Raconteur: From Intent to Stories

Pei-Yu Chi, Henry Lieberman
MIT Media Laboratory
20 Ames St., Cambridge, MA, USA
{peggychi, lieber}@media.mit.edu

ABSTRACT

When editing a story from a large collection of media, such as photos and video clips captured from daily life, it is not always easy to understand how particular scenes fit into the intent for the overall story. Especially for novice editors, there is often a lack of coherent connections between scenes, making it difficult for the viewers to follow the story.

In this paper, we present *Raconteur*, a story editing system that helps users assemble coherent stories from media elements, each annotated with a sentence or two in unrestricted natural language. It uses a Commonsense knowledge base, and the AnalogySpace Commonsense reasoning technique. Raconteur focuses on finding *story analogies* – different elements illustrating the same overall "point", or independent stories exhibiting similar narrative structures.

Author Keywords

Storytelling, media editing, story goal, story analogy, commonsense computing, video, photograph.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Human Factors

INTRODUCTION

When presenting or editing a large set of material, such as photos and video clips captured from daily life, it is not easy to understand how particular scenes fit into the overall story. Most people therefore choose to present the story by events in chronological order [6], or by location or characters. Although there is software for automating categorization or suggesting keyword tags, it is still challenging to create a coherent higher-level presentation that tells an entertaining story. Novice users often do not pay attention to the "point" being made by showing a given scene, or provide meaningful connections between scenes, making it difficult for the viewers to follow the story. We believe that an intelligent interface that provides assistance in relating the concrete elements of the scene to the overall story intent, will result in more effective story composition.

We present *Raconteur*, a story editing system that helps users think about material in a story, by showing related scenes or other stories with similar goals. The word "raconteur", by definition, is a person who is skilled in relating stories and anecdotes meaningfully. Similarly, our system tries to understand the narrative goal presented by the user, and find analogous story elements related to the goal.

We aim to create a system that helps a user tell a story by selecting a sequence of media items from a corpus of video, stills, narrations, and other media. We are assuming that each media element is annotated with a sentence or two in unrestricted natural language. The annotation may describe people, events, actions, and intent of the scene. Some annotations may be generated by metadata, transcription of audio, or other means.

A user at first presents a story goal in natural language, for example, "a 4-hour biking challenge on Cape Cod". The objective of the system is to provide a selection of possible matches of the annotated media elements to the story goal, that best help to tell the story. Note that this objective is *not* the same as simple search, keyword matching, subject relevance, or other conventional retrieval problems.

The tools we use for doing this are a large Commonsense knowledge base, Open Mind Common Sense; state-of-the art natural language parsing; and our own unique AnalogySpace inference technique for analogical reasoning. We can only provide a brief description of the knowledge base and inference in this paper; we refer the reader to the references for more detailed explanation of the tools.

For the "biking challenge" example, the system is able to suggest video clips that support relevant themes, such as anticipation and worries, preparation, difficulties, and, finally, the result of success or failure of the trip.

Analogous Story Thinking

We are inspired by how humans understand stories using analogies, which are partial similarities between different situations that support further inferences [4]. Schank proposed the idea of "story skeleton" to explain how we construct and comprehend a story under a certain structure to communicate with each other [11].

IUI'10, February 7–10, 2010, Hong Kong, China.

For example, a camping, hiking, and biking trip may include challenges (putting a tent up, going through a long and steep path), difficulties or worries (cannot assemble the tent poles, unable to finish the path, get lost on the way, bad weather), and enjoyable experiences (learning the setup, arriving a new place, meeting new friends) of a physical activity with a group of people. Similar points may be presented again and again through the process of developing a story. Superficially different events may illustrate analogous themes, so the ability to make analogies helps tell a story in a coherent way to the audience.

FORMATIVE USER STUDY

Before we present the Raconteur system itself, we describe a formative study to see if users would find presentation of analogous story elements helpful in story construction. In [12], we presented a more extensive user study reporting concrete experience with a previously implemented system for story construction via Commonsense knowledge. The present study is concerned with the value of the new analogical inference. We designed a story-editing interface that shows both the raw set of selected material and the analogous elements we found.

Experimental setup

We asked one participant to collect media (photos or videos) documenting her life experiences for three months. We observed the 30 collected stories, selected and compared similar topics, and specified the key shots in each story.

We summarize three main categories: 1) stories with a clear procedure as a story pattern, e.g. birthday parties that people sing the "Happy Birthday" song, make wishes, cut the cake, etc.; 2) Stories without a clear procedure but with certain expected events; 3) Stories without a clear procedure and expected events, e.g. a camping, hiking, and biking trips that include difficult challenges and new experience of an activity. We chose one story from each of these categories as test cases, including "Hsien's birthday party with a potluck dinner", "Mike's commencement party for his first master's degree", and "A 4-hour biking challenge in Cape Cod". The collector annotated each media element with a sentence or two in English.

5 participants were invited, including 3 males and 2 females, aged between 20-30, experienced with digital media. Participants were asked to edit stories for sharing with their friends. The facilitator first helped them familiarize themselves with the test cases, and then introduced our editing interface and conducted the 3 editing sessions.

Results

We found that when the size of the corpus was large and the story was relatively complex (test case 1&3), presenting the analogous story helped users follow a story pattern better. Especially for test case 3 (a biking trip), participants found the story complex for them to edit, and reported the analogous examples helped them to design the story development. Most participants spent considerable time on observing the similarity of story content. One participant said, "It was interesting to see how the system presented a new perspective to the story I wanted to tell"; another explained, "The system helped me rethink the similarity and differences between experiences, which I would rarely think to do from just browsing a bunch of files." These findings encouraged us that the analogical reasoning mechanism would prove useful to users in story construction.

DESIGN OF THE RACONTEUR SYSTEM

In this section we present a concrete example of a story of a vacation trip from the collected story base, along with the description of our knowledge representation and inference methods in Raconteur, and the user interface.

Finding Elements Analogous to the Story Goal

A user starts by inputting a story goal, in unrestricted natural language. A simple example is, "*a 4-hour biking challenge on Cape Cod*". First of all, we need to determine the possible patterns to present the goal. Table 1 lists sample narrative goals and media annotations that Raconteur finds for this Cape Cod biking trip example.

Both goal statements and media item annotations are processed using conventional natural language tools such as part of speech tagging. The result is related to knowledge stored in our Open Mind Common Sense knowledge base, and the ConceptNet semantic network [9]. Assertions from that knowledge base relevant to the concept "challenge", after excluding anything that has no connection to "biking" include,

Anticipations and Worries
• "Cape Cod, a peninsula with stunning scenes, is famous for vacation and outdoor activities such as biking." (A1) • "Before this, I had only experienced long biking trip once. When I promised to take this challenge, I was a little bit nervous and afraid that I couldn't finish the trip." (A2)
Preparation
• "To start a day, we need energy; so first, we went to have a luxurious brunch." (P1) • "It is important to rent a good bike for the challenge. We came to the rent shop and pick our own bikes." (P1) • "Then, we biked all the way to the end of the rail trail. This was about six miles, the most difficult part." (P2)
Difficulties
• "It took us a while to find the correct way between the branches. Thanks to the map, or we would get lost." (D1) • "It was a really long trail… I almost wanted to give up on the half way, especially I had no idea about how long I had biked." (D2)
Results: Successes or Failures
• "As you can see, we were really thrilled when we arrived the beach. Although I already felt one mile was long!" (R1) • "Surprisingly, it was easier than what I imagined. We were so excited when we arrived the destination." (R2)

Table 1. Multiple story units with similar patterns that Raconteur finds. The number indicates the elements in the same set of patterns.

- Desires (challenge, anticipate something)
- MotivatedByGoal (challenge, test oneself)
- HasProperty (challenge, difficulty)
- Causes (challenge, success)
- Causes (challenge, failure)

One pattern for presenting a challenge involves "*anticipations and worries*", "*preparation*", "*difficulties*" and "*results*" (successes or failures). When Raconteur finds more than one pattern to present the goal, multiple patterns can be considered.

Analogical Inference in Raconteur

Second, Raconteur analyzes the annotation of each media element in natural language, and makes analogical inferences. It uses the AnalogySpace Commonsense reasoning technique [13].

AnalogySpace represents the entire space of OMCS's knowledge through a sparse matrix whose rows are ConceptNet concepts (noun phrases and verb phrases), and whose columns are *features*, one-argument predicates that can be applied to those concepts. A feature generally consists of one of 20 or so two-place relations (kind_of, part_of, etc.) together with another concept. Inference is performed by Principal Component Analysis on this matrix, using the Singular Value Decomposition. This transforms the space by finding those axes that best account for the variation in the matrix. These axes are often semantically meaningful. The reason this is good for computing analogy is that concepts that have similar Commonsense assertions true about them wind up close to each other in the transformed space. Unlike logical approaches to analogy, it is computationally efficient, and tolerant of vagueness, noise, redundancy, and contradiction.

For example, A1 in Table 1 with descriptions of "*Cape Cod*", "*stunning*", "*famous*", "*vacation*", and "*biking*" infers this piece of material indicates the user's anticipation of the trip; P1 infers the preparation including having brunch, and renting a bike; D1 explains the difficulty of finding the way to avoid getting lost; then R1 shows the excitement of the arrival.

Finally, by finding the elements related to the story goal, we provide different perspectives on telling a story within a set of material. If there is more than one generated pattern that matches the story goal, Raconteur will go through this process to different patterns and present one that contains the most analogous elements in the set.

Raconteur User Interface

The final results of analogous story elements will be shown through the web-based user interface as Figure 1. A user can see the unorganized, sequential material in chronological order as in Figure 1a. The user decides a story goal in English, and then the analogous elements will be shown, as in Figure 1b. The user can drag and drop photos or video clips as desired to create a story.

Discussion

We'd also like to discuss how Raconteur might enable new kinds of storytelling activities. Our model encourages users to think about the story goal instead of directly composing individual elements. To help participants reason how the

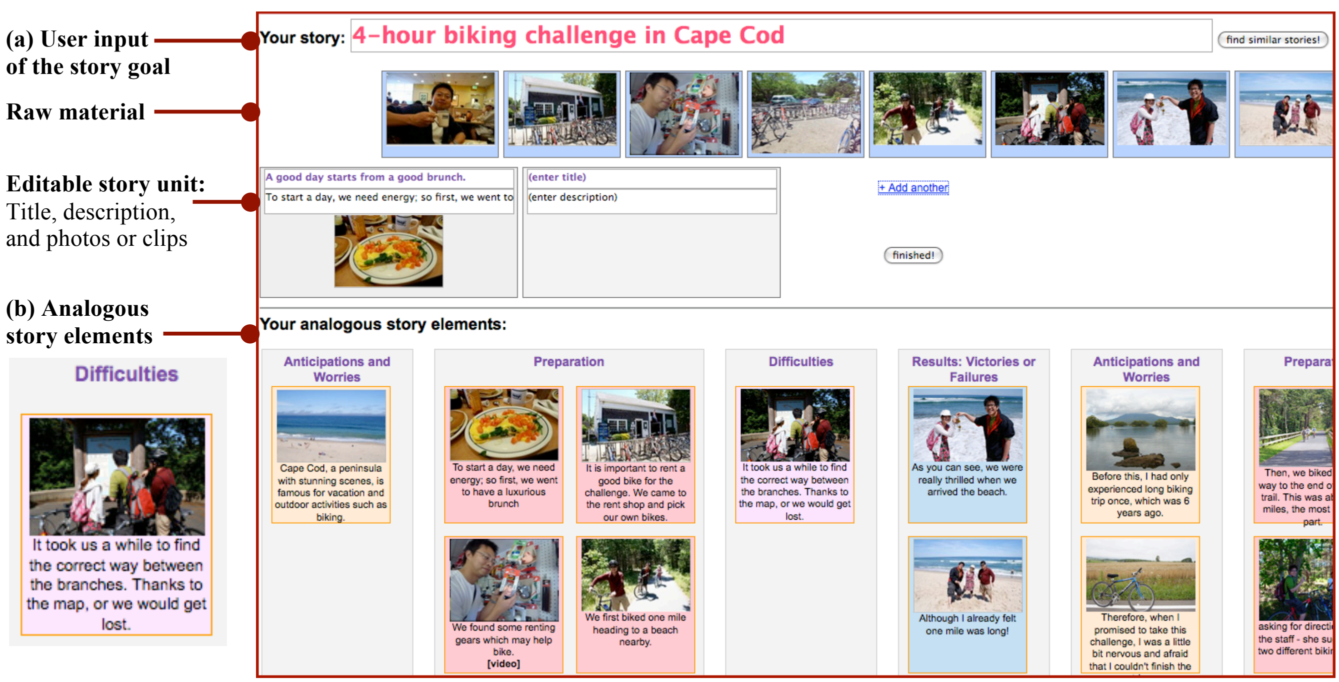

Figure 1. Raconteur interface: (a) the upper part presents the raw material of the unorganized story, and provides the editing interface for users to decide the story goal and the sequence of scenes; (b) the lower part shows the sets of analogous story elements in a pattern that matches the story goal.

pattern and the results were generated, we highlight the keywords that relate to system inferences, allowing users to revise the suggested pattern and explore the space of results. Furthermore, to make the storytelling process closer to the experience of daily conversation, we are also considering integrating our technique in a chat scenario. The storyteller can interact with a partner to talk about the stories, and Raconteur can suggest both story topics and media elements.

RELATED WORK

Cooper *et. al.* designed a similarity-based analysis to cluster photos by timestamps and content [3]. Joshi and Luo presented a method to infer events and activities from geographical information [5]. However, most of the research work on automatic media organization focuses on analyzing the basic attributes such as time and location; few of them consider the overall story development and story thinking with digital media.

Another emerging research area is to interact with digital media on the level of story composition. ARIA is a software agent that dynamically retrieves related photos based on the content of an email or web page [8]. Barry presented a system that presents contextual information during the process of video capture [1]. The closest system to the present one is Storied Navigation [12], which shares the goal of composing stories from annotated media clips. Our work here differs in the use of the analogy inference technique, and is focused on instantiating narrative goals directly through analogical inference to create better story structures.

We are also aware of several relevant narrative systems such as the storytelling and planning system "Universe," which models the story structure as a set of hierarchical plans and generates plot outlines based on the author's story goal [7]; Riedl and León's story analogous generation system is able to transform existing stories to a novel context [10]; Cheong *et. al.* presents an authoring interactive narrative framework to help users construct branching story structure [2]. Although rare of these projects incorporate digital media as our goal, they provide insights of story analysis to our work.

CONCLUSION AND FUTURE WORK

We have presented Raconteur, a story editing system that helps users think about material in a story by showing related scenes or other stories with similar goals. We suggest that presenting analogous stories can provide a helpful guideline for users to tell their stories. Our formative user study shows that this kind of analogy finding is particularly helpful in the case where users have large libraries or complex stories. Future work will focus on increasing relevance of suggestions, improving interactivity of media selection and output previews, and conducting detailed evaluation. We also are exploring augmenting the media capture experience as well as post-production editing. We aim for providing a fun and productive environment for storytelling. Maybe it will help your friends become more interested in watching your vacation movies, after all.

REFERENCES

1. Barry, B. and Davenport, G. Documenting Life: Videography and Common Sense. In *Proc. ICME2003*: the 2003 IEEE Intl. Conf. on Multimedia and Expo, IEEE Press (2003), Baltimore, MD, USA.
2. Cheong, Y., Kim, Y. Min, W. and Shim, E. PRISM: A Framework for Authoring Interactive Narratives. In *Proc. ICIDS 2008*: the 1st Joint Intl. Conf. on Interactive Digital Storytelling: Interactive Storytelling, ACM Press (2008), Erfurt, German.
3. Cooper, M., Foote, J., Girgensohn, A., and Wilcox, L. Temporal Event Clustering for Digital Photo Collections. In *Proc. MM'03*: the 11th ACM Intl. Conf. on Multimedia, ACM Press (2003), Berkeley, CA, USA.
4. Gentner, D. Analogy. In *A Companion to Cognitive Science*, Oxford University Press (1998), pp. 107-113.
5. Joshi, D. and Luo, J. Inferring Generic Activities and Events from Image Content and Bags of Geo-tags. In *Proc. CIVR'08*: the 2008 Intl. Conf. on Content-based image and video retrieval, ACM Press (2008), Niagara Falls, Canada.
6. Kirk, D., Sellen, A., Rother, C., and Wood, K. Understanding Photowork. In *Proc. CHI2006*: the 24th Intl. Conf. on Human factors in computing systems, ACM Press (2006), Montréal, Québec, Canada.
7. Lebowitz, M. Story Telling as Planning and Learning. *Poetics 14* (1985), pp. 483-502.
8. Lieberman, H. and Liu, H. Adaptive Linking Between Text and Photos using Common Sense Reasoning. In *Proc. AH2002*: the 2nd Intl. Conf. on Adaptive hypermedia and adaptive web-based systems, ACM Press (2002), London, UK.
9. Liu, H. and Singh, P. ConceptNet: a Practical Commonsense Reasoning Toolkit. In *BT Technology Journal* (2004), 22, 4, 211-226.
10. Riedl, M. and León, C. Generating Story Analogues. In *Proc. AIIDE09*: the 5th Conf. on Artificial Intelligence for Interactive Digital Entertainment, AAAI Press (2009), Palo Alto, CA, USA.
11. Schank, R. Tell Me a Story: A New Look at Real and Artificial Intelligence. Northwestern University Press (1991).
12. Shen, Y.-T., Lieberman, H., and Davenport G. What's Next? Emergent Storytelling from Video Collections. In *Proc. CHI2009*: the 27th Intl. Conf. on Human factors in computing systems, ACM Press (2009), Boston, MA, USA.
13. Speer, R., Havasi, C., and Lieberman, H. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge. In *Proc. AAAI2008*: the 23rd AAAI Conf. on Artificial intelligence, AAAI Press (2008), Chicago, IL, USA.