July 18, 2019

Generating a digest memorable highlights from unstructured messaging conversation

Joseph Edwin Johnson Jr.

Generating a digest memorable highlights from unstructured messaging conversation

ABSTRACT

Many social media and photo collection applications offer features to generate a personalized digest of photos or structured text. A digest typically includes highlights of content from a larger collection. Such digest generation and viewing capabilities are not currently available for freeform unstructured conversational interactions, such as those that take place via messaging or chat applications. This disclosure describes techniques to automatically generate a digest of such interactions with permission from the conversing parties. The conversational interactions may contain a mixture of text, images, audio, and video. A user can review and save the digest offline for later personal use, share the digest with the other conversation parties, or post it on social media. The described techniques can be applied to direct one-to-one conversations as well as to multi-party group conversations.

KEYWORDS

- Messaging conversation
- Chat summary
- Conversation summary
- Conversation digest
- Conversation highlights
- Natural Language Processing (NLP)
- Unstructured text

BACKGROUND

Many social media and photo collection applications offer features to generate a personalized digest of photos or structured text, e.g., hashtags. A digest typically includes

highlights of content from a larger collection such as all the photos shared with another individual, photos liked by a subgroup of social media connections, photos at a given location, photos taken during a given time interval etc. Such a digest often serve as a means to relive memories. A digest can be constructed manually by the user or generated automatically. These digests can then be viewed via quick and enjoyable mechanisms, such as slideshows, short videos, etc. and can possibly be shared with other users.

Such digest generation and viewing capabilities are not currently available for freeform unstructured conversational interactions, such as those that take place via messaging or chat applications. While such interactions sometimes contain photos, videos, and audio, most conversational data includes freeform unstructured text.

DESCRIPTION

This disclosure describes techniques to automatically generate a digest of such interactions with permission from the conversing parties. The conversational interactions may contain a mixture of text, images, audio, and video. A user can review and save the digest offline for later personal use, share the digest with the other conversation parties, or post it on social media. The described techniques can be applied to direct one-to-one conversations as well as to multi-party group conversations.

Accessing conversation data, programmatically analyzing it to identify content suitable for a digest, and generating the digest are performed with specific user permission from participants in a conversation. Users are provided with options to selectively provide permission to access portions of a conversation, deny permission for certain conversations, or disable digest generation entirely. Further, users can select the parties that can generate and/or view a digest, e.g., whether the digest can be viewed by the user only, by other participants in a conversation,

or by other parties; whether the digest is shareable via social media; whether the digest is available for download; etc.

If conversational parties permit, the conversational digest is generated by applying summarization algorithms to the text portion within the conversation, which often forms the bulk of the conversational content. The summarization algorithms are applied to rank paragraphs, sentences, and phrases within the text and subsequently, to select a subset of the ranked text artifacts. For instance, the digest can be formed by concatenating *k* highest ranked artifacts. While presenting the artifacts within the digest, the temporal order of occurrence is usually the most appropriate. However, the digest formation algorithm can also use an alternate ordering of artifacts for increased coherence. For instance, within a specific time interval, artifacts connected to the same theme or topic can be grouped together. Similarly, artifacts can be grouped based on machine learned sentiment, if use of sentiment analysis techniques is permitted by the user.

With permission from the conversing parties, topics within the conversational text can be identified by applying natural language processing (NLP) algorithms used for topic clustering, such as Latent Dirichlet Allocation (LDA). Alternatively, topics can be identified by the application of deep learning techniques by processing the conversational text at the word and/or sentence levels followed by the use of clustering algorithms, such as k-means. Some deep learning algorithms, such as recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent units (GRUs), transducers, etc., can directly generate conversational summaries from text subsets grouped by time or topic.

Dictionaries can be used to assist in topic selection and clustering. For example, such dictionaries can indicate that ‘breakup’ and ‘separation’ are conceptual synonyms, ‘boyfriend’ may be a hypernym or hyponym of ‘lover,’ ‘wheels’ in informal conversations can be a

meronym of car, etc. Use of such relationships can aid in clustering similar themes and topics. For instance, words that are synonyms, hypernyms, or meronyms can be replaced with a common key and grouped within the same cluster.

In addition to binary classification of text as positive or negative, further analysis can be carried out to detect specific emotional connotations, with permission from the parties involved in the conversation. The detected emotions can be used to select and/or filter artifacts and topics included in the digest. For example, users may benefit from a digest that memorizes grievance, such as the loss of a loved one, despite the unhappy nature of the topic. Other topics such as a relationship breakup or financial discussions, can be excluded from digest generation. Additionally, if the conversation parties permit, appropriate heuristics and/or machine learning techniques can be used to exclude certain types of content, e.g., inappropriate language or media.

If the parties in the conversation permit, images shared within the conversation can be analyzed and ranked in the digest generation process. The top ranked images can be included in the conversational digest along with the selected text artifacts. Further, the analysis can be performed to determine whether specific pieces of text within the conversation are associated with a given image. For instance, text occurring within a short time frame before or after an image is shared can be estimated to be related to the image. Alternatively, or in addition, potentially associated text and image pairs can be ranked by a machine learning algorithm trained on relevant corpus of text and image data, such as social media comments on photos. For example, it may be determined that long paragraph-length messages are less likely to be associated with an image than responses that are short. Moreover, text content that is likely to offer little information about the image, such as stop words and common phrases such as “Cool,” “Whoa,” etc. is ranked low and filtered out of consideration for the digest.

The application of the trained machine learning algorithm can be based on formulating a representation of the image such that comments in the corpus can be selected based on images most similar to the one within the conversation. Similarly, appropriate representations of the comments within the conversation as well as those in the data corpus can be used for comparing similarity and judging relevance.

When the generated digest includes an image from the conversation, one or more of the pieces of text within the conversation determined to be associated with the image can be included in the digest alongside the image. The process can also work in reverse; the inclusion a piece of text in the digest with permission from the conversing parties can result in one or more images determined to be associated with that piece of text being selected for the digest. In cases where only a subset of comments associated with an image selected for the digest are included in the digest, the comments are selected based on ranking. Ranking can be based on criteria such as, e.g., prioritizing the user's own direct comments, time separation between the image and the message, topic similarity with other images and/or text selected for the digest, etc.

If the user permits, the digest creation algorithm can be used to select relevant content, e.g., photos, that are not shared in the conversation but are present on the user device. With user permission, selection of such photos can be based on heuristics such as identities, demographics, and relationships of people in the photo, time and location of capture of the photo, etc. The selection can be further based on associations between the relevant characteristics of the photo and the properties of the conversation, individual messages within the conversation, parties in the conversation, etc. For instance, with user permission, a photo external to the conversation may be selected for the digest if it was taken around the same time as a conversation message included in the digest and includes all of the parties involved in the conversation.

The digest creation can be provided as part of a messaging application, or as a feature of a virtual assistant application. For example, a virtual assistant can be used to select content for inclusion in the digest.

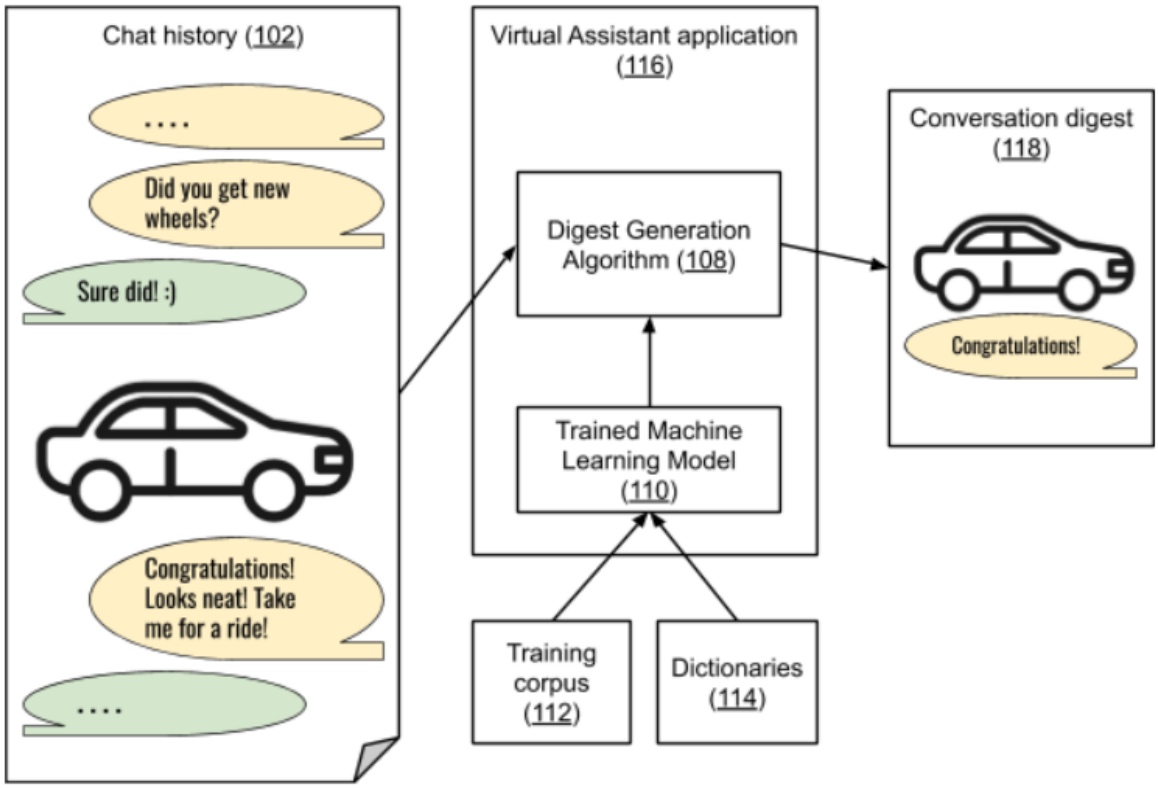

Fig. 1: Generating a digest from freeform unstructured conversational messages

Fig. 1 shows generation of a digest from an unstructured freeform messaging conversation, with chat history (102). With permission of the relevant conversation parties, the chat history is provided as input to a digest generation algorithm (108), e.g., implemented as part of a virtual assistant application (116). Topics and themes for the individual pieces of content within the conversation are detected, e.g., using a trained machine learning model (110). The model is trained using a training corpus of online interactions (112) and dictionaries that include information on relationships among words and phrases (114). With user permission, the generated digest can include photos, e.g., from the conversation or otherwise associated with the

user. A digest that highlights the most memorable aspects of the original conversation (118) is generated and presented to the user for review. The digest can also be interactive.

In the example illustrated in Fig. 1, a portion of the chat history is shown where two users exchange messages related to one of the users purchasing a car and the other user offering congratulations. As can be seen in the digest, salient aspects such as the photo of the car and the congratulatory message are included in the digest.

The parameter *k* used to select the highest ranked content for the digest can be determined by the developers of the digest generation algorithm, specified by the user, or dynamically determined during the digest creation process. Further, the user is given the opportunity to review the content of the generated digest. Upon review, the user can decide to delete or edit any piece of content within the digest, rearrange the order in which the content appears within the digest, or insert additional content at any place within the digest. The user can choose to retain the reviewed digest solely for personal use or choose to share the digest with other parties, such as those involved in the digested conversation, those connected to the user on social media, etc. The stored or shared digest can be replayed as necessary.

The described techniques can be extended by using text-to-speech technology to read the text content summarized within a digest. The text can be read with a system-generated voice or any other voice. For example, if the users permit, a user-specific voice model can be used to generate the speech for the digest. Where appropriate, a hybrid approach can be employed such that the digest is composed of human-generated and system-created audio clips.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social

actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to automatically generate a digest of conversational interactions in messaging applications with permission from the conversing parties. The conversational interactions may contain a mixture of text, images, audio, and video. A user can review and save the digest offline for later personal use, share the digest with the other conversation parties, or post it on social media. The described techniques can be applied to direct one-to-one conversations as well as to multi-party group conversations.