Generating Narratives from Personal Digital Data: Using Sentiment, Themes, and Named Entities to Construct Stories

Elaine Farrow[1]([⊠]), Thomas Dickinson[2], and Matthew P. Aylett[1]

[1] School of Informatics, University of Edinburgh, Edinburgh, UK
Elaine.Farrow@ed.ac.uk, matthewa@inf.ed.ac.uk
[2] Knowledge Media Institute, The Open University, Milton Keynes, UK
thomas.dickinson@open.ac.uk

Abstract. As the quantity and variety of personal digital data shared on social media continues to grow, how can users make sense of it? There is growing interest among HCI researchers in using narrative techniques to support interpretation and understanding. This work describes our prototype application, ReelOut, which uses narrative techniques to allow users to understand their data as more than just a database. The online service extracts data from multiple social media sources and augments it with semantic information such as sentiment, themes, and named entities. The interactive editor automatically constructs a story by using unit selection to fit data units to a simple narrative structure. It allows the user to change the story interactively by rejecting certain units or selecting a new narrative target. Finally, images from the story can be exported as a video clip or a collage.

Keywords: Social media · Narrative · Triptych · Multi-media

1   Introduction

Social media use has led to such an explosion in personal digital data that users can easily become lost and overwhelmed, and a common challenge for HCI researchers is to help users cope. Selection and summarization tasks (such as identifying a list of hashtags which are currently popular or reporting which Facebook posts got the most 'likes') represent a typical approach to improving understanding – treating it as a database to be queried.

In contrast, we often interpret our own experiences, desires, and motivations using stories not statistics. Both in the commercial world and in academia, interest is growing in bridging the gap between viewing personal digital data as a database, and viewing the data as (elements of) personal narratives. The use of narrative to make sense of our everyday lives is considered to be a fundamental human behavior [2], and the exploration of narrative formats for data presentation can help HCI researchers understand how meaning is constructed through stories and how stories can be used to interpret data. There have been some recent popular examples of automatic narrative generation

from social media data. For example, Facebook's *A Look Back*[1] compiles users' most popular posts into a short film. However, as noted in previous research [3–5], the resulting film is very much a finished product. Users can replace chosen posts with others from a limited selection, but have very little overall control.

The ReelOut application is part of a wider project in which we seek to build a novel text-driven software system that can automatically generate film-like life documentaries from personal digital data, and to explore the human experience and response to such systems. We aim to empower users by allowing them to interpret their data to suit their own vision of the narratives they see within their lives. ReelOut allows users to build stories from primitive units (such as tweets and Facebook posts) with reference to a particular narrative target: as a simple example for illustration, consider a story on the theme of 'food and drink' that starts with a negative sentiment and ends with a positive sentiment.

Our process is illustrated in Fig. 1. We extract text and metadata from a user's social media posts and augment it with semantic tags. Next, we employ a unit selection process borrowed from speech synthesis to fit the data to the desired narrative target. The user can change the story by rejecting certain units or selecting a new narrative target. Finally, the images associated with the generated story can be exported as a short movie clip or a collage of images.
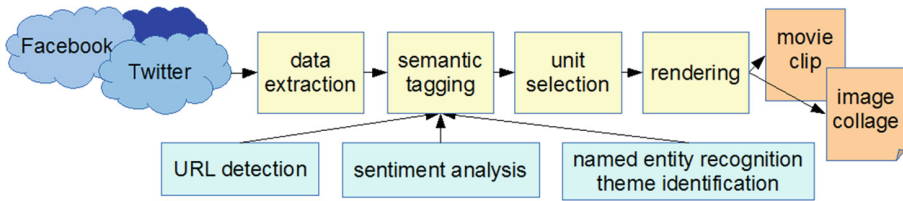

Fig. 1. The processing pipeline

2 Generating Narratives from Personal Digital Data

At the time of writing, our online service can extract data from Twitter and Facebook, with limited support for Instagram. We intend to add further platforms to this list in the near future. The sentiment of each post is calculated using Sentistrength,[2] a popular sentiment analysis tool for short web texts. Entities such as locations are extracted using AlchemyAPI's[3] entity extraction endpoint, while themes are identified using AlchemyAPI's taxonomy endpoint. Some social media posts, such as comments and replies, form a conversational thread. For these, we note which other units come before and after them in the thread. Finally, the data is passed to the interactive editor as a series of RLUnits, an XML format consisting of marked-up text and semantic tags (Fig. 2).

---

[1] facebook.com/lookback.
[2] sentistrength.wlv.ac.uk.
[3] www.alchemyapi.com.

The interactive editor (Fig. 3) provides a graphical user interface for story creation. By default, generated stories are sequences of three linked units (a *triptych*), corresponding to the classic 3-act structure of setup, confrontation, and resolution. The generated picture sequence can be exported as a short movie clip or saved as a collage of images.

```
<rlunit id="15028" src="facebook" time="20091101:190400">
   <image>...</image>
   <messagetext>Happy after dinner face</messagetext>
   <sentiment>positive</sentiment>
   <theme>food and drink</theme>
</rlunit>
```

Fig. 2. An RLUnit representing a Facebook post with a positive sentiment and a theme of 'food and drink' (the image URL is omitted due to space constraints).
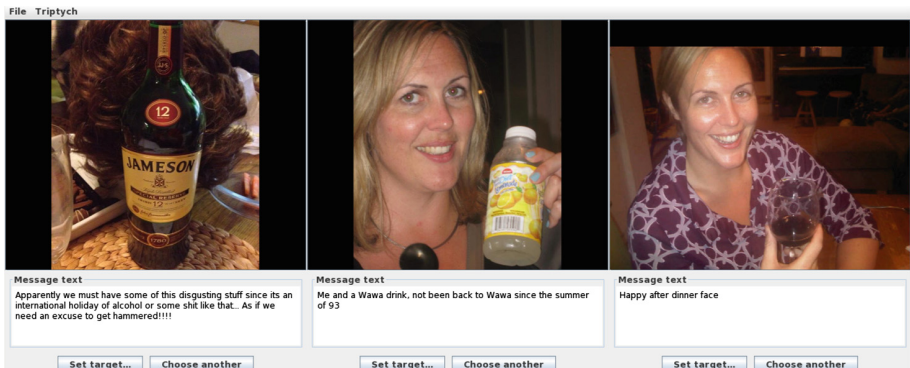
Fig. 3. The interactive editor, showing a story on the theme of 'food and drink' that starts with a negative sentiment and ends with a positive sentiment.

Our story generation algorithm is inspired by the unit selection process in speech synthesis, where many thousands of units are fitted to an utterance structure. We use dynamic programming to fit the marked-up data units (RLUnits) to a predefined narrative target, represented as an ordered collection of slots with associated semantic tags which constrain their contents.

Two cost functions are optimized to produce the output. The first, *target cost*, represents the fit with the narrative target: units which share semantic tags with the target slot will have a low target cost. The second, *join cost*, represents how connected two adjacent RLUnits are: units which have a similar set of semantic tags and appear in the correct order will have a low join cost.

The editor allows the user to set the target for each story slot interactively, using the semantic tags found in the extracted data. It also offers unit reselection – rejecting a particular unit and automatically selecting the next-best unit to fit that slot – a powerful method borrowed from speech synthesis, where it enables users to modify automatically synthesized utterances without requiring an understanding of the linguistic or phonetic structure of speech.

3   Conclusion and Future Work

Our novel end-to-end automatic narrative generation application, ReelOut, augments personal digital data from social media sites with semantic tags such as sentiment, themes, and named entities. It uses unit selection to build a story that fits a specified narrative target, and allows users to change the story interactively by rejecting particular units or selecting a new narrative target. It is extensible for further research and development.

We have run an initial evaluation of the story-generation algorithm by asking participants to look at picture sequences generated from public data (not their own), some generated by our system and others chosen at random [1]. We found that our system produced output which users rated significantly higher than random when asked "How much does this sequence of pictures tell a story?" We are currently working towards a user trial of the full system, allowing people to create and evaluate stories using their own personal digital data.

Our future plans include developing a data-driven event detection algorithm which will allow us to classify individual units as belonging to a larger event, such as 'starting school' or 'getting married'. We are also experimenting with ways to represent units which have text but no associated image, such as rendering the text itself as an image or sourcing a new image from an external source.

Our application demonstrates one important way HCI researchers can use narrative to help users to make sense of the growing mass of personal digital data which threatens to overwhelm them – by automatically constructing stories.

References

1. Aylett, M.P., Farrow, E., Pschetz, L., Dickinson, T.: Generating narratives from personal digital data: Triptychs. In: CHI EA 2015. ACM, New York, NY, USA, pp. 1875–1880 (2015). http://doi.acm.org/10.1145/2702613.2732702
2. Sarbin, T.R.: Narrative Psychology: the Storied Nature of Human Conduct. Praeger Publishers/Greenwood Publishing Group, Westport (1986)
3. Schwanda Sosik, V., Zhao, X., Cosley, D.: See friendship, sort of: how conversation and digital traces might support reflection on friendships. In: CSCW 2012. ACM, New York, pp. 1145–1154 (2012). http://doi.acm.org/10.1145/2145204.2145374

4. Thiry, E., Lindley, S., Banks, R., Regan, T.: Authoring personal histories: exploring the timeline as a framework for meaning making. In: CHI 2013. ACM, New York, NY, USA, pp. 1619–1628 (2013). http://doi.acm.org/10.1145/2470654.2466215
5. Zhao, X., Lindley, S.E.: Curation through use: understanding the personal value of social media. In: CHI 2014. ACM, New York, pp. 2431–2440 (2014). http://doi.acm.org/10.1145/2556288.2557291