SYSTEM, METHOD, AND COMPUTER PROGRAM PRODUCT FOR RECOMMENDING A SET OF ITEMS TO A USER

MARYAM MOOSAEI
VISA

HAO YANG
VISA

YUSAN LIN
VISA

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
MOOSAEI, MARYAM; YANG, HAO; and LIN, YUSAN, "SYSTEM, METHOD, AND COMPUTER PROGRAM PRODUCT FOR RECOMMENDING A SET OF ITEMS TO A USER", Technical Disclosure Commons, (August 24, 2020)
https://www.tdcommons.org/dpubs_series/3535

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
TITLE “SYSTEM, METHOD, AND COMPUTER PROGRAM PRODUCT FOR RECOMMENDING A SET OF ITEMS TO A USER”

APPLICANT: VISA INTERNATIONAL SERVICE ASSOCIATION

ADDRESS: P.O. Box 8999, San Francisco, CA, 94128, USA

Nationality: USA

INVENTORS: MARYAM MOOSAEI;
HAO YANG; AND
YU-SAN LIN
TECHNICAL FIELD

[0001] This disclosure relates to systems, methods, and computer program products for recommending a set of items to a user.

BACKGROUND

[0002] Recommending products to consumers has been a widely and thoroughly studied research problem. In recent years, recommending fashion clothing items to consumers based on personal tastes of the consumers has received a relatively large amount of attention. Although many approaches have been proposed, more successful methodologies have typically leveraged visuals of fashion clothing items. For example, some techniques focus on recommending single fashion clothing items to users. However, when shopping for fashion clothing items, consumers may be more tempted to purchase items that can be viewed together as a compatible fashion outfit. Accordingly, some techniques focus on learning a compatibility of fashion clothing items.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] Additional advantages and details are explained in greater detail below with reference to the exemplary embodiments that are illustrated in the accompanying schematic figures, in which:

[0004] FIGURE 1 illustrates a diagram of non-limiting embodiments or aspects of an environment in which systems, devices, products, apparatuses, and/or methods, described herein, can be implemented.

[0005] FIGURE 2 illustrates a diagram of non-limiting embodiments or aspects of components of one or more devices and/or one or more systems of FIGURE 1.

[0006] FIGURE 3 shows a flowchart illustrating a method of recommending a set of items to a user in accordance with some embodiments of the present disclosure.
FIGURE 4 illustrates a block diagram of an exemplary computer system for implementing embodiments consistent with the present disclosure.

DESCRIPTION OF THE DISCLOSURE

[0008] It is to be understood that the present disclosure may assume various alternative variations and step sequences, except where expressly specified to the contrary. It is also to be understood that the specific devices and processes illustrated in the attached drawings and described in the following specification are simply exemplary and non-limiting embodiments or aspects. Hence, specific dimensions and other physical characteristics related to the embodiments or aspects disclosed herein are not to be considered as limiting.

[0009] For purposes of the description hereinafter, the terms “end,” “upper,” “lower,” “right,” “left,” “vertical,” “horizontal,” “top,” “bottom,” “lateral,” “longitudinal,” and derivatives thereof shall relate to the disclosed subject matter as it is oriented in the drawing figures. However, it is to be understood that the disclosed subject matter may assume various alternative variations and step sequences, except where expressly specified to the contrary. It is also to be understood that the specific devices and processes illustrated in the attached drawings, and described in the following specification, are simply exemplary embodiments or aspects of the disclosed subject matter. Hence, specific dimensions and other physical characteristics related to the embodiments or aspects disclosed herein are not to be considered as limiting unless otherwise indicated.

[0010] No aspect, component, element, structure, act, step, function, instruction, and/or the like used herein should be construed as critical or essential unless explicitly described as such. Also, as used herein, the articles “a” and “an” are intended to include one or more items and may be used interchangeably with “one or more” and “at least one.” Furthermore, as used herein, the term “set” is intended to include one or more items (e.g., related items, unrelated items, a combination of related and unrelated items, and/or the like) and may be used interchangeably with “one or more” or “at least one.” Where only one item is intended, the term “one” or similar language is used. Also, as used herein, the terms “has,” “have,” “having,” or the like are intended to be open-ended terms. Further, the phrase “based on” is intended to mean “based at least partially on” unless explicitly stated otherwise.
As used herein, the terms “communication” and “communicate” may refer to the reception, receipt, transmission, transfer, provision, and/or the like of information (e.g., data, signals, messages, instructions, commands, and/or the like). For one unit (e.g., a device, a system, a component of a device or system, combinations thereof, and/or the like) to be in communication with another unit means that the one unit is able to directly or indirectly receive information from and/or transmit information to the other unit. This may refer to a direct or indirect connection (e.g., a direct communication connection, an indirect communication connection, and/or the like) that is wired and/or wireless in nature. Additionally, two units may be in communication with each other even though the information transmitted may be modified, processed, relayed, and/or routed between the first and second unit. For example, a first unit may be in communication with a second unit even though the first unit passively receives information and does not actively transmit information to the second unit. As another example, a first unit may be in communication with a second unit if at least one intermediary unit (e.g., a third unit located between the first unit and the second unit) processes information received from the first unit and communicates the processed information to the second unit. In some non-limiting embodiments or aspects, a message may refer to a network packet (e.g., a data packet and/or the like) that includes data. It will be appreciated that numerous other arrangements are possible.

As used herein, the term “computing device” may refer to one or more electronic devices that are configured to directly or indirectly communicate with or over one or more networks. A computing device may be a mobile or portable computing device, a desktop computer, a server, and/or the like. Furthermore, the term “computer” may refer to any computing device that includes the necessary components to receive, process, and output data, and normally includes a display, a processor, a memory, an input device, and a network interface. A “computing system” may include one or more computing devices or computers. An “application” or “application program interface” (API) refers to computer code or other data sorted on a computer-readable medium that may be executed by a processor to facilitate the interaction between software components, such as a client-side front-end and/or server-side back-end for receiving data from the client. An “interface” refers to a generated display, such as one or more graphical user interfaces (GUIs) with which a user may interact, either directly or indirectly (e.g., through a keyboard, mouse, touchscreen, etc.). Further, multiple computers, e.g., servers, or other computerized devices, such as an autonomous vehicle including a vehicle.
computing system, directly or indirectly communicating in the network environment may constitute a “system” or a “computing system”.

[0013] It will be apparent that systems and/or methods, described herein, can be implemented in different forms of hardware, software, or a combination of hardware and software. The actual specialized control hardware or software code used to implement these systems and/or methods is not limiting of the implementations. Thus, the operation and behavior of the systems and/or methods are described herein without reference to specific software code, it being understood that software and hardware can be designed to implement the systems and/or methods based on the description herein.

[0014] Some non-limiting embodiments or aspects are described herein in connection with thresholds. As used herein, satisfying a threshold may refer to a value being greater than the threshold, more than the threshold, higher than the threshold, greater than or equal to the threshold, less than the threshold, fewer than the threshold, lower than the threshold, less than or equal to the threshold, equal to the threshold, etc.

[0015] Recommending sets of items (e.g., fashion outfits including clothing items, furniture sets including furniture items, landscaping area designs including landscaping items, graphical user interface (GUI) designs including GUI items, architectural designs including architectural items, artwork designs including artwork items, food or culinary designs including food items, etc.) to users involves several challenges. For example, different sets of items (e.g., different fashion outfits, etc.) may include different numbers of items (e.g., different fashion outfits may include different numbers of clothing items, such as a different number of (or no number of) a shirt, a pair of pants, a pair of shoes, a purse, a coat, a jacket, a watch, a skirt, a bracelet, a belt, and/or the like, etc.) and/or different types of a same item (e.g., different fashion outfits may include different types of a same clothing item, such as different types of a shirt, different types of a pair of pants, different types of a pair of shoes, different types of a purse, different types of a coat, different types of a jacket, different types of a watch, different types of a skirt, different types of a bracelet, different types of a belt, and/or the like, etc.). As an example, a user may have different personal preferences for different types of items in sets of items (e.g., a particular user may put a greater emphasis on a pair shoes when determining her personal preference for a fashion outfit including the pair of shoes and one or more other clothing items, etc.). In such examples, user preferences for sets of items may only be available
for sets of items that a user is known to prefer (e.g., like, etc.). For example, data associated with user preferences for sets of items (e.g., fashion outfits, etc.) may only indicate that a user prefers or likes the set of items itself as a whole (e.g., a fashion outfit itself as a whole, etc.), but may not provide an indication as to whether (and/or an indication as to an amount, how much, etc.) the user prefers or likes an individual item in the plurality of items included in the set of items (e.g., a user preference for an individual clothing item of a plurality of clothing items included in a fashion outfit liked by the user, etc.). As an example, data associated with user preferences for sets of items (e.g., fashion outfits, etc.) may be unavailable with respect to relationships between users and the sets of items (e.g., with respect to relationships between users and fashion outfits, etc.). For example, data associated with user preferences for sets of items may indicate only what a user prefers or likes, but not what users do not prefer or dislike.

[0016] Provided are improved systems, devices, products, apparatuses, and/or methods for recommending a set of items to a user.

[0017] Existing computing systems for recommending an item or items to a user may not capture a personal emphasis or attention on one or more items of a set of items by a user when determining a preference of the user toward the set of items. For example, existing computing systems may not use or implement a neural network or model for which (i) processing of inputs is order invariant; (ii) inputs to one or more layers for learning personal emphasis or attention on one or more items of a set of items by a user are independent of each other; and (iii) weighting for parameters of one or more layers for learning a personal emphasis or attention on one or more items of a set of items by a user is personalized to the user. In this way, existing computing systems may not (i) use different numbers of items in various sets of items for learning a preference of a user with respect to the sets of items; (ii) learn a preference of a user toward a labeled set of items as well as a preference of the user toward unlabeled items; (iii) recommend a set of items without using negatively labeled items and/or negatively labeled sets of items; and (iv) address a problem of different emphasis imposed on different items of sets of items across different users.

[0018] Non-limiting embodiments or aspects of the present disclosure are directed to systems, devices, products, apparatus, and/or methods for recommending a set of items to a user that obtain training data including: a user identifier of a user; one or more first images including a first set of multiple items; and one or more second images including a second set
of multiple items; and train a neural network based on the training data by: processing, with at least one first fusing layer of the neural network, the user identifier and the first set of multiple items to concatenate each item of the multiple items in the first set of multiple items with the user identifier in a first set of fused embeddings; processing, with at least one second fusing layer of the neural network, the user identifier and the second set of multiple items to concatenate each item of the multiple items in the second set of multiple items with the user identifier in a second set of fused embeddings; processing, with at least one first scoring layer of the neural network, the first set of fused embeddings to determine a first score associated with the first set of multiple items for the user identifier; processing, with at least one second scoring layer of the neural network, the second set of fused embeddings to determine a second score associated with the second set of multiple items for the user identifier; and modifying, using an objective function of the neural network that depends on the first score, the second score, and a margin between the first score and the second score, one or more parameters of the neural network.

[0019] In this way, non-limiting embodiments or aspects of the present disclosure capture a personal emphasis or attention on one or more items of a set of items by a user when determining a preference of the user toward the set of items. For example, non-limiting embodiments or aspects of the present disclosure provide for (i) processing of inputs that is order invariant; (ii) inputs to one or more layers for learning personal emphasis or attention on one or more items of a set of items by a user that are independent of each other; and/or (iii) weighting for parameters of one or more layers for learning a personal emphasis or attention on one or more items of a set of items by a user that is personalized to the user. Accordingly, non-limiting embodiments or aspects of the present disclosure may (i) use different numbers of items in various sets of items for learning a preference of a user with respect to the sets of items; (ii) learn a preference of a user toward a labeled set of items as well as a preference of the user toward unlabeled items; (iii) recommend a set of items without using negatively labeled items and/or negatively labeled sets of items; and/or (iv) address a problem of different emphasis imposed on different items of sets of items across different users.

[0020] Referring now to Figure 1, Figure 1 illustrates a diagram of an example environment 100 in which systems, devices, products, apparatuses, and/or methods, described herein, can be implemented. As shown in Figure 1, environment 100 includes recommendation system 102, communication network 104, image data system 106, and/or user device 108. Systems
and/or devices of environment 100 can interconnect via wired connections, wireless connections, or a combination of wired and wireless connections. For example, systems and/or devices of environment 100 may interconnect and/or communicate information and/or data (e.g., image data, text data, user identifiers, user preference scores, etc.) via communication network 104.

[0021] Recommendation system 102 may include one or more devices capable of receiving information and/or data from image data system 106 and/or user device 108 via communication network 104 and/or communicating information and/or data to image data system 106 and/or user device 108 via communication network 104. For example, recommendation system 102 may include one or more computing systems including one or more processors (e.g., one or more computing devices, one or more mobile computing devices, one or more servers, etc.).

[0022] In some non-limiting embodiments or aspects, recommendation system 102 includes one or more devices capable of obtaining training data including: a user identifier of a user; one or more first images including a first set of multiple items; and one or more second images including a second set of multiple items; and training a neural network based on the training data by: processing, with at least one first fusing layer of the neural network, the user identifier and the first set of multiple items to concatenate each item of the multiple items in the first set of multiple items with the user identifier in a first set of fused embeddings; processing, with at least one second fusing layer of the neural network, the user identifier and the second set of multiple items to concatenate each item of the multiple items in the second set of multiple items with the user identifier in a second set of fused embeddings; processing, with at least one first scoring layer of the neural network, the first set of fused embeddings to determine a first score associated with the first set of multiple items for the user identifier; processing, with at least one second scoring layer of the neural network, the second set of fused embeddings to determine a second score associated with the second set of multiple items for the user identifier; and modifying, using an objective function of the neural network that depends on the first score, the second score, and a margin between the first score and the second score, one or more parameters of the neural network.

[0023] In some non-limiting embodiments or aspects, an image includes a matrix (e.g., a grid, a rectangular array, a multi-dimensional grid, a multi-dimensional array, a set of rows and columns, etc.) that has a plurality of elements (e.g., units, cells, pixels, etc.). Each element of
the matrix includes image data (e.g., a value of image data, a value of geographic location image data, a value of color data, a value of a prediction score, etc.) associated with the image. In some non-limiting embodiments or aspects, one or more elements (e.g., each element, etc.) of an image is associated with at least three dimensions. For example, a first dimension of the element may include a width of the element, a second dimension of the element may include a length of the element, and a third dimensions of the element may include a value associated with the image data of the element (e.g., a value of the image data stored by the element, etc.). In some non-limiting embodiments, training data includes one or more images.

[0024] Communication network 104 may include one or more wired and/or wireless networks. For example, communication network 104 may include a cellular network (e.g., a long-term evolution (LTE) network, a third generation (3G) network, a fourth generation (4G) network, a fifth generation (5G) network, a code division multiple access (CDMA) network, etc.), a public land mobile network (PLMN), a local area network (LAN), a wide area network (WAN), a metropolitan area network (MAN), a telephone network (e.g., the public switched telephone network (PSTN)), a private network, an ad hoc network, an intranet, the Internet, a fiber optic-based network, a cloud computing network, and/or the like, and/or a combination of these or other types of networks.

[0025] Image data system 106 may include one or more devices capable of receiving information and/or data from recommendation system 102 and/or user device 108 via communication network 104 and/or communicating information and/or data to recommendation system 102 and/or user device 108 via communication network 104. For example, image data system 106 may include one or more computing systems including one or more processors (e.g., one or more computing devices, one or more mobile computing devices, one or more servers, etc.). In some non-limiting embodiments or aspects, image data system 106 includes one or more devices capable of storing, retrieving, accessing, and/or providing image data associated with one or more images including one or more items. For example, image data system 106 may provide or make available one or more websites and/or one or more databases from which one or more images of one or more items are available for access and/or retrieval, and recommendation system 102 can access and/or web crawl the one or more websites and/or query the one or more databases to access and/or retrieve the one or more images to create a dataset of image data. As an example, one or more of the images may be associated with one or more users (e.g., one or more of the images may include an indication
that the one or more users likes or prefers one or more items in one or more images, a specific item of a plurality of different items in a single image including the plurality of different items, a specific image including one or more items, etc.). In such an example, image data system 106 may include a fashion website (e.g., Polyvore, Lookbook.nu, etc.) from which images of clothing items, images of fashion outfits that each include multiple clothing items (e.g., a shirt, a skirt, a hat, a purse, etc.), images of single clothing items (e.g., a single shirt, a single skirt, a single hat, a single, purse, etc.) are available, and a fashion outfit, clothing items in the fashion outfit, and/or a single clothing item may be associated with a user that has indicated that the user likes the fashion outfit, one or more clothing items in the fashion outfit, and/or the single clothing item to the fashion website. However, non-limiting embodiments or aspects are not limited thereto, and an item in an image may include one or more items of any type of item, such as a clothing item, a furniture item, a landscaping item, a GUI item, an architectural item, an artwork item, a food or culinary item, and/or the like.

[0026] User device 108 may include one or more devices capable of receiving information and/or data from recommendation system 102, image data system 106 and/or another user device 108 via communication network 104 and/or communicating information and/or data to recommendation system 102, image data system 106 and/or another user device 108 via communication network 104. For example, user device 108 may include one or more computing systems including one or more processors (e.g., one or more computing devices, one or more mobile computing devices, one or more servers, etc.). In some non-limiting embodiments or aspects, user device 108 may include a client device and/or the like. For example, user device 108 may be associated with a user. As an example, user device 108 can receive information associated with one or more items and/or one or more sets of one or more items from recommendation system 102 and/or image data system 106. In such an example, recommendation system 102 can generate, for a user of user device 108, a recommendation for a set of items (e.g., a recommendation for a particular set of items of a plurality of different sets of items processed with the trained neural network by recommendation system 102, such as a highest ranked or scored set of items relative to the plurality of different sets of items, one or more sets of items of the plurality of different sets of items that satisfy one or more thresholds, and/or the like, etc.).

[0027] The number and arrangement of devices, systems, and/or networks shown in Figure 1 are provided as an example. There may be additional devices, systems, and/or networks,
fewer devices, systems, and/or networks, different devices, systems, and/or networks, or differently arranged devices, systems, and/or networks than those shown in Figure 1. Furthermore, two or more devices, systems, and/or networks shown in Figure 1 may be implemented within a single device, a single system, and/or a single network, or a single device, a single system, and/or a single network shown in Figure 1 may be implemented as multiple, distributed devices, systems, and/or networks. Additionally, or alternatively, a set of devices (e.g., one or more devices, etc.), a set of systems (e.g., one or more systems, etc.), and/or a set of networks (e.g., one or more networks, etc.) of environment 100 may perform one or more functions described as being performed by another set of devices, systems, and/or networks of environment 100.

[0028] Referring now to Figure 2, Figure 2 illustrates a diagram of example components of a device 200. Device 200 may correspond to one or more devices of recommendation system 102, one or more devices of communication network 104, one or more devices of image data system 106, and/or user device 108 (e.g., one or more devices of a system of user device 108, etc.). In some non-limiting embodiments or aspects, one or more devices of recommendation system 102, one or more devices of communication network 104, one or more devices of image data system 106, and/or user device 108 (e.g., one or more devise of a system of user device 108, etc.) can include at least one device 200 and/or at least one component of device 200. As shown in Figure 2, device 200 may include a bus 202, a processor 204, memory 206, a storage component 208, an input component 210, an output component 212, and a communication interface 214.

[0029] Bus 202 may include a component that permits communication among the components of device 200. In some non-limiting embodiments or aspects, processor 204 may be implemented in hardware, firmware, or a combination of hardware and software. For example, processor 204 may include a processor (e.g., a central processing unit (CPU), a graphics processing unit (GPU), an accelerated processing unit (APU), etc.), a microprocessor, a digital signal processor (DSP), and/or any processing component (e.g., a field-programmable gate array (FPGA), an application-specific integrated circuit (ASIC), etc.) that can be programmed to perform a function. Memory 206 may include random access memory (RAM), read only memory (ROM), and/or another type of dynamic or static storage device (e.g., flash memory, magnetic memory, optical memory, etc.) that stores information and/or instructions for use by processor 204.
Storage component 208 may store information and/or software related to the operation and use of device 200. For example, storage component 208 may include a hard disk (e.g., a magnetic disk, an optical disk, a magneto-optic disk, a solid state disk, etc.), a compact disc (CD), a digital versatile disc (DVD), a floppy disk, a cartridge, a magnetic tape, and/or another type of computer-readable medium, along with a corresponding drive.

Input component 210 may include a component that permits device 200 to receive information, such as via user input (e.g., a touch screen display, a keyboard, a keypad, a mouse, a button, a switch, a microphone, etc.). Additionally, or alternatively, input component 210 may include a sensor for sensing information (e.g., a global positioning system (GPS) component, an accelerometer, a gyroscope, an actuator, etc.). Output component 212 may include a component that provides output information from device 200 (e.g., a display, a speaker, one or more light-emitting diodes (LEDs), etc.).

Communication interface 214 may include a transceiver-like component (e.g., a transceiver, a separate receiver and transmitter, etc.) that enables device 200 to communicate with other devices, such as via a wired connection, a wireless connection, or a combination of wired and wireless connections. Communication interface 214 may permit device 200 to receive information from another device and/or provide information to another device. For example, communication interface 214 may include an Ethernet interface, an optical interface, a coaxial interface, an infrared interface, a radio frequency (RF) interface, a universal serial bus (USB) interface, a Wi-Fi® interface, a cellular network interface, and/or the like.

Device 200 may perform one or more processes described herein. Device 200 may perform these processes based on processor 204 executing software instructions stored by a computer-readable medium, such as memory 206 and/or storage component 208. A computer-readable medium (e.g., a non-transitory computer-readable medium) is defined herein as a non-transitory memory device. A memory device includes memory space located inside of a single physical storage device or memory space spread across multiple physical storage devices.

Software instructions may be read into memory 206 and/or storage component 208 from another computer-readable medium or from another device via communication interface 214. When executed, software instructions stored in memory 206 and/or storage component
208 may cause processor 204 to perform one or more processes described herein. Additionally, or alternatively, hardwired circuitry may be used in place of or in combination with software instructions to perform one or more processes described herein. Thus, embodiments or aspects described herein are not limited to any specific combination of hardware circuitry and software.

[0035] The number and arrangement of components shown in Figure 2 are provided as an example. In some non-limiting embodiments or aspects, device 200 may include additional components, fewer components, different components, or differently arranged components than those shown in Figure 2. Additionally, or alternatively, a set of components (e.g., one or more components) of device 200 may perform one or more functions described as being performed by another set of components of device 200.

[0036] Figure 3 shows a flowchart illustrating a method of recommending a set of items to a user in accordance with some embodiments of the present disclosure.

[0037] As illustrated in Figure 3, the method 300 includes one or more blocks for recommending a set of items to a user. The method 300 may be described in the general context of computer executable instructions. Generally, computer executable instructions can include routines, programs, objects, components, data structures, procedures, modules, and functions, which perform particular functions or implement particular abstract data types.

[0038] The order in which the method 300 is described is not intended to be construed as a limitation, and any number of the described method blocks can be combined in any order to implement the method. Additionally, individual blocks may be deleted from the methods without departing from the scope of the subject matter described herein. Furthermore, the method can be implemented in any suitable hardware, software, firmware, or combination thereof.

[0039] At block 301, a user identifier of a user, one or more first images including a first set of multiple items, and one or more second images including a second set of multiple items are obtained. In one embodiment, at least one processor of recommendation system 102 obtains the user identifier of the user, the one or more first images including the first set of multiple items, and the one or more second images including the second set of multiple items for training data. The first set of multiple items includes a different number of items than the second set of
multiple items. A preference of the user for the first set of multiple items is known, and the preference of the user for the second set of multiple items is unknown. Subsequently, at least one processor of recommendation system 102 encodes the user identifier into a one-hot encoding vector in which an index of the vector associated with the user identifier is 1 and a remainder of the index values is 0. The at least one processor of recommendation system 102 processes the one or more first images including the first set of multiple items to generate, for each item of the multiple items in the first set, a visual item embedding in a first set of visual item embeddings using a first pre-trained convolutional neural network (CNN). The at least one processor of recommendation system 102 processes the one or more second images including the second set of multiple items to generate, for each item of the multiple items in the second set, a visual item embedding in a second set of visual item embeddings using a second pre-trained CNN.

[0040] At block 303, a neural network is trained based on the training data i.e. the user identifier of the user and the one or more first images including the first set of multiple items. In one embodiment, at least one processor of recommendation system 102 trains the neural network based on training data by processing, with at least one first fusing layer of the neural network, the user identifier and the first set of multiple items to concatenate each item of the multiple items in the first set of multiple items with the user identifier in a first set of fused embeddings. In detail, the at least one processor of recommendation system 102 processes, with at least one user embedding layer of the neural network, the one-hot encoding vector to generate a lower-dimensional user embedding. Subsequently, the at least one processor of recommendation system 102 processes, with at least one first embedding layer of the neural network, the first set of visual item embeddings to generate a lower-dimensional first set of visual item embeddings. The at least one processor of recommendation system 102 processes, with the at least one first fusing layer of the neural network, the lower-dimensional user embedding and the lower-dimensional first set of visual item embeddings to generate the first set of fused embeddings. The first set of fused embeddings includes each visual item embedding in the lower-dimensional first set of visual item embeddings concatenated with the lower-dimensional user embedding. The processing, with the at least one first fusing layer of the neural network, the user identifier and the first set of multiple items is order invariant.

[0041] At block 305, the neural network is trained based on the training data i.e. the user identifier of the user and the one or more second images including the second set of multiple
items. In one embodiment, at least one processor of recommendation system 102 trains the neural network based on training data by processing, with at least one second fusing layer of the neural network, the user identifier and the second set of multiple items to concatenate each item of the multiple items in the second set of multiple items with the user identifier in a second set of fused embeddings. In detail, the at least one processor of recommendation system 102 processes, with at least one second embedding layer of the neural network, the second set of visual item embeddings to generate a lower-dimensional second set of visual item embeddings. The at least one processor of recommendation system 102 processes, with the at least one second fusing layer of the neural network, the lower-dimensional user embedding and the lower-dimensional second set of visual item embeddings to generate the second set of fused embeddings. The second set of fused embeddings includes each visual item embedding in the lower-dimensional second set of visual item embeddings concatenated with the lower-dimensional user embedding. The processing, with at least one second fusing layer of the neural network, the user identifier and the second set of multiple items is order invariant.

[0042] At block 307, a first score associated with the first set of multiple items for the user identifier is determined. The at least one processor of recommendation system 102 processes, with at least one first scoring layer of the neural network, the first set of fused embeddings to determine the first score associated with the first set of multiple items for the user identifier. In detail, in one embodiment, the at least one processor of recommendation system 102 processes, with a first fully connected layer of the at least one first scoring layer of the neural network, each fused embedding of the first set of fused embeddings to generate a personalized item score for that fused embedding. The personalized item score is associated with the user identifier and an item of the multiple items in the first set of multiple items used to generate that fused embedding. Subsequently, the at least one processor of recommendation system 102 processes, with a first pooling layer of the at least one first scoring layer of the neural network, the lower-dimensional user embedding and each personalized item score for the first set of fused embeddings to determine the first score associated with the first set of multiple items for the user identifier. In another embodiment, the at least one processor of recommendation system 102 processes, with a first pooling layer of the at least one first scoring layer of the neural network, the lower-dimensional user embedding and each fused embedding of the first set of fused embeddings to generate a first contextual embedding. Subsequently, the at least one processor of recommendation system 102 processes, with a first fully connected layer of the at
least one first scoring layer of the neural network, the first contextual embedding to determine the first score associated with the first set of multiple items for the user identifier.

[0043] At block 309, a second score associated with the second set of multiple items for the user identifier is determined. The at least one processor of recommendation system 102 processes, with at least one second scoring layer of the neural network, the second set of fused embeddings to determine the second score associated with the second set of multiple items for the user identifier. In detail, in one embodiment, the at least one processor of recommendation system 102 processes, with a second fully connected layer of the at least one second scoring layer of the neural network, each fused embedding of the second set of fused embeddings to generate a personalized item score for that fused embedding. The personalized item score is associated with the user identifier and an item of the multiple items in the second set of multiple items used to generate that fused embedding. Subsequently, the at least one processor of recommendation system 102 processes, with a second pooling layer of the at least one second scoring layer of the neural network, the lower-dimensional user embedding and each personalized item score for the second set of fused embeddings to determine the second score associated with the second set of multiple items for the user identifier. In another embodiment, the at least one processor of recommendation system 102 processes, with a second pooling layer of the at least one second scoring layer of the neural network, the lower dimensional user embedding and each fused embedding of the second set of fused embeddings to generate a second contextual embedding. Subsequently, the at least one processor of recommendation system 102 processes, with a second fully connected layer of the at least one second scoring layer of the neural network, the second contextual embedding to determine the second score associated with the second set of multiple items for the user identifier.

[0044] At block 311, one or more parameters of the neural network is modified. In one embodiment, at least one processor of recommendation system 102 modifies the one or more parameters of the neural network using an objective function of the neural network that depends on the first score, the second score, and a margin between the first score and the second score.

[0045] Figure 4 illustrates a block diagram of an exemplary computer system for implementing embodiments consistent with the present disclosure.
In an embodiment, the computer system 400 may be used to implement the recommendation system 102. The computer system 400 may include a central processing unit ("CPU" or "processor") 402. The processor 402 may include at least one data processor for recommending a set of items to a user. The processor 402 may include specialized processing units such as, integrated system (bus) controllers, memory management control units, floating point units, graphics processing units, digital signal processing units, etc.

The processor 402 may be disposed in communication with one or more input/output (I/O) devices (412 and 413) via I/O interface 401. The I/O interface 401 employ communication protocols/methods such as, without limitation, audio, analog, digital, monoaural, radio corporation of America (RCA) connector, stereo, IEEE-1394 high speed serial bus, serial bus, universal serial bus (USB), infrared, personal system/2 (PS/2) port, bayonet neill-concelman (BNC) connector, coaxial, component, composite, digital visual interface (DVI), high-definition multimedia interface (HDMI), radio frequency (RF) antennas, S-Video, video graphics array (VGA), IEEE 802.11b/g/n/x, Bluetooth, cellular e.g., code-division multiple access (CDMA), high-speed packet access (HSPA+), global system for mobile communications (GSM), long-term evolution (LTE), worldwide interoperability for microwave access (WiMax), or the like, etc.

Using the I/O interface 401, the computer system 400 may communicate with one or more I/O devices such as input devices 412 and output devices 413. For example, the input devices 412 may be an antenna, keyboard, mouse, joystick, (infrared) remote control, camera, card reader, fax machine, dongle, biometric reader, microphone, touch screen, touchpad, trackball, stylus, scanner, storage device, transceiver, video device/source, etc. The output devices 413 may be a printer, fax machine, video display (e.g., cathode ray tube (CRT), liquid crystal display (LCD), light-emitting diode (LED), plasma, plasma display panel (PDP), organic light-emitting diode display (OLED) or the like), audio speaker, etc.

In some embodiments, the processor 402 may be disposed in communication with a communication network 409 via a network interface 403. The network interface 403 may communicate with the communication network 409. The network interface 403 may employ connection protocols including, without limitation, direct connect, ethernet (e.g., twisted pair 10/100/1000 Base T), transmission control protocol/internet protocol (TCP/IP), token ring, IEEE 802.11a/b/g/n/x, etc. The communication network 409 may include, without limitation,
a direct interconnection, local area network (LAN), wide area network (WAN), wireless network (e.g., using Wireless Application Protocol), the Internet, etc. Using the network interface 403 and the communication network 409, the computer system 400 may communicate with a database 414, which may be the image data system 106. The network interface 403 may employ connection protocols include, but not limited to, direct connect, ethernet (e.g., twisted pair 10/100/1000 Base T), transmission control protocol/internet protocol (TCP/IP), token ring, IEEE 802.11a/b/g/n/x, etc.

[0050] The communication network 409 includes, but is not limited to, a direct interconnection, a peer to peer (P2P) network, local area network (LAN), wide area network (WAN), wireless network (e.g., using Wireless Application Protocol), the Internet, Wi-Fi and such. The communication network 409 may either be a dedicated network or a shared network, which represents an association of the different types of networks that use a variety of protocols, for example, hypertext transfer protocol (HTTP), transmission control protocol/internet protocol (TCP/IP), wireless application protocol (WAP), etc., to communicate with each other. Further, the communication network 409 may include a variety of network devices, including routers, bridges, servers, computing devices, storage devices, etc.

[0051] In some embodiments, the processor 402 may be disposed in communication with a memory 405 (e.g., RAM, ROM, etc. not shown in Figure 4) via a storage interface 404. The storage interface 404 may connect to memory 405 including, without limitation, memory drives, removable disc drives, etc., employing connection protocols such as, serial advanced technology attachment (SATA), integrated drive electronics (IDE), IEEE-1394, universal serial bus (USB), fiber channel, small computer systems interface (SCSI), etc. The memory drives may further include a drum, magnetic disc drive, magneto-optical drive, optical drive, redundant array of independent discs (RAID), solid-state memory devices, solid-state drives, etc.

[0052] The memory 405 may store a collection of program or database components, including, without limitation, user interface 406, an operating system 407, etc. In some embodiments, computer system 400 may store user/application data, such as, the data, variables, records, etc., as described in this disclosure. Such databases may be implemented as fault-tolerant, relational, scalable, secure databases such as Oracle or Sybase.
The operating system 407 may facilitate resource management and operation of the computer system 400. Examples of operating systems include, without limitation, Apple™ Macintosh™ OS X™, UNIX™, Unix-like system distributions (e.g., Berkeley Software Distribution (BSD), FreeBSD™, Net BSD™, Open BSD™, etc.), Linux distributions (e.g., Red Hat™, Ubuntu™, K-Ubuntu™, etc.), International Business Machines (IBM™) OS/2™, Microsoft Windows™ (XP™, Vista/7/8, etc.), Apple iOS™, Google Android™, Blackberry™ operating system (OS), or the like. In some embodiments, the computer system 400 may implement web browser 408 stored program components. Web browser 408 may be a hypertext viewing application, such as Microsoft™ Internet Explorer™, Google Chrome™, Mozilla Firefox™, Apple™ Safari™, etc. Secure web browsing may be provided using secure hypertext transport protocol (HTTPS), secure sockets layer (SSL), transport layer security (TLS), etc. Web browsers 408 may utilize facilities such as AJAX, DHTML, Adobe™ Flash, Javascript, Application Programming Interfaces (APIs), etc.

According to some non-limiting embodiments or aspects, a computer program product including at least one non-transitory computer-readable medium including one or more instructions for recommending a set of items to a user that, when executed by at least one processor, cause the at least one processor to: obtain training data including: a user identifier of a user; one or more first images including a first set of multiple items; and one or more second images including a second set of multiple items; and train a neural network based on the training data by: processing, with at least one first fusing layer of the neural network, the user identifier and the first set of multiple items to concatenate each item of the multiple items in the first set of multiple items with the user identifier in a first set of fused embeddings; processing, with at least one second fusing layer of the neural network, the user identifier and the second set of multiple items to concatenate each item of the multiple items in the second set of multiple items with the user identifier in a second set of fused embeddings; processing, with at least one first scoring layer of the neural network, the first set of fused embeddings to determine a first score associated with the first set of multiple items for the user identifier; processing, with at least one second scoring layer of the neural network, the second set of fused embeddings to determine a second score associated with the second set of multiple items for the user identifier; and modifying, using an objective function of the neural network that depends on the first score, the second score, and a margin between the first score and the second score, one or more parameters of the neural network.
The illustrated steps are set out to explain the exemplary embodiments shown, and it should be anticipated that ongoing technological development will change the manner in which particular functions are performed. These examples are presented herein for purposes of illustration, and not limitation. Further, the boundaries of the functional building blocks have been arbitrarily defined herein for the convenience of the description. Alternative boundaries can be defined so long as the specified functions and relationships thereof are appropriately performed. Alternatives (including equivalents, extensions, variations, deviations, etc., of those described herein) will be apparent to persons skilled in the relevant art(s) based on the teachings contained herein. Such alternatives fall within the scope and spirit of the disclosed embodiments. Also, the words "comprising," "having," "containing," and "including," and other similar forms are intended to be equivalent in meaning and be open ended in that an item or items following any one of these words is not meant to be an exhaustive listing of such item or items, or meant to be limited to only the listed item or items. It must also be noted that as used herein, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise.

Furthermore, one or more computer-readable storage media may be utilized in implementing embodiments consistent with the present disclosure. A computer readable storage medium refers to any type of physical memory on which information or data readable by a processor may be stored. Thus, a computer readable storage medium may store instructions for execution by one or more processors, including instructions for causing the processor(s) to perform steps or stages consistent with the embodiments described herein. The term “computer readable medium” should be understood to include tangible items and exclude carrier waves and transient signals, i.e., are non-transitory. Examples include random access memory (RAM), read-only memory (ROM), volatile memory, nonvolatile memory, hard drives, CD ROMs, DVDs, flash drives, disks, and any other known physical storage media.

Finally, the language used in the specification has been principally selected for readability and instructional purposes, and it may not have been selected to delineate or circumscribe the inventive subject matter. Accordingly, the disclosure of the embodiments of the disclosure is intended to be illustrative, but not limiting, of the scope of the disclosure.
SYSTEM, METHOD, AND COMPUTER PROGRAM PRODUCT FOR RECOMMENDING A SET OF ITEMS TO A USER

ABSTRACT

Systems, methods, and computer program products obtain training data and train a neural network based on the training data by concatenating a user identifier to each item of a first set of multiple items in a first set of fused embeddings, concatenating the user identifier to each item of a second set of multiple items in a second set of fused embeddings, determining a first score associated with the first set of multiple items based on the first set of fused embeddings, determining a second score associated with the second set of multiple items based on the second set of fused embeddings, and modifying, using an objective function of the neural network that depends on the first score, the second score, and a margin between the first score and the second score, one or more parameters of the neural network.

FIGURE 1
Fig. 1
Fig. 2
Fig. 3

1. Obtain training data including a user identifier of a user, one or more first images including a first set of multiple items, and one or more second images including a second set of multiple items.

2. Train a neural network based on the training data by processing the user identifier and the first set of multiple items to concatenate each item of the multiple items in the first set of multiple items with the user identifier in a first set of fused embeddings.

3. Process the user identifier and the second set of multiple items to concatenate each item of the multiple items in the second set of multiple items with the user identifier in a second set of fused embeddings.

4. Process the first set of fused embeddings to determine a first score associated with the first set of multiple items for the user identifier.

5. Process the second set of fused embeddings to determine a second score associated with the second set of multiple items for the user identifier.

6. Modify using an objective function of the neural network that depends on the first score, the second score, and a margin between the first score and the second score, one or more parameters of the neural network.
Fig. 4