uniqueness and emotional impact. Autobiographical memories may be accessed more easily and with greater frequency in old age, precisely because they are more robust and less likely to dissipate than memories of everyday commonplace experiences such as what you had for dinner last week. Autobiographical memories include multi-sensory information about the experiential context, including sights, sounds and other sensory and perceptual information. A song, a scent, or simply a word can evoke autobiographical memories.

RT can also be conducted on a one-to-one level but is distinct from life review therapy (LRT). LRT typically involves individual sessions, in which the person is guided chronologically through life experiences, encouraged to evaluate them, and may produce a life story book as a result. Although the procedures are different, both RT and LRT involve the recollection of past experiences (events, emotions and relationships).

Facilitated reminiscence exploits the relatively well preserved autobiographical memories to enhance communication opportunities for older adults who may differ in abilities, cultural background, and life experiences [16]. As a reminiscence facilitator, the Speech and Language Therapist possesses knowledge of the developmental aspects of typical aging and understanding of cognitive-communicative disorders in aging [1]. In facilitating a traditional RT group the SLT manages the selection of topics, scheduling, group composition and communicative interactions between and among group participants. An understanding of the participants' shared historical experiences is the starting place for topic selection [16]. This is achieved by firstly, considering the personal interests, likes and dislikes of individuals. Secondly, the flashbulb memories shared by particular age cohorts and thirdly, universally experienced developmental life events such as childhood, schooldays, adulthood, marriage, work life, retirement.

Creative therapeutic approaches are required to facilitate the socialization needs of residents and to appeal to an increasingly culturally and linguistically diverse population. Facilitated RT programmes therefore need to be simultaneously engaging, relevant, cost effective and culturally sensitive [15]. Mismatches can arise in age, life experience and culture between majority culture clinicians and older adults from non-mainstream populations [31] or vice versa. Hence the need for detailed group member profiling is important to enable positive and successful reminiscence facilitation. REMPAD builds on our previous research in which the use of video and other digital multisensory content to stimulate conversation and social interaction was found to be a feasible in group reminiscence therapy sessions [30].

A comprehensive approach towards the person with dementia that takes into account their life history is essential. The *person-centred approach* to dementia situates the person with dementia at the centre of all aspects of caregiving [18] [9]. The focus is on identifying and meeting the needs of the person, in contrast to the medical model which focuses on identifying and treating symptoms. The person-centred approach aims to enhance well-being by improving relationships and communication between people with dementia, their families and professional caregivers. This is achieved by taking into account the life experiences and the likes and dislikes of each person with dementia in order to develop a greater understanding of the individual. This in turn allows for care tailored specifically to the individual to take place. *Person-centredness* is achieved when carers and family members focus more on the individual than on the illness.

To address the needs of the residential care population and their associated activity coordinators REMPAD proposes a solution to enhance facilitator knowledge and provide access to personalized reminiscence material for the benefit of aiding conversation and memory recollection amongst nursing home participant users in a group context.

2.2 Recommender Systems

In this section we provide background and related work in the area of recommender systems. There are broadly three categories of recommender system: those based on user matching (collaborative), those based on learning content preferences (content-based) and those that use a knowledge base approach (case-based reasoning). We describe each of these in turn and how they relate to the REMPAD system.

Much work in recent years in the area of recommender systems has focussed on user-item rating prediction through inference over large datasets. A common approach is to make predictions of user-item ratings based on the previous ratings for that item of similar users, known as *collaborative recommendation*. Perhaps the most salient example is the Netflix prize [3] which pushed forward the state of the art in large-scale collaborative recommendations systems. A characteristic of collaborative recommender systems is that they rely on the availability of large amounts of data. Also, a collaborative approach relies solely on user-item rating information, rather than any information about the items themselves. These user ratings may not be able to model certain aspects of the recommendation task.

A second popular category of recommendation is *content-based recommendation*. In content-based recommender systems, a user's preferences are stored based on their previous interactions or ratings of items. The system then learns from these preferences so that they may identify new items to recommend. Content-based recommenders rely on the system being able to explicitly model properties of objects. The advantages of content-based systems include transparency in the recommendation decisions and the ability to recommend new items never seen before by the system, provided the necessary features can be extracted. A drawback is the uncertainty when a new user uses the system and the limitations in terms of how items can be modelled, sometimes referred to as the *semantic gap*. Pazzani and Billsus provide an overview of content-based recommenders [32] and Lops et al. provide a recent review of the state of the art [21].

A third approach to recommender systems is based on *case-based reasoning* (CBR). CBR approaches are those which rely on a knowledge base representation of known items and item context. Although CBR approaches can vary, in particular to the extent that they implement the full CBR process, the most common CBR Recommenders use a stated preference from a user and use a similarity function to match the parameters in that preference to the item descriptors in the knowledge base [22]. This process could also involve other information relevant to the recommendation task such as user profile, preference refinement and previous uses of the system. A limitation of CBR systems is the need to create and maintain a knowledge base of items and as with content-based recommenders, the semantic gap. An advantage is the intuitiveness with which a user can express their preference and if necessary, refine their requirements, in many ways similar to interactive search systems. CBR systems are of particular use in e-commerce systems where users are looking for products to purchase. Overviews of the adaptation of the CBR process to recommendation tasks are provided by Bridge et al. [8] and by Smyth [36].

For group-based RT, a collaborative approach is not feasible due to the small size of the user group. Our system uses a hybrid approach consisting of a CBR and a content-based recommender, supported by a traditional search feature for query refinement, and a novelty multiplier. To mitigate the limitations of these approaches we use the CBR approach to bootstrap the content-based approach. In order to create our knowledge base of users and items we adapt traditional methods for profiling RT participants and use an efficient curation and annotation process to produce low-cost item descriptors.

Some recent works have examined the more complex task of recommending content for groups of individuals. In groups with disparate sets of preferences, it is not clear how to optimally recommend content for a given group context. Popular approaches seek to minimize misery, maximize individual utility or use an aggregated measure of group satisfaction.

McCarthy et al.'s work has tackled the group problem from a case-based perspective using iterative interactive critiquing of cases among group members to reach and optimal solution [24], [25], [26]. An early review of group recommenders is provided by Jameson and Smyth, outlining the significant challenges in moving from individual to group recommendation [17]. Another early work from O'Connor et al. uses collaborative filtering to produce lists of movie recommendations for groups to watch [29]. They introduce a *minimum misery* strategy i.e. the overall satisfaction in a group is directly related to the satisfaction of the least happy group member. Later we will see this is a principle we employ in REMPAD. Recently, Masthoff has compared group recommender systems from the literature, noting the different strategies used for aggregating individual profiles [23]. Although many systems use relatively straightforward strategies to simulate group recommender systems using *individual* recommenders, more complex approaches have been tried to explicitly model *group* preferences [10]. However, perhaps due to the typical dearth of group-level ratings, or the complexity of the task, most approaches use an array of individual recommenders.

Although we are aware of some recent works which investigated using digital systems for RT [2] [14] [20], to the best of our knowledge the REMPAD system is the first system to implement an algorithm to recommend content in the context of group RT. We take inspiration from the aformentioned CBR and content-based approaches, but design our system with specific considerations for the RT application domain such as minimizing interactivity and task complexity, and maintaining tight constraints on preventing dissatisfaction among group members and recommendation dead-ends.

3. APPROACH

In our approach we model a system for use in a care setting with a group of people with mild-moderate dementia and an activity co-ordinator. In this section we describe our approach to recommendation. There are two types of users in our system: the activity co-ordinator, or clinician, who facilitates the session, and the therapy participants themselves. We use *item* to refer to a video indexed by our system; *user* to refer to a therapy participant; *group* to refer to a therapy group, consisting of a set of users; and *facilitator* to refer to the clinician who runs the session and physically interacts with the system.

3.1 The REMPAD System

The REMPAD system is a cloud-based service which is accessed through a mobile device such as a tablet. This interface controls the application flow, interpreting participant requirements, selecting content to display on a second larger screen and providing online feedback to the system.

A typical session flow is as follows:

1. The user creates a new session and selects which group members are present.
2. The system executes a recommendation query and the user is presented with a list of suggested videos, a results list. Initially the top two ranked videos are displayed. If they wish, the user can browse through the results list two videos at a time.
3. At this stage, the user can (i) select a video to play, (ii) select two results from the list to suggest to the group[1], (iii) issue a search query to refine the results list.
4. Upon selecting a video to watch, the video is displayed on the shared viewing screen. At this time the facilitator is presented with a feedback screen where they note perceived user and group satisfaction with the content.
5. After each video the system may return to (2) or end the session.
6. At the end of the session, the facilitator enters overall user and group feedback for the session.

The system is designed to support sessions with minimal intrusion on the role of the user who must also monitor and engage with the group participants during the sessions. Typically a session lasts about 45 minutes and a group will watch several videos in a session.

3.2 Data Curation and Annotation

The data we use in our system is from the popular video sharing website, YouTube[2]. YouTube has been previously used successfully in reminiscence therapy by the authors [30]. By its nature, YouTube is suitable for use in our system. There is an abundance of content available through standard APIs and each video is accompanied with rich metadata. The content itself is diverse and esoteric, reflecting the variety of uploaders and sharing needs on YouTube. This content is useful for RT as there is often content relevant to niche subjects, people, places, events which may not be covered in more mainstream content sources.

Although we had intended using YouTube metadata for organising and presenting videos in the REMPAD system, initial testing revealed that the quality and consistency of metadata were not of sufficient standard to support the system requirements. To address this we used a curation and annotation process. The project team consisting of research assistants, clinicians, and postdoctoral researchers, curated content using a custom curation interface. This interface offers a search functionality which uses the YouTube search API to find videos relevant to areas of interest, times and locations which are suggested to the curator. We provided curators with subject matter targets reflecting a broad range of media types and content. The curator then previews videos and if happy with the content can queue the videos for annotation.

The index used in the system contains a wide range of video content. Examples include documentary excerpts, home

[1] providing therapy participants with a binary choice is a common approach in RT

[2] `http://www.youtube.com`

videos, music recordings, interviews and sports. Curators were advised to search for videos ideally less than 5 minutes and no more than ten minutes so that they were appropriate for use in RT.

An important concept in RT is orientation towards people, places and times. To offer a personalised experience, we also wish to model a user's preferences and interests. The metadata produced by the annotation process for a video includes title, description, location(s), date, people, seasons/holidays as well as vectors describing relevance to a variety of genres, media, music, interests and sports.

Initially the authors annotated 343 videos. This can be a time-intensive task, taking approximately 3 or 4 minutes per video. To reduce the cost of indexing content, we obtained a further 258 video annotations using the crowdsourcing service, CrowdFlower[3].

The crowdsourced annotations were added to the video index at approximately the halfway stage in the trials to prevent staleness of content. In order for the system to perform effectively, it needs to provide usable recommendations amongst the top results (ideally top 2) or otherwise risk slowing the facilitator and disrupting the momentum of the RT session. Even though the index we use in these trials is relatively small, it is still a significant challenge to produce useful recommendations at the very top of the results list. We have designed the system and processes to be scalable as significantly expanding the user base and index is a goal of future work.

3.3 User Profiling

The user profiles are gathered through short interviews with users before the first use of the system. This is inspired by existing practices in care settings where a record is often made of people's life history and interests. Similar to video metadata, the metadata we collect for users includes date of birth, locations lived in, and interest vectors related to genre, media, music, interests, sports, similar to the video vectors. A key difference with users is we allow them to also express dislike using a 5-point Likert scale whereas the equivalent for video was either categorical or on a 3-point relevance scale: not relevant, relevant, highly relevant. In the following section we refer to the concatenation of the genre, medium, music, interests, sports vectors as simply the *feature vector* for users and items.

3.4 A Recommender Model for RT

Our recommender algorithm consists of a scoring function which is used to proactively rank items for a given recommendation context consisting of a group of users, their previous item ratings and interactions, and optionally a search query.

We model a user u as having three features: a location, a date of birth and a feature vector whose values are normalised to between -1 and 1.

$$u = < u_l, u_d, u_f > \tag{1}$$

Similarly, we model an item i as having four features: a location i_l, a date i_d, an interest vector with values normalised between 0 and 1 i_f, and a textual description i_t.

$$i = < i_l, i_d, i_f, i_t > \tag{2}$$

[3]http://www.crowdflower.com

A search query q is given by two optional fields: a text query q_t and a decade q_d.

$$q = < q_t, q_d > \tag{3}$$

The scoring function for an item i, given a group of users, G, and an optional search query, q is:

$$S(i, G, q) = \left(\frac{w_1 S_{CBR}(i, G) + w_2 S_C(i, G) + w_3 S_{Rel}(i, q)}{\sum\limits_{j=1,2,3} w_j} \right) * N \tag{4}$$

where S_{CBR} is the CBR scoring function; S_C is the content-based scoring function; S_{Rel} is the relevance function; and N is a novelty multiplier. In our system, we present two options for S_{CBR}. In the first, $S_{CBRlate}$, we aim to aggregate individual preferences at a late stage using a minimum misery approach.

$$S_{CBRlate}(i, G) = \min_{u \in G}(S_{CBRlate}(i, u)) \tag{5}$$

For each individual user the function uses a linear combination of three similarity functions:

$$S_{CBRlate}(i, u) = Sim_{date}(i_d, u_d) + Sim_{loc}(i_l, u_l) + Sim_{feat}(i_f, u_f) \tag{6}$$

In line with the priorities of good reminiscence content, the date similarity function upweights items related to recent events or to events that occurred when the user was aged below 30. We also provide a small bonus to items from before the user was born which may be of historical or cultural interest.

$$Sim_{date}(i_d, u_d) = \begin{cases} 1 & \textit{when } i_d - u_d < 30 \textit{ yrs} \\ 0.75 & \textit{when now} - i_d < 10 \textit{ yrs} \\ 0.25 & \textit{when } i_d < u_d \\ 0 & \textit{otherwise.} \end{cases} \tag{7}$$

Similarly, the location similarity function upweights the best specific matches between user and item:

$$Sim_{loc}(i_l, u_l) = \begin{cases} 1 & \textit{when regions match} \\ 0.5 & \textit{when countries match} \\ 0.1 & \textit{when continents match} \\ 0 & \textit{otherwise.} \end{cases} \tag{8}$$

The similarity between feature vectors is given by the Cosine Similarity between the feature vectors:

$$Sim_{feat}(i_f, u_f) = \textit{Cosine Similarity}(\vec{i_f}, \vec{u_f}) \tag{9}$$

The second of our CBR scoring functions, we aggregate preferences into a single meta profile for the group from the outset. $S_{CBRearly}$, consists of a linear combination of similarity functions, but this time interpreted at a group level:

$$S_{CBR_{early}}(i, G) = Sim_{date}(i_d, G_d) + Sim_{loc}(i_l, G_l) + Sim_{feat}(i_f, G_f) \tag{10}$$

$$\textit{where } G_x = \{u_x : u \in G\} \tag{11}$$

This can be seen as treating the group as a *meta-user*. For date, we simply model the date for the group as the mean point in time, given the range of dates of birth:

$$Sim_{date}(i_d, G_d) = Sim_{date}(i_d, \overline{u_d}) \tag{12}$$

$$where\ \overline{u_d} = \frac{1}{|G_d|} \sum_{u_d \in G_d} u_d \tag{13}$$

For locations we use the best match for a common location in the group:

$$Sim_{loc}(i_l, G_l) = \max_{u_l \in G_l} (i_l, u_l) \tag{14}$$

where $u_l \in G_l$ and u_l is common to 2 or more members of group G.

To compare features at a group level we consider positive features and common negative features:

$$Sim_{feat}(i_f, G_f) = Common_{pos}(i_f, G_f) - Common_{neg}(i_f, G_f) \tag{15}$$

In order to identify the common positive features we rank the features according to the number of users in the group who have declared each feature as an interest or strong interest and take the top m features, F_{pos}. Similarly, in order to identify the common negative features we rank the features according to the aggregate score from the users in the group, and take the top n features, F_{neg}. For the negative ranking we assign 1 to a *dislike* and 2 to a *strong dislike*, thus emphasising extreme negative preferences. We also create a set of relevant features for each item, F_{rel}. The commonality scores are then given by:

$$Common_{pos}(i_f, G_f) = \frac{|F_{pos} \cap F_{rel}|}{m} \tag{16}$$

$$Common_{neg}(i_f, G_f) = \frac{|F_{neg} \cap F_{rel}|}{n} \tag{17}$$

In our experiments we set m to 40 and n to 20.

$S_C(i, G)$ is given by the output classification probability of the positive class from a multinomial naive Bayes classifier trained on positive and negative examples for the group G. An item i is a positive example for group G if it satisfies the following criteria:

- There has been no negative item ratings from group G for item i.
- There has been no negative item ratings for user u for item i, $u \in G$.
- There has previously been a positive item rating from group G, or from $u \in G$, for item i.

There is just a single criterion for an item to become a negative example:

- There has been negative group-level or individual feedback from Group G for item i.

If the number of examples in the positive set is below a threshold r, we bootstrap the process by adding the r top-ranked examples by S_{CBR} to the positive set. Similarly if the size of the negative set is less than r, we add r lowest-ranked examples by S_{CBR} to the negative set. In our experiments we set r to 5. The features used for classification are the item feature vector i_f.

In a case where a user has chosen to enter a search query, the search query-item relevance is given by:

$$S_{Rel}(i, q) = \frac{w_4 Rel_{text}(i_t, q_t) + w_5 Rel_{date}(i_d, q_d)}{\sum_{j=4,5} w_j} \tag{18}$$

where $Rel_{text}(i_t, q_t)$ is the score given by a search over an index item text fields (title, description, people), i_t, using the search platform Solr[4]. We reward queries if they are from the same decade or a neighbouring decade as a candidate items:

$$Rel_{date}(i_d, u_d) = \begin{cases} 1 & \textit{when from same decade} \\ 0.5 & \textit{when from neighbouring decades} \\ 0 & \textit{otherwise.} \end{cases} \tag{19}$$

We set $w_4 = 2$ and $w_5 = 1$, emphasising the specificity of a text query, particularly as a common search task is known-item search, where the facilitator is trying to find an item they are aware is in the index.

Novelty often has an important role in recommender systems [37]. In order to prevent the results list becoming predictable and familiar, we penalise results if they have been recently browsed or played. This novelty function has a decay so as to allow familiar videos to move back up the results list as the time since they were last browsed or played increases. In REMPAD there is both a requirement to show novel results in the list *and* to ensure that known familiar and useful content is rediscoverable[5].

Let $n_b(i, G)$ be the number of queries since item i was last browsed in a results list for group G. Let $n_p(i, G)$ be the number of queries since item i was last played in group G. We define the novelty multiplier N then to be:

$$N(i, G) = \frac{w_6 \log(\min(n_p(i, G), h)) + w_7 \log(\min(n_p(i, G), k))}{\sum_{j=6,7} w_j} \tag{20}$$

We set $h = 5$ and $k = 10$ in our experiments, and upweight the importance of playing an item over browsing, with $w_6 = 2$ and $w_7 = 1$.

4. EXPERIMENTS

We trialled our system over a period of several weeks involving over 50 users in 7 therapy groups across 6 locations, residential care homes. See Table 1 for details of groups and sessions for those groups.

Our evaluation has two focuses. First we wish to ascertain the degree to which the recommender has supported the reminiscence therapy sessions for the groups in our study. Secondly, we wish to investigate the comparative performances of different configurations of our algorithm[6]. The

[4]`http://lucene.apache.org/solr/`

[5]For this reason we also provide *favourites* and *history* functions which are sometimes used

[6]As this is not a controlled study, our ethical approval does not extend to using a control as one of our experimental con-

Group	Members	Sessions Completed	Videos Played	Videos per Session
A	7	4	21	5.25
B	8	9	59	6.56
C	8	9	61	6.78
D	7	6	55	9.17
E	8	11	72	6.55
F	7	5	26	5.2
G	11	10	68	6.8
Total	56	54	362	6.7

Table 1: Session and video play counts for trial groups.

four configurations we use are (i) $S_{CBRearly}$ without S_C, (ii) $S_{CBRearly}$ with S_C, (iii) $S_{CBRlate}$ without S_C, (iv) $S_{CBRlate}$ with S_C. These configurations were assigned to sessions for groups in a latin squares arrangement[7].

In both cases, our evaluation focuses on three aspects: (i) accuracy, (ii) utility and (iii) perceived usefulness. Unlike some recommenders, our multimedia system is based on ranked recommendation lists, akin to a search system. For accuracy we compare system-ranked lists to reference rank lists as rated using a given group of annotators, using Spearman's rank correlation coefficient, ρ. In this approach, we construct ideal lists for users and groups given knowledge of their item ratings. We then use these as references with which we correlate a given ranking produced by the system [35].

For utility we use *R-Score*. R is appropriate in scenarios like ours, where the user can only use a small set of items and the user is unlikely to be exposed to the majority of the items in the ranked list. R incorporates a half-life, α, which is equivalent to approximately the rank at which the user has a 0.5 chance of browsing the item, thus incorporating likelihood of observation of a given recommendation [35] [7]. In our experiments we set α to 5. For calculating both ρ and R we use user-item ratings and group-item items and present them as mean values over a given set of ranked lists returned by the system.

Recently there has been an emphasis on the importance of user experience and the perceived usefulness in evaluation of recommender systems [33] [19]. To reflect this in our evaluation we also use end-of-session group and user ratings ratings. For reporting these scores, we conflated any Likert or other ordinal scales to a three-point scale: positive, neutral, negative. We then average these values assigning +1 to positive, -1 to negative, 0 to neutral, giving an average score, r, in the range (-1,1) for a rating of feedback values.

It is worth noting that in our multimedia system, these ratings are important as they are the clinician's interpretation of the satisfaction of the individuals, and group, in the therapy sessions. This is a natural extension of the facilitator's role in terms of monitoring, interpretting reacting to therapy participants' reaction to stimulus.

ditions. This has precluded us from exposing people with dementia to potentially weak experimental conditions such as randomly selected content which might not suit their tastes.

[7]In practice this was difficult to maintain as users often created impromptu sessions for training and testing purposes which were later removed from the trial data.

Group	R user	R group	mean ρ user	mean ρ group
A	17.51	28.13	0.11	0.10
B	11.22	9.34	0.19	0.13
C	3.51	4.89	0.20	0.23
D	6.49	0.62	0.06	0.04
E	11.47	12.33	0.09	0.09
F	3.40	3.64	0.16	0.20
G	13.30	12.87	0.08	0.10
All	9.62	9.42	0.13	0.13

Table 3: Utility (R Score) and accuracy (Spearman's ρ) scores for groups and total.

5. RESULTS AND DISCUSSION

The results overall from our trials are positive and show that the system is effectively supporting the content discovery task for the facilitator during a group RT session. 69% of queries successfully resulted in a played video. Typically, unsuccessful queries resulted in the facilitator either refreshing to obtain a new list of recommendations, refining the query using the search function, or playing a previously viewed video from favourites or history. Inspection of our logs reveals that search was only used in a minority of cases. The search query terms suggest that the most common search need was to find a result either viewed previously or previously browsed in a results list, a pattern sometimes called *known-item search*.

In 43% of successful queries, the video chosen was on the first screen (top two recommendations), 73% in the first 3 screens (top six recommendations) and 86% in the first 5 screens (top ten recommendations, see Figure 1). Facilitators appear to be comfortable choosing from near the top of the results list, consistent with a high level of satisfaction and trust in the recommendations.

Looking closer at the explicit online ratings that the facilitators provide to the system, we see they are overall very positive (see Table 2). 62% of user-item ratings were positive, with just 1% negative. Similarly, 49% of group-item ratings were positive, with just 3% negative. The reason the group-item ratings were not as positive as the user-item ratings likely reflects the comparative difficulty in recommending items for a group rather than an individual. Looking at end-of-session feedback, we observe the same pattern.

For 6 of the 7 groups, the user session-ratings were more positive than item-ratings. This pattern also holds for group ratings for 5 of the 7 groups. This is interesting as it agrees with the intuition that the probability of overall satisfaction is higher if the user or individual is evaluating over a series of recommendations, as they may be tolerant of some inaccuracies i.e. a user may be satisfied with a session without necessarily giving a positive rating for each video in that session.

Unlike the ratings, R and ρ show no significant difference between groups and users when it comes to either accuracy or utility (see Table 3). The R-Score for groups does vary in some cases, showing much higher group utility than user utility for group A and a lower group utility than user utility for group D. In the former case, group A has by far the lowest proportion of non-neutral item ratings, so perhaps this has an effect, although how is unclear. For accuracy, we see that each of the rank correlations are positive, although

	user-session					group-session					user-item					group-item				
	n	+1	0	-1	r	n	+1	0	-1	r	n	+1	0	-1	r	n	+1	0	-1	r
A	25	0.36	0.64	0.00	0.36	4	0.25	0.75	0.00	0.25	132	0.35	0.64	0.02	0.33	59	0.34	0.66	0.00	0.34
B	56	0.64	0.36	0.00	0.64	9	0.33	0.67	0.00	0.33	348	0.57	0.41	0.02	0.55	72	0.63	0.38	0.00	0.63
C	51	0.71	0.22	0.08	0.63	9	0.56	0.44	0.00	0.56	317	0.61	0.35	0.04	0.56	55	0.55	0.45	0.00	0.55
D	29	0.83	0.17	0.00	0.83	6	0.67	0.17	0.17	0.50	255	0.70	0.29	0.01	0.69	21	0.24	0.76	0.00	0.24
E	63	0.86	0.11	0.03	0.83	11	0.91	0.09	0.00	0.91	417	0.71	0.28	0.00	0.71	61	0.33	0.52	0.15	0.18
F	26	0.73	0.27	0.00	0.73	5	0.80	0.20	0.00	0.80	128	0.77	0.23	0.00	0.77	26	0.77	0.23	0.00	0.77
G	70	0.61	0.39	0.00	0.61	10	0.60	0.40	0.00	0.60	478	0.56	0.43	0.01	0.56	68	0.57	0.41	0.01	0.56
All	320	0.69	0.29	0.02	0.67	54	0.61	0.37	0.02	0.59	2075	0.62	0.37	0.01	0.60	362	0.49	0.48	0.03	0.47

Table 2: Ratings for each group and total.

S_{CBR}	S_C	mean R		ρ	
		user	group	user	group
late	no	11.42	10.20	0.08	0.09
early	no	8.37	8.20	0.19	0.19
late	yes	9.80	10.53	0.08	0.08
early	yes	9.03	8.78	0.12	0.13

Table 4: Utility (R Score) and accuracy (Spearman's ρ) scores for 4 system configurations.

relatively weak. It should be noted that novelty has had negative effect on both accuracy and utility as we report it here. The novelty multiplier deliberately pushes recently seen videos far down the results list. As we have seen, the majority of these will have had a positive rating, and R and ρ will be negatively affected as a result.

With the recommender configurations, we wish to compare the two forms of S_{CBR} and to look at the impact of including S_C. Thus, two important questions in our experiments are (a) does altering the method of computing S_{CBR} have an effect? and (b) does integrating S_C into the scoring function have an effect? In Table 5 we examine the difference in four system configuration comparisons: comparing early aggregation CBR with late aggregation CBR (i and ii); and comparing CBR with and without content-based recommendation (iii and iv). See Table 6 for user and group ratings according to system configuration and Table 4 for ρ and R.

For the base case (i) we find $S_{CBRearly}$ performs better than $S_{CBRlate}$ for ρ, but lower on all other measures. Thus our method for combining profiles into a meta-profile before employing CBR similarity functions does not perform as well as a minimum misery late aggregation approach in terms of utility or ratings, but has a higher accuracy. Integrating S_C (ii) appears to reduce the disparity between the CBR approaches. In this case, ρ is still significantly higher for $S_{CBRearly}$ than $S_{CBRlate}$, but the difference is smaller. We also see the gap lessen for R and ratings, particularly user-session ratings, where $S_{CBRearly}$ performs significantly better, producing the highest ratings for user-session, group-session and group-item. It would appear that $S_{CBRearly}$ with S_C is somewhat of a sweet spot, balancing the individual and group preferences in S_C with the meta-profile used for $S_{CBRearly}$.

Adding S_C to $S_{CBRlate}$ (iii) appears to significantly hurt performance from a user perspective, but not for groups. This is the standout case in which we observed a difference in how users and groups respond to different experimental conditions.

Our results show that ρ is at odds with some of the other measures. For example, configurations using $S_{CBRlate}$ have

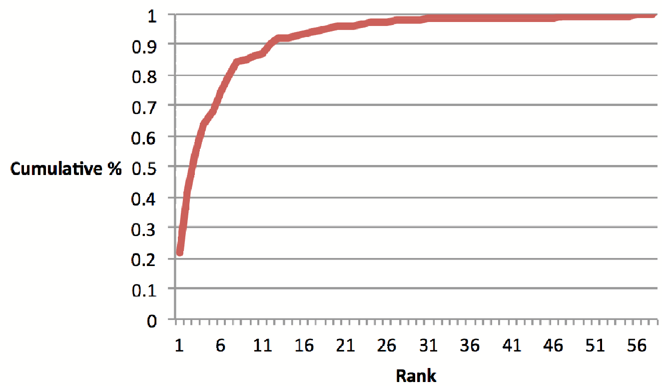

Figure 1: Cumulative rank of selected item for successful queries

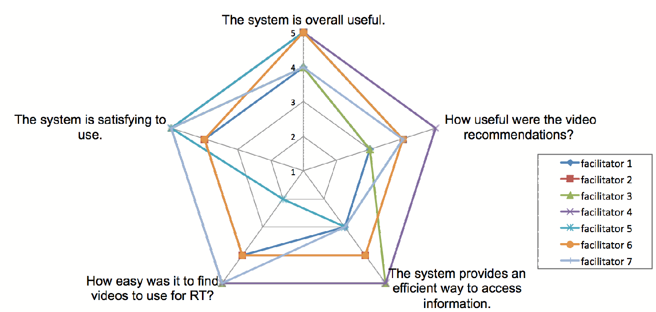

Figure 2: Facilitators' responses to usability questions

a higher R but a lower ρ; adding S_C to $S_{CBRearly}$ disimproves ρ but performance improves across other measures. This intriguing observation suggests it is possible that switching between late aggregation and early aggregation CBR, or indeed using a weighted combination, would enable us to tune the system by trading accuracy for utility.

After our trials, the facilitators participated in a semi-structured interview. Two aspects we focused on were ease of use and perceived usefulness of the recommendations (see Figure 2). The responses were positive, with all facilitators agreeing that the system was satisfying and useful. They were also predominantly positive about the ease and efficiency with which they could find those items and the usefulness of those items. Some unstructured feedback emphasised the requirement that speed, efficiency, novelty and accuracy are important, and even the smallest delay or frustration

Recommender Configuration (A,B)					ρ		R		$r_{session}$		r_{item}	
	S_{CBR}	S_C	S_{CBR}	S_C	user	group	user	group	user	group	user	group
i	late	no	early	no	0.10**	0.11**	-3.00**	-2.00	-0.21*	-0.20	-0.08*	-0.09
ii	late	yes	early	yes	0.04**	0.04**	-0.77	-1.75	0.23*	0.20	0.05	0.07
iii	late	no	late	yes	0.00	0.00	-1.62	0.32	-0.19*	-0.06	-0.07*	0.04
iv	early	no	early	yes	-0.07**	-0.07**	0.61	0.58	0.24*	0.33	0.07*	0.20

Table 5: Difference in recommender configurations (B-A) with statistical significance at p<0.05 (*) and p<0.001 (**)

S_{CBR}	S_C	user-session				group-session				user-item				group-item			
		+1	0	-1	r	+1	0	-1	r	+1	0	-1	r	+1	0	-1	r
late	no	0.78	0.22	0.00	0.78	0.64	0.36	0.00	0.64	0.67	0.31	0.02	0.64	0.90	0.00	0.10	0.79
early	no	0.59	0.39	0.02	0.58	0.50	0.44	0.06	0.44	0.57	0.43	0.01	0.56	0.85	0.01	0.14	0.70
late	yes	0.63	0.34	0.04	0.59	0.57	0.43	0.00	0.57	0.59	0.39	0.02	0.58	0.90	0.04	0.06	0.83
early	yes	0.83	0.16	0.01	0.81	0.77	0.23	0.00	0.77	0.64	0.35	0.01	0.63	0.94	0.01	0.04	0.90

Table 6: Ratings for 4 system configurations, altering method for computing S_{CBR}, and optionally including S_C.

with the system can have a negative effect, unlike other applications where users are perhaps more tolerant.

6. CONCLUSION

In this paper we have contributed a novel approach to recommending multimedia content to use in group RT. We provided background and related work in the fields of RT and recommender systems, motivating the work and outlining the limitations of existing approaches. We introduced a method based on a hybrid system using CBR recommendation, content-based recommendation and search to satisfy the system requirements. We developed and trialled this system over a period of several weeks in residential care homes and have reported on the efficacy of the proposed approach in terms of accuracy, utility and perceived usefulness.

We find, in general, a higher proportion of positive item ratings for individual users than groups, reflecting the greater difficulty in recommending for groups. We also find that session ratings are higher both for groups and users than individual item ratings. These observations suggest that although it is harder to recommend for groups than individuals, recommending a set or sequence of videos (as in our sessions) may have significant advantages over single recommendations. We see some variance for utility across groups and in general we find accuracy and utility to be consistent between group and user ratings.

Our best performing system configuration uses a combination of an early aggregation CBR and a learning-based content method, possibly reflecting the richest representation of user and group preferences. We also observe that a late aggregation CBR approach with minimum misery appears to favour utility, wheres an early aggregation CBR approach favours accuracy. This potentially gives scope to build a system which is tuneable for accuracy versus utility.

Finally in interview feedback from users we learn of a unanimous satisfaction with the system and a reinforcement of the initial requirements for a responsive, accurate, efficient and easy-to-use system to support facilitation of RT sessions. This is perhaps a strong motivation to focus on a utility as evaluation measure for systems in this area.

For the work we have done to date, the features we used are quite specific to the RT application and somewhat heuristic and so for future work we intend to enrich our preference representations further, in particular using text features such as TF-IDF. We will also expand our content collection and investigate the possibility of introducing collaborative recommendation approaches. In order to enrich our content collection we also will explore using other sources of video and other forms of content.

Overall we find recommending content for use in group RT challenging task and one that is naturally suited to a recommender systems approach. A discussion point that naturally falls out of our work is one of the relationship between modelling group and user preferences. There is evidently an interplay between the two, as, although the ultimate goal is individual therapy participant satisfaction and successful reminiscence, this may not be possible without achieving group satisfaction. Similarly, group satisfaction is likely not attainable without individual satisfaction. This is an important question to address and provides an interesting avenue for future research both for group RT systems and more generally in the area of group recommender systems.

Acknowledgements

This work is supported by Science Foundation Ireland under grant 07/CE/I1147 and Enterprise Ireland under grant CF/2011/1318.

7. REFERENCES

[1] American Speech-Language-Hearing Association. Evidence-based practice in communication disorders: An introduction. *Retrieved May*, 13:2006, 2004.

[2] A. J. Astell, N. Alm, G. Gowans, M. Ellis, R. Dye, and P. Vaughan. Developing an engaging multimedia activity system for people with dementia. In *International Workshop on Cognitive Prostheses and Assisted Communication (CPAC 2006)*, page 16, 2006.

[3] J. Bennett and S. Lanning. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35, 2007.

[4] J. E. Birren and D. E. Deutchman. *Guiding autobiography groups for older adults.* Johns Hopkins Press, 1991.

[5] E. Bohlmeijer, M. Roemer, P. Cuijpers, and F. Smit. The effects of reminiscence on psychological well-being

in older adults: A meta-analysis. *Aging and Mental Health*, 11(3):291–300, 2007.

[6] C. Bowlby. *Therapeutic activities with persons disabled by Alzheimer's disease and related disorders*. Aspen Publishers Baltimore, 1993.

[7] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.

[8] D. Bridge, M. H. Göker, L. McGinty, and B. Smyth. Case-based recommender systems. *The Knowledge Engineering Review*, 20(03):315–320, 2005.

[9] D. Brooker. What is person-centred care in dementia? *Reviews in clinical gerontology*, 13(3):215–222, 2004.

[10] Y.-L. Chen, L.-C. Cheng, and C.-N. Chuang. A group recommendation system with consideration of interactions among group members. *Expert systems with applications*, 34(3):2082–2090, 2008.

[11] M. A. Conway. *Flashbulb memories*. Lawrence Erlbaum Associates, Inc, 1995.

[12] M. A. Conway, A. F. Collins, S. E. Gathercole, and S. J. Anderson. Recollections of true and false autobiographical memories. *Journal of Experimental Psychology: General*, 125(1):69, 1996.

[13] M. Cotelli, R. Manenti, and O. Zanetti. Reminiscence therapy in dementia: A review. *Maturitas*, 2012.

[14] D. Harley and G. Fitzpatrick. Youtube and intergenerational communication: the case of geriatric1927. *Universal access in the information society*, 8(1):5–20, 2009.

[15] J. L. Harris. Reminiscence: A culturally and developmentally appropriate language intervention for older adults. *American Journal of Speech-Language Pathology*, 6(3):19, 1997.

[16] J. L. Harris and A. F. R. Plan. Signatures: Speaking up about memories. 2012.

[17] A. Jameson and B. Smyth. Recommendation to groups. In *The adaptive web*, pages 596–627. Springer, 2007.

[18] T. Kitwood, K. Bredin, et al. Towards a theory of dementia care: personhood and well-being. *Ageing and society*, 12(3):269–287, 1992.

[19] J. A. Konstan and J. Riedl. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2):101–123, 2012.

[20] N. Kuwahara, K. Kuwabara, and S. Abe. Networked reminiscence content authoring and delivery for elderly people with dementia. In *Proceedings of International Workshop on Cognitive Prostheses and Assisted Communication*, pages 20–25, 2006.

[21] P. Lops, M. de Gemmis, and G. Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*, pages 73–105. Springer, 2011.

[22] F. Lorenzi and F. Ricci. Case-based recommender systems: A unifying view. In *Intelligent Techniques for Web Personalization*, pages 89–113. Springer, 2005.

[23] J. Masthoff. Group recommender systems: Combining individual models. In *Recommender Systems Handbook*, pages 677–702. Springer, 2011.

[24] K. McCarthy, L. McGinty, B. Smyth, and M. Salamó. The needs of the many: A case-based group recommender system. In *Advances in Case-Based Reasoning*, pages 196–210. Springer, 2006.

[25] K. McCarthy, M. Salamó, L. Coyle, L. McGinty, B. Smyth, and P. Nixon. Cats: A synchronous approach to collaborative group recommendation. In *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference, Melbourne Beach, FL*, pages 86–91, 2006.

[26] K. McCarthy, M. Salamó, L. Coyle, L. McGinty, B. Smyth, and P. Nixon. Group recommender systems: a critiquing based approach. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 267–269. ACM, 2006.

[27] S. Morgan and B. G. University of Wales. *The Impact of a Structured Life Review Process on People Wih Memory Problems Living in Care Homes*. 2000.

[28] A. Norris. *Reminiscence: With Elderly People*. Winslow, 1986.

[29] M. O'Connor, D. Cosley, J. A. Konstan, and J. Riedl. Polylens: a recommender system for groups of users. In *ECSCW 2001*, pages 199–218. Springer, 2002.

[30] J. O'Rourke, F. Tobin, S. O'Callaghan, R. Sowman, and D. Collins. 'youtube': a useful tool for reminiscence therapy in dementia? *Age and ageing*, 40(6):742–744, 2011.

[31] J. C. Payne. Cultural competence in treatment of adults with cognitive and language disorders. 2011.

[32] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

[33] P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 157–164. ACM, 2011.

[34] D. L. Schacter. *Searching for memory: The brain, the mind, and the past*. Basic Books, 1996.

[35] G. Shani and A. Gunawardana. Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer, 2011.

[36] B. Smyth. Case-based recommendation. In *The adaptive web*, pages 342–376. Springer, 2007.

[37] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 109–116. ACM, 2011.

[38] H. L. Williams, M. A. Conway, and G. Cohen. 2 autobiographical memory. *Memory in the real world*, page 21, 2008.

[39] B. Woods, A. Spector, C. Jones, M. Orrell, and S. Davies. Reminiscence therapy for dementia. *Cochrane Database Syst Rev*, 2, 2005.