COHERENCE AND PERSONALIZATION OF SOCCER HIGHLIGHTS

C. Dolbear[1], J. M. Brady[1], J. Teh[2] and P. Hobson[2].

[1] Robotics Research Group, Department of Engineering Science, University of Oxford.
[2] Motorola UK Research Laboratory

ABSTRACT

We present an evaluation of the personalization accuracy of a system to automatically generate soccer highlights from a full length match according to the preferences of a soccer fan. This is measured objectively in three ways: firstly, how well the duration of the summary meets users' requests; secondly, via the introduction of a utility function to measure fulfillment of user requirements, and thirdly by measuring the coherence of the summary. We introduce the concept of a *context group* of soccer events for modelling the causal relationships between events, and show that including whole context groups in the summary, rather than just one event at a time, both increases coherence and mitigates the inherent reduction in coherence due to personalization.

1. INTRODUCTION

This paper describes a system to automatically generate soccer highlights from a full length match according to the preferences of a soccer fan. The massive interest in soccer makes commercialization of filtering techniques in this domain attractive. Manually edited soccer highlights are already being marketed as a major application for third generation mobile phones, but since substantial expertise and time is required to edit soccer highlights by hand, an advantage of an automatic system is to allow a user to receive personalized highlights. For instance, user requirements research [1] has shown that fans wish to see specific events involving particular players, and are very keen on viewing summaries of soccer matches when mobile, as the key activities are only a small part of the game. A knowledge elicitation study with sports editors [2] has shown that soccer highlights must tell the story of the match, and that flow of play and event causality, for example the events leading up to a goal, are very important elements of a good highlights package.

Young [3] notes that coherence comes from the selection of actions whose causal and temporal relationships highlight an underlying plot. User interaction, for example in automatic narrative generation tools, allowing the user to alter the state of the world at any given point in a story, can so radically alter the world that even the most accommodating plot lines cannot survive. This raises the question of how far we can personalize a summary before losing the sense of coherence. For example, including only events involving the soccer fan's favourite player may result in a meaningless sequence of disjoint events, providing no understanding to the viewer of what actually happened in the game. We need to make sure that personalization only takes place within a framework of coherent summarization, in order to avoid this problem.

Using our automatic soccer highlights generation system, we address the issue of how to measure users' satisfaction with the content they are provided with and how this relates to the length of summary they might pay for. We also investigate the trade-off between personalization and coherence in a summary, by developing a novel quantitative measure of summary coherence based on the causal relationships between events. We begin this paper by discussing previous work on soccer highlights generation and personalization of summaries. In section 3 we describe our personalization method for a soccer highlights application and report on results in section 4. Finally, conclusions are presented in section 5.

2. BACKGROUND

Research on multimedia summarization has mainly concentrated on key frame identification and video skim generation for browsing purposes, for example, [4]. Soccer highlights, however, are summaries presented as the final product for user consumption. The information extraction problem has been addressed in the soccer domain using audio and video features such as colour density analysis, slow motion replay detection, penalty-box detection and speech-band energy to identify semantic events using machine learning techniques such as Bayesian Belief Networks [5]. With such systems, *any* event that can be recognised is deemed important enough to include in the summary. This leaves the generation of more meaningful summaries, containing only events relevant to a particular user, as an open area of research.

It has long been recognized in natural language processing research that an accurate summary includes all the narrative elements of the original text [6], and the importance of a text unit depends directly on the number and quality of causal relations that the unit has to other text units [7]. More recently, narrative coherence, modelled using tree-depth measurement in rhetorical structure theory trees, has been the basis for sentence selection algorithms for text summarization [8]. While the authors have shown that modeling textual coherence improves summarization results, the causal relationships in all of these previous methods have been manually annotated. Our system is able to identify these causal relationships automatically, and we then use them to personalize the summaries.

From interviews with soccer fans asking them to rank events in priority order [1], we have a clear idea of user preferences in the soccer domain. Goals were found to be the most important, followed by major referee decisions, sendings off, fouls, the build up to and celebrations following a goal, interviews with goal scorers or man of the match, and finally controversial incidents. This insight is used in the design of our user profile ontology in section 3.

A frequently used method for personalizing a multimedia summary, for example [9], is to assign a weight to each of the user's preferences, and use these weightings to vary the scores of the multimedia content entities, so that a resource allocation agent can then determine which content should be included in the personalized summary. An alternative is to use a collaborative filtering technique [10]. These are mainly employed in recommender systems providing personalized suggestions about items that a user may find interesting. Neighbourhoods of users with similar tastes (specified via user profiles) are formed and used to generate recommendations of items that a particular user may be interested in. Neither of these personalization approaches allow for the *coherent* combination of a number of items into a summary, and so are insufficient for our purposes.

3. METHOD

3.1. Soccer ontology and 'neutral' summarization

Since our primary focus is on information summarization rather than extraction, we sidestep the need to extract information from the audio or video representation of soccer matches, and use the minute-by-minute "ticker-tape" reports widely available on many sports' websites.
While work has been done on information extraction from free-text soccer reports [11] we avoid this complexity by using a template mining technique to extract information directly from text where there is an automatically recognizable pattern. For example on websites such as the BBC's, the number of event classes that are described is limited, so we can simply search for expected words and phrases such as "Goal", "by" (followed by a player's name) and "from left half". We use a soccer ontology containing 20 classes representing common soccer events such as Assist, Booking, Corner, Foul, and Goal. Each event class has a *start time*, *extra time, duration* and *player* property associated with it. Our training set consists of 126 examples of full length soccer match descriptions and their corresponding summaries, generated by manual annotation of those events shown in the highlights broadcast on television, which we use as a "ground truth" benchmark. Events are clustered into causally-related groups (which we term *context groups*). This is either done using the groupings chosen by the editor of the original web page (when the events are grouped into paragraphs on the web page) or they are clustered using a Markov chain to estimate the joint probability of those events occurring as a group. The transition probability matrix of the Markov chain is estimated by counting the frequency of occurrence of each event class and pairs of event classes. For example, P(Booking|Foul) is high, as many Fouls are followed by Bookings in the training set. To cluster events in the test problem into a group, events are added to the group in turn, and the joint probability of the group's occurrence is calculated. The group is terminated when this joint probability falls below a certain threshold. The relative priority of each context group is the probability of a context group being included in the summary, given that it has occurred in the full length sequence. This is estimated using a second Markov chain, whose transition probability matrix derives from the frequency of occurrence of events in the summaries of the training set (rather than full length sequences, as for the context group clustering). By introducing the concept of these context groups, we can generate a summary not consisting solely of disjoint, unrelated events, but which makes sense as a whole, and explains to the viewer, for example, what caused a player to be sent off the pitch, or how a goal came about.

3.2. Personalized summaries and user profile design

Figure 1 shows our user profile ontology, along with two instances, representing example users *Simon* and *Sarah*.

User profile property	Instance 1	Instance 2
Name	Simon	Sarah
Summary length	60 seconds	5 minutes
Favourite club	Manchester City	Everton
Secondary clubs	-	Arsenal
Favourite player(s)	David Seaman, Nicolas Anelka	Wayne Rooney, Thierry Henry
Favourite event	Goal	Goal
Second favourite event	Sending off	Penalty
Third favourite event	Foul	Shot
Fourth favourite event	Penalty	Save
Fifth favourite event	Booking	Assist

Figure 1. Properties of the user profile class, along with values of two instances used in the personalization experiments.

The property values in these user profiles were chosen to reflect two users, one more interested in controversial events, and the other in skill and goal-related events. We bias the summaries for the two different users towards different narrative episodes by changing the priority of different context groups, based on their content. Since we know the duration of each event, we can limit the duration of the summary to the user's preferred length by choosing only the highest priority context groups until the required duration is reached.

4. RESULTS

4.1. Duration accuracy

The first question to answer is whether our system can produce summaries of the right length for different users. We varied the requested length of the summary between 30 seconds and 20 minutes. Then the duration error between this request and the actual summary output was measured for the 126 soccer matches, in a leave-one-out fashion. The experiment was first carried out using summaries based on the priorities of individual events, and then repeated using summaries where whole context groups of events were included at a time, so that the differences in the two methods could be evaluated. Figure 2 shows a graph of the mean percentage error in duration accuracy for different summary lengths, and it can be seen that single-event based summarization is more accurate than context-group based summarization, especially for shorter summaries since an event's duration is of finer granularity than a context group's. The mean duration of a single event is 19.2 seconds, compared with 27.1 seconds for a context-group. However, beyond about 300 seconds there is little advantage in using single-event based summaries, in terms of duration accuracy, and the advantage of context-group based summaries is in the additional coherence they provide to the overall summary, measured quantitatively in section 4.3.

4.2 Utility

To offer the user choices like, "We know you're an Everton supporter, would you pay for an extra five minutes to see Everton scoring from a penalty?" we introduce a measure of how well the content presented to the user fulfills their requirements, which we term *utility*. Our utility function for a summary *S* and user profile *U* is defined as:

$$Utility\ (S,U) = \sum_{i=0}^{N} w_i * frequency\ (events\quad of\quad class\quad U(i) \in S)$$

where *i* is the index of user profile properties, and *N* is the number of properties in the user profile.

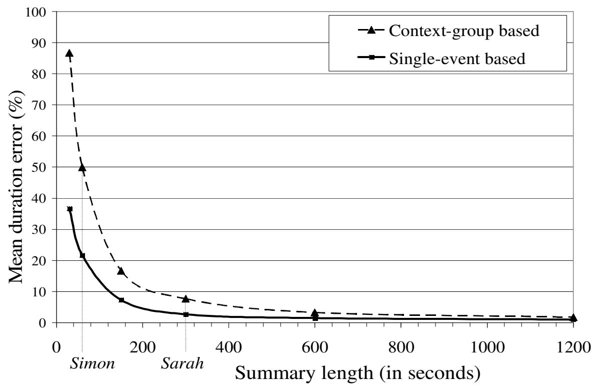

Figure 2. Personalized summary duration error against summary length, with the two users *Simon* and *Sarah*'s preferred summary lengths marked.

The weightings w_i give higher priority to the more preferred events, and those involving a favourite player or club. Figure 3 shows how utility increases with summary length; for *Sarah* the rate of increase decreases with summary length, while for *Simon* it increases. *Simon* is a tougher customer to please than *Sarah*, although this difference is less noticeable at shorter summary lengths. This is because *Sarah'*s favourite events are included more often in the neutral (non-personalized) summaries than Simon's favourites.

4.3 Coherence

We now investigate the trade-off between coherence and personalization. To what extent is our suggestion valid that constraining personalization to the context group level improves coherence? Our coherence measure for a summary *S*, consisting of events E_t, E_{t-1},... E_1 is calculated as:

$$Coherence(S) = \sqrt[t]{P(E_t, E_{t-1} ... E_2 \mid E_1)}$$

That is, coherence is based on the causal relationships between the summary events, as calculated using the conditional probability of occurrence of the sequence, given the first event. Coherence is calculated here for summaries of the same length as the ground truth summaries, because coherence was found to decrease with summary length. Therefore, we do not vary the summary length for Simon and Sarah's preferences in this experiment in order to make useful comparisons of their coherence. Figure 4 shows the mean coherence of the summaries broadcast on television (the "ground truth"), compared with our neutral summaries generated using both single-event based and context-group based summarization; and summaries personalized for *Simon* and *Sarah.*

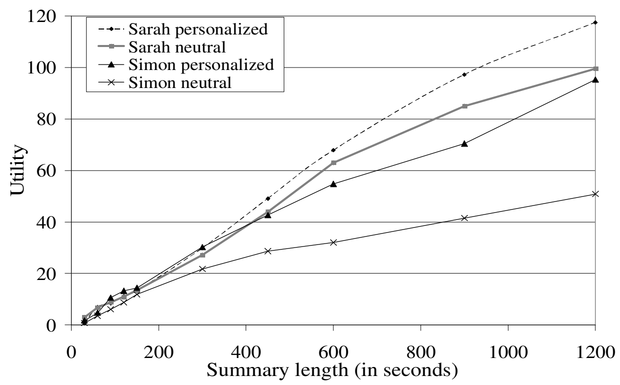

Figure 3. Personalized summary utility against summary length

Figure 4 shows that coherence is much higher for context-group based summarization than when single events are included in the summaries. The small difference in *Simon* and *Sarah*'s results can be attributed to the small variations in summary lengths. As suggested in [3], personalization reduces coherence in the summary, but the drop is much smaller for the context-group based summaries than the single-event based ones, which shows the advantages of the context group idea in retaining coherence, even in a personalized summary.

5. CONCLUSIONS

In the context of a system for automatically generating personalized soccer highlights, we have looked at how accurately we can generate a summary of the user's preferred duration: to our knowledge, the first study of its kind. We found that while single-event based summarization has a smaller duration error than context-group based summarization, since an event's duration is of finer granularity than a context group's, this advantage decreases significantly with summary length. The mean percentage error between the actual and preferred summary length also decreases as the summary length increases.

To entice a user to pay for extra content, or help them save time, we have developed a utility measure to quantify the additional benefit a particular user would gain from an increment in summary length. We found that utility increases with summary length, and that some users have higher utility than others, even for the neutral summaries. Finally, we investigated how our use of context groups of causally related events contributed to summary coherence. We found that including whole context groups in the summary, rather than just one event at a time, not only increased coherence, but mitigated the reduction in coherence due to personalization.

Experiment	Coherence
Ground truth	0.112
Neutral, single event based	0.018
Simon, event-based, personalized	0.006
Sarah, event-based, personalized	0.009
Neutral, context-group based	0.117
Simon, context-group based, personalized	0.112
Sarah, context-group based, personalized	0.113

Figure 4: Mean coherence of various summaries; comparing ground-truth with neutral and personalized summaries

6 ACKNOWLEDGEMENTS

C Dolbear is the recipient of a Motorola University Partnerships in Research Grant.

7 REFERENCES

[1] A. Evans, “User-Centred Design of a Mobile Football Video Database,” *Second International Conference on Mobile and Ubiquitous Multimedia*, Norrkoping, Sweden, Dec 2003
[2] C. Dolbear, and J. M. Brady “Soccer Highlights Generation using A Priori Semantic Knowledge” *IEE International Conference on Visual Information Engineering*, VIE 2003, University of Surrey, UK. July 2003.
[3] R. M. Young “Creating Interactive Narrative Structures: The Potential for AI Approaches” *AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment* AAAI, Stanford, CA, March 2000.
[4] H. Sundaram and S. F. Chang “Condensing Computable Scenes using Visual Complexity and Film Syntax Analysis”. *Proceedings of the International Conference on Multimedia and Expo, ICME'01.* IEEE Tokyo, Japan, August 2001.
[5] A. Ekin, A. M. Tekalp and R. Mehrotra “Automatic Soccer Video Analysis and Summarization”, *IEEE Transactions on Image Processing*, vol 12, no 7, 2003.
[6] W. G. Lehnert “Plot Units: A Narrative Summarization Strategy” *Cognitive Science* vol 4, pp 293-331 1981
[7] T. Trabasso and L. L. Sperry “Causal relatedness and importance of story events” *Journal of Memory and Langugage* vol 24 pp 595-61 1985
[8] I. Mani, E. Bloedorn, and B. Gates. “Using cohesion and coherence models for text summarization” *Proceedings of the AAAI Spring Symposium on Intelligence Text Summarization*, pp 69-76 1998.
[9] A. M. Ferman, J. H. Errico, P. Van Beek and M. I Sezan “Content-based filtering and personalization using structured metadata” *Proceedings of the Joint Conference on Digital Libraries* p393, ACM Portland, Oregon. July 2002.
[10] U. Shardanand and P. Maes “Social Information Filtering: Algorithms for Automating "Word of Mouth"” Proceedings of Computer Human Interaction CHI'95 p 210-217 1995
[11] H. Saggion, H. Cunningham, D. Maynard, K. Bontcheva, O. Hamza, C. Ursu. and Y. Wilks “Extracting Information for Information Indexing of Multimedia Material” *Third Language Resources and Evaluation Conference* 2002