November 2022

Local Detection of Topical Entities Using Machine Learning

Hakim Sidahmed

Sushant Prakash

Raphael Fuchs

Ellie Zhou

Marius Dragus

See next page for additional authors

Inventor(s)

Hakim Sidahmed, Sushant Prakash, Raphael Fuchs, Ellie Zhou, Marius Dragus, and Felix Schmiedl

LOCAL DETECTION OF TOPICAL ENTITIES USING MACHINE LEARNING

Introduction

As users interact more and more with mobile phones and other devices, it is imperative for service providers that utilize user data to do so in privacy preserving ways. For example, service providers may need to understand screen content in a private way, without sending user data from a local device to remote server for processing. While providers may have server-side technology that allows entity recognition, current techniques may not allow privacy preserving manners of understanding on-device content.

Summary

Computer-implemented systems and methods for detecting topical entities on device using locally stored machine learned models provides for understanding of content being displayed on the device with enhanced user data privacy. A device may display textual content which is input to an entity extraction model on the device to determine entities (e.g., people, places, things, situations, etc.) presently displayed. The entities are input to a topicality machine learned model stored on the device to determine a confidence score for each entity input. The confidence score directly correlates to the likelihood that an entity is the topic of the textual content being displayed. The confidence score of each entity and the entity itself may then be used by the device to tailor features or functions to a user based on the topics being displayed. For example, a user may use a personal device to search for movie tickets. The personal device may be able to extract the textual content being displayed at a given moment and determine the user is looking at movie tickets for a recently released movie. The device may then generate voice recognitions and prompts using phrases such as, for example “this movie” or “that movie” that map to the identified movie. The user may then interact with the device in a more natural

communication format by saying things such as, for example, “Please by tickets for this movie” or “Show me movie times near me for that movie.”

Computer-implemented systems and methods for providing on device topic detection with the disclosed technology may provide for real-time feature adjustments or additions that are directly tailored to content displayed on a device. A device may include a display and a computing system. The computing system may extract content presented on the display and input the content to an entity detection model that determines entities (e.g., people, places, things, etc.) present in provided content and a confidence score that denotes a confidence or likelihood in the determined entity. The determined entities and their confidence scores may be input to a topicality model which may output a confidence score for each entity signifying how likely it is that each entity is the topic of the previously input content. The determined entities and their respective topicality scores may be used by the computing system to enable or generate tailored user features on the device such as, for example, tailored voice response, shopping recommendations, predictive text, or similar features that may benefit from knowing what topics are currently being displayed on the device.

Consider an example of a user viewing a webpage including a review of a new movie. The personal computing device may extract the content currently displayed on the screen and input it to a set of machine-learned models to determine that a movie review of movie “ABC” is currently being displayed on the user computing device. The user computing device may then tailor pronouns or definite articles (“this”, “these”, “that”, etc.) to the movie the review is discussing such that it may successfully interpret a voice command such as, for example, “show me movie times for this movie near me” or “what time is that movie playing near me?” or any other command a user may issue using pronouns or definite articles.

In one embodiment, a user computing device may use determined topics being displayed to generate accurate promotions or product recommendations for a user. For example, a user may view a product review. The user computing device may determine that a topic of the review is the product and provide a link to purchase the product or provide information for the product. It should be appreciated that any number of features or capabilities may be created using the entities and topic confidence scores output from machine-learned models stored on a device, and the examples disclosed herein are for descriptive purposes only.

Topics corresponding to currently displayed content may be determined using a first machine-learned model that uses extracted content currently displayed on the computing device to determine entities being displayed and entity confidence scores. The entity confidence scores indicate how confident the first machine-learned model is in a given entity prediction. A second machine-learned model determines a topic and confidence score for each entity indicating a likelihood the entity is the topic of the content currently displayed on the computing device. The machine-learned models may be stored locally on the computing device after trained by a server computing system which can provide model updates to the user computing devices. Initial or subsequent training may be performed on-device.

In one embodiment, machine-learned models are stored on a user computing device for detecting topicality of content being displayed on the user computing device. The user computing device may send feedback to a server computing system for model training. The server computing system may comprise two machine-learned models trained on content extracted from web pages. The server computing system may use the feedback data from the user computing device to further train the machine-learned models. The server computing system may send new machine-learned models or updates to the user computing device for detecting topicality of content being display on the user computing device.

In one embodiment, content may be used by a local machine-learned model on the user computing device to perform topic analysis. A determined topic may be used by the user computing device and/or be provided to a server computing system. The server computing system may use the topic from the user computing device for a plurality of functions or searches and return data to the user computing device. The user computing device may use the data from the server computing system to enable user features or functionality pertaining to the topic(s). In this manner, privacy of the user data can be maintained while providing user-specific features based on user data.

Detailed Description

Figure 1 depicts an example computing system 100 in which systems and methods in accordance with the present disclosure can be executed. The computing system includes a user computing device 102 containing one or more processors 112, memory 114 which may contain data 116 and instructions 118 configured to conduct the methods disclosed herein, and a user input component 122. The user input component can be, for example, a touch display or physical buttons within the user computing device 102. The computing system 100 further includes a network 180 and a server computing system 130. The server computing system 130 includes one or more processors 132, and memory 134 which may include data 136 and instruction 138 configured to conduct the methods disclosed herein. For example, a user may view content displayed on the user computing device 102. The displayed content may be extracted by processor(s) 112 and input to machine-learned models stored in memory 116. The machine-learned models may then output data which may be sent across network 180 to server computing system 130. The server computing system 130 may retrieve data stored in memory 134 or use network 180 to receive data from other computing systems (not shown) and return the data to the user computing device 102. The data returned to the user computing device 102 pertains to the output sent from the user computing device 102 to the server computing system 130.

Figure 2 depicts an example computing system 200 according to aspects of the present disclosure. The computing system 200 comprises a user computing device 202 comprising a display 206, processor(s) 210, a first machine learned model set 208, and a second machine-learned model set 212. The computing system 200 further comprises a remote computing system 204 comprising memory 214, processor(s) 218, a first machine-learned model set 216, and a second machine-learned model set 220. In one embodiment, the processor(s) 210 may extract content from display 206 and provide it as input to the first machine-learned model set 208 or the second machine-learned model set 212. In some instances, the user computing device 202 may send data pertaining to the machine-learned model sets to the remote computing system 204. The remote computing device 204 may use the data from the user computing device 202 to further train the first and second machine-learned model sets 216, 218. The remote computing system 204 may send data or copies pertaining to the machine-learned model sets 216, 218 to the user computing device 202 to replace or modify the machine-learned model sets 208, 212. In one embodiment, the first machine-learned model set 216 and second machine-learned model set 218 within the remote computing system 204 may be continuously trained on content obtained via processor(s) 218.

The user computing device may call an on-device OCR system to extract the text from the screen. The OCR returns the text extracted from the screen, as well as position of each word, and an approximation of its height.

The user computing device can use the entity embeddings to detect entities on the screen (using the output of the OCR). These embeddings enable the differentiation between entities such as "Taylor Swift" (singer), and text, like "swift" (fast, or programming language), and "Taylor" (name). The embedding detection comes with a confidence score (how sure is the entity model that the entity it detects is what it thinks it is). The output of this extraction step is a set of entities, the associated confidence score, as well as their position on the screen.

The extracted signals may include entity embeddings, confidence scores, entity heights, positions on the screen and the like. These can be provided to a transformer model, which outputs the topicality score of each entity. The transformer model can be trained on the server, using, as golden labels, the topicality scores output by the server-side model, which itself is trained using signals that are not available on device.

Referring now to Figure 3, an example method 300 for extracting topics is provided according to aspects of the present disclosure. The method incorporates a user computing device 302, a first machine-learned model 306, and a second machine-learned model 310. The user computing device 302 may display content 304 for a user and extract said content 304 to input to the first machine-learned model 306. The first machine-learned model 306 may then output a first output 308 comprising one or more entities and respective confidence scores. The respective confidence scores may correspond to how confident the first machine-learned model 306 is in each entity (e.g., people, place, things, etc.) prediction. The first output 308 comprising entities and respective confidence scores may then be input to the second machine-learned model 310 to output a final output 312 comprising topical entities and respective confidence scores. The topical entities in the final output 312 may be the same entities within the first output 308 from the first machine-learned model 306, however the confidence scores in the final output 312 may correspond to how confident the second machine-learned model is that each topical entity of the topic of the content 304. The final output 312 may then be stored on the user computing device 302 to enable tailored features or capabilities within the device. For example, the user computing device may utilize the topical entities from the final output 312 to enable use of definite articles (e.g., this, that, those, etc.) or pronouns (e.g., him, her, them, etc.) in voice commands in reference to content 304 displayed by the user computing device 302. It should be appreciated that the enabled features or capabilities described herein are for descriptive purposes only, in practice the final output 312 may be used to enable a plurality of features or capabilities.

Figure 4 depicts an example method 400 according to aspects of the present disclosure. At 402, the method comprises extracting content from a device's display. The content may be any text interpretable data, melodic data, or any other medium for providing data to a user. At 404, the extracted content is input to a first machine-learned model. The first machine-learned model may be trained to determine entities within given data such as, for example, people, places, things, situations or similar. At 406, the machine-learned model outputs entities and respective confidence scores based on the input data from the device's display. At 408, the output entities and respective confidence scores from the first machine-learned model are input to a second machine-learned model. The second machine-learned model may be trained to determine a topic of input content and entities extracted from said content. At 410, the second machine-learned model may output entities and topicality confidence scores, the topical confidence scores pertaining to a likelihood that a given entity is the topic of the extracted content. At 412, the device may configure features on said device to utilize the determined topic entities and topicality confidence scores. For example, the device may enable shopping recommendations based on products being mentioned, displayed, or alluded to by content on the device. It should be appreciated that the method described herein is laid out in no particular order and any steps may be rearranged, omitted, repeated, etc. within the method of determining topics from content.

Figures

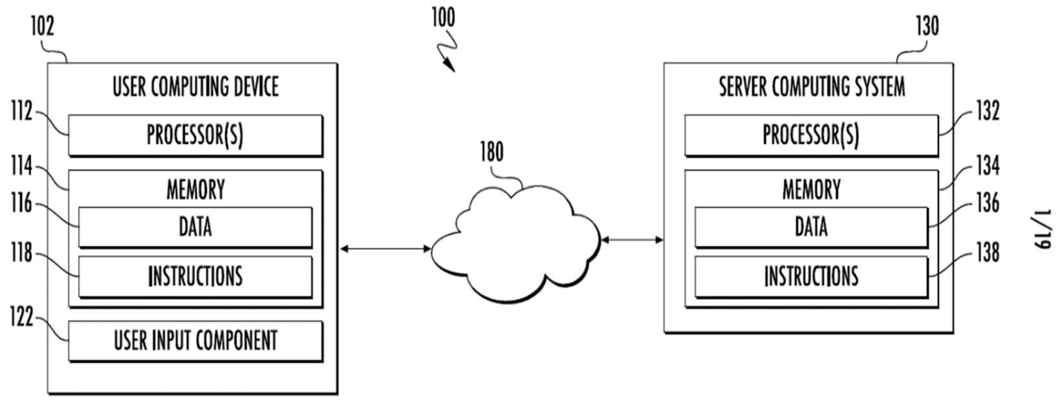

FIG. 1

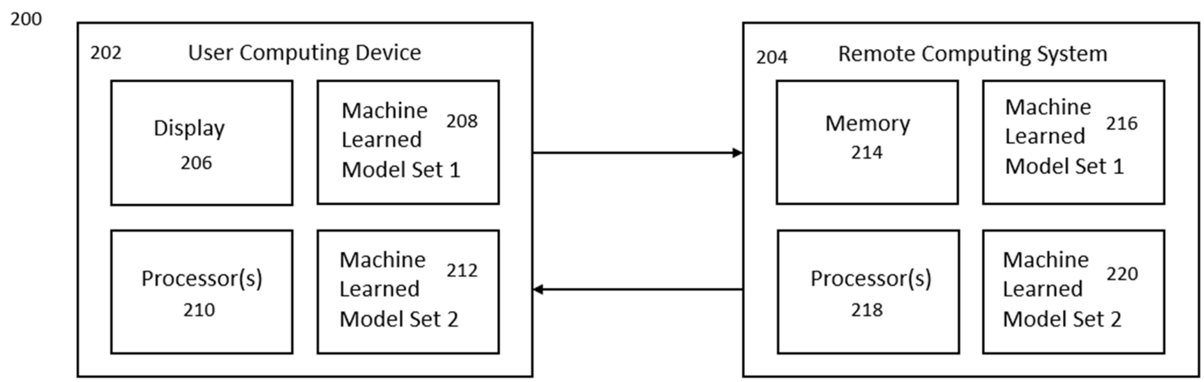

FIG. 2

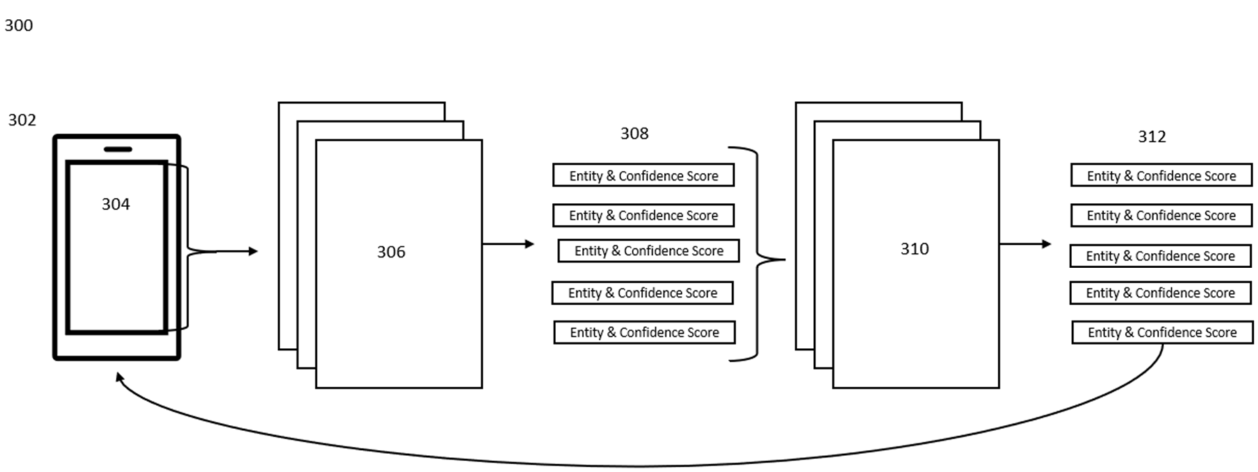

FIG. 3

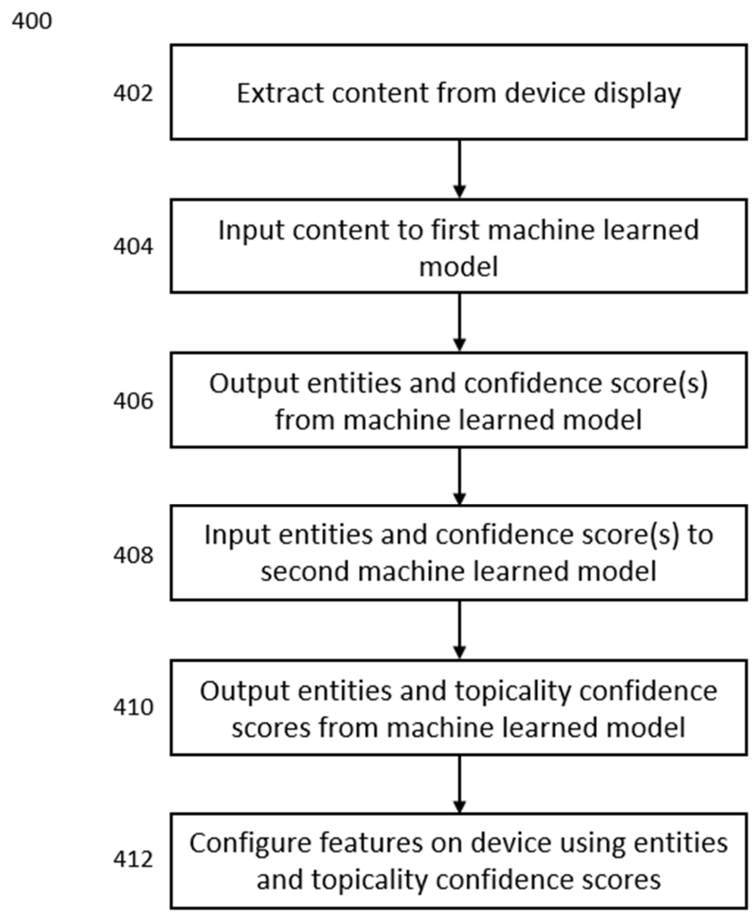

FIG. 4

Abstract

Computer-implemented systems and methods for determining topics of displayed content are provided while maintaining user data privacy and security. Entity identification and topic determination models may be stored within a user computing device such that the user computing device may perform topic detection of content presently displayed on the user computing device to maintain user data privacy. Once a topic(s) is determined from the content, features within the user computing device may be enabled or tailored to a user based on the content being displayed.