Audience selection based on natural language

Technical Field

[0001] This disclosure generally relates to methods and systems for the selection of an audience for targeted message communication.

Background

[0002] This background description is provided for the purpose of generally presenting the context of the disclosure. Contents of this background section are neither expressly nor impliedly admitted as prior art against the present disclosure.

[0003] A significant volume of activity data is generated by individuals through their everyday online interactions. For example, individuals using e-commerce services or ride-hailing services or other online services generate a significant volume of data through their transactions or interactions on such platforms or online services. The significant volume of data comprises meaningful insights about the preferences and behaviours of such individuals. The extraction of such meaningful insights presents a significant technical challenge. This technical challenge is exacerbated by the continuous generation of such activity data and new sources of such activity data. The technical challenge is also exacerbated by the challenges of scalable infrastructure required to effectively process the activity data being generated continuously to generate meaningful inferences.

[0004] The activity data being generated from various sources has varying formats controlled by the systems that generate or capture the activity data. The lack of consistency further complicates the ability to effectively analyze the activity data and generate actionable inferences. Often because of the complexity of the activity data, various systems use tags assigned to the activity data to assist the analysis. However, the accuracy and effectiveness of the analysis is limited by the accuracy and completeness of the tags assigned to the activity records. With the nature of online commerce changing and the steady creation of new sources of activity data, any tag-based system quickly becomes outdated and inaccurate for analysis of activity data.

Summary

[0005] <To be completed after the claims are approved>.

Brief Description of the Drawings

[0006] Some embodiments of systems and methods for target audience selection or for training models for audience targeting, in accordance with the present disclosure, will now be described, by way of non-limiting example only, with reference to the accompanying drawings in which:

[0007] Figure 1 illustrates a block diagram of an overview of the audience selection system;

[0008] Figure 2 illustrates a part of a model for audience selection;

[0009] Figure 3 illustrates a block diagram of a part of a system for audience selection;

[0010] Figure 4 illustrates a user interface for receiving input regarding audience selection; and

[0011] Figure 5 illustrates a block diagram of a system for audience selection and some associated components;

[0012] Figure 6 illustrates a flowchart of a method for audience selection; and

[0013] Figure 7 illustrates a flowchart of a method for training models for audience section.

Detailed Description

[0014] Embodiments relate to systems and methods for audience selection or targeting from a pool of audiences. In particular, embodiments allow the selection of an audience subset based on natural language description of the desired characteristics of the audience. For example, a natural language description is "A young urban food enthusiast, enjoys spicy dishes, often orders from Mala restaurant, and ranks in the top 10% of monthly spenders in their city" (220 in Figure 2). The audience selection is performed from a pool of prospective audiences. Results of the audience selection include a list of individuals among the audience pool that fit the natural language description. Some embodiments also allow the selection of an audience subset using a combination of natural language description and one or more seed audience identifiers.

[0015] Figure 1 illustrates a high-level schematic diagram of the operations of system 500 according to the disclosure. The natural language text 104 summarized an intended audience of interest. The intended audience may be of interest to an advertiser. In some embodiments, the system 500 may also be provided details of a seed audience 106. The system 500 processes the received information to select a target audience 101 that is a subset of the larger audience pool 102. System 500 may provide a user interface to receive inputs 104 and 106. A user computing device 540 (illustrated in Figure 5) communicates with the system 500 through a user interface exemplified in Figure 4. In user interface region 402, the user may enter the natural language description 104 of the target audience. In region 404, the user may provide details of the seed audience. The details of the seed audience may be provided using a file with the seed audience identity details populated in a predefined format.

[0016] The embodiments advantageously allow users with lesser technical competence to perform insightful queries on complex databases of user activity records and generate meaningful insights or subsets of audiences of interest based on the query. Audience selection tasks would otherwise require a thorough and time-intensive analysis of various complex databases, conception of query parameters, testing of sample queries, validation of test results, and iterative re-execution and adaptation of queries and combinations of results from various databases. All of the noted steps require significant technical expertise. Several conventional audience selection systems also require effective tagging of activity records to effectively operate. The tagging operations introduce further overheads in audience selection operations. Inaccuracy in tagging operations may result in inaccurate audience selection outcomes. The disclosed embodiments address one or more of such problems by providing a description encoder model 522 that transforms a natural language description of the characteristics of an intended audience into a vector representation/embedding. In addition, the embodiments also provide an activity encoder model 524 that transforms the audience activity records into a vector representation/embedding. Both the description encoder model and the activity encoder model generate embeddings in a common vector space. This allows the embeddings generated by the description encoder model to be compared with pre-generated embeddings based on audience activity records. The comparison allows the identification of target audiences based on the natural language description.

[0017] Figure 5 illustrates a block diagram of a system 500 for audience selection and some related components. An audience activity database 530 comprises records 531

associated with the activity of audiences. Each audience may be uniquely identified based on their contact details (email, phone number, address etc.) or their identity records (unique identification number such as a driving license number, passport number etc.).

[0018] The audience activity records include records of transactions performed by the audience. The audience activity records include records of transactions performed by the audience. The transactions may include transactions performed by the audience one-commerce platforms, food delivery platforms, ridesharing platforms etc. The audience activity records can also include the point of interest (POI), which refer to specific locations or destinations potentially appealing to a user. Examples of POIs include restaurants, landmarks, tourist attractions, and frequent pick-up or drop-off sites. The transactions may include transactions performed by the audience on e-commerce platforms, food delivery platforms, ridesharing platforms etc. The transaction records may include clickstream records of the interactions of the audience with a website or an app. Clickstream records include details of elements that the audience had interacted with on a website or an app. The elements may include merchants or products listed on the website or app. The clickstream events may include clicking over elements or viewing actions related to elements etc. The clickstream event data includes information relating to elements the audience has interacted with. For example, if the audience had clicked an item in a marketplace the clickstream event data includes details of the item clicked on by the user. In some embodiments, the audience activity database comprises records associated with the user's ridesharing activities. The records may include details of the points of origin or destination of the user. The various real-world objects/events, such as products or destinations referred to in the audience activity database are represented in the form of vector representations or embeddings that encode information of the objects/events referred to in each record. Records in the audience activity database are progressively updated as more information regarding the audience becomes available. The records associated with an individual in the audience activity collective represent the preferences and interests of the individual. While such records are available in varying formats, with distinctive sets of attributes and originate from different sources, such records comprise valuable information that is processed by the activity encoder model 524 to generate audience activity embedding. Audience activity records may also be referred to as audience behaviour records.

[0019] Figure 2 illustrates a schematic example of audience activity records. Records 202 relate to merchants clicked or purchased by an individual. Collectively, multiple items

are represented as an embedding 203 based on the attribute information of individual items purchased. This consolidation can be achieved, for example, by pooling merchant representations for an individual. For example, in 202, each merchant that an individual interacts with has a representation – e.g. vector or matrix representation – determined *a priori* (e.g., derived from pre-training techniques). Subsequent to this, standard deep neural network methodologies like sum pooling, average pooling, or attention pooling can be employed to formulate a consolidated representation – e.g. consolidated vector representation or consolidated matrix representation – referred to as 203. Records 204 relate to items from an individual's clickstream events, app interactions or purchase behaviors. Collectively, multiple clickstream events or app interactions are represented as an embedding 205. Records 206 relate to locations or points of interest associated with the ride-hailing activities of an individual. Collectively, multiple points of interest are represented as an embedding 207 based on the attribute information of individual points of interest.

[0020] Based on the embeddings 203, 205 and 207, the trained activity encoder model 524 generates an audience activity/behaviour embedding 232. Similarly, based on the natural language text 220, the description encoder model 522 generates an audience description embedding 234. The audience activity embeddings are generated for all individuals in the audience pool 102. The description encoder model 522 and the activity encoder model 524 are trained to generate the embeddings in a common vector space. The training methodology has been described with reference to Figure 7. The embeddings are generated such that the description embedding is close in the vector space to the activity embeddings of individuals that fit the intended target audience on the natural language description 220. A similarity analysis between the audience description embedding and the audience activity embeddings is performed to identify individuals in the target audience subset within the audience pool.

[0021] Figure 3 illustrates a block diagram of an embodiment wherein the target audience is identified using a combination of seen audience information 106 and natural language text. Seed audience embedding 302 is retrieved from the audience embedding database 532 based on the seed audience information 106. The seed audience embedding 302 and the description embedding 304 are combined to perform similarity analysis with the audience embedding database 532 to identify the target audience 101.

[0022] Figure 6 illustrates a flowchart of a method of target audience identification using system 500. At step 610, system 500 receives natural language text 104. The natural language text comprises a description of the intended target audience. The description may include adjectives describing the target audience or numeric criteria to specifically target a particular type of audience. In some embodiments, the system 500 may also receive at step 610 one or more seed audience records 106. The description encoder model 522 processes the natural language text 104 to generate a target audience description embedding at step 620. In embodiments wherein the input at step 610 comprises seed audience identifiers, the system retrieves corresponding seed audience embedding from the audience embedding database 532. System 500 then generates a composite embedding based on the target audience description embedding and the retrieved embeddings corresponding to the seed audiences. The composite embedding may be generated by an averaging operation on the target audience description embedding and the retrieved embeddings corresponding to the seed audiences.

[0023] After obtaining the target audience description embedding or the composite embedding at 620, system 500 performs a similarity analysis at step 630 to identify the target audience. The similarity analysis is performed with respect to the records in the audience embeddings database. The similarity analysis may be performed by computing a similarity metric which may include: Euclidean distance, cosine distance or dot product etc. The outcome of the similarity analysis allows the identification of records on the audience embeddings database 532 that are most similar to the target audience description embedding or the composite embedding determined at 620. The records with the greatest similarity values are included in a target audience subset. A calibrated similarity threshold may be employed to define a similarity cutoff and narrow down the target audience subset. At step 640, a message directed to the identified target audience is transmitted by system 500. The system 500 may have access to details of corresponding audience computing devices 550 to direct targeted messages to the target audience. In some embodiments, the targeted messages may comprise recommendations or suggestions of prospective interest to the target audience. The targeted messages may be displayed on a user interface or display of the audience computing device. In embodiments wherein the audience computing device is a smartphone, the targeted messages may be presented as advertisements within a smartphone application.

[0024] Figure 7 illustrates part of a flowchart for training the Description Encoder Model 522 to process natural language text and generate an embedding in a common vector space as the embeddings in the audience embedding database. The training is driven using a keyword generation model 528. The keyword generation model may be a rule-based model comprising a plurality of rules represented in the form of propositional and/or conditional logic rules that are applied to the audience activity records 531 to generate keywords 702. The keyword generation model may comprise a plurality of predefined keywords and based on the outcome of the application of the rules, a subset of keywords are assigned to an individual. The rule-based method involves creating a set of predefined rules to analyze and categorize user behaviors. These rules are typically based on patterns or specific criteria that are observed in the data. To that end, the rules may be produced manually, or may be automatically produced by analyzing parameters describing user behaviours and extracting parameters that are most consistently determinative of user behavior, such as user characteristics or demographic information (e.g. age, home suburb, sex) frequency of use of a service, location when service is used, type or subtype of service used (e.g. a 'type' may be a car booking or food order, and 'subtype' may be standard or premium car or one of various non-overlapping price ranges of restaurants). When user behavior matches a particular rule, a corresponding keyword or label is generated. For a food ordering and car booking platform, these rules can be based on: food preferences: analyze the type and frequency of food orders to determine user's food preferences; travel patterns: monitor the frequency and destinations of car bookings to understand user's travel habits; location-based insights: identify the areas where users frequently order food or book rides to understand their living or working zones; time-based patterns: check the time of day users typically order food or book rides to determine their routine; loyalty to restaurants or locations: track repeated orders from the same restaurant or repeated rides to the same location. For example, after processing audience activity records, the following keywords may be generated for an exemplary individual: "frequently ordering fast food, male, 23 years old, living in CBD." The generated keywords 702 are processed by a text generation model 526 to generate an audience persona description 704. The text generation model 526 generates natural language text describing an individual based on the received keywords. Continuing the above example, a description may be "This user is a young, 23-year-old male who resides in the bustling Central Business District. He showcases a clear preference for fast food, often opting for quick and convenient meal options. Living in the heart of the city, he frequently uses our platform to order his favorite fast-food items and book rides to

navigate through his urban environment. His patterns suggest a fast-paced lifestyle typical of young professionals in metropolitan areas." The text generation model 526 may be implemented using a generative adversarial network (GAN) based model, or a Markov process-based model or any other alternative machine learning model for generating natural language text. In some embodiments, the GPT-4 LLM may be used as the text generation model. The audience persona description 704 serves as a proxy for modelling the natural language description 104 received by the system 500 after it is trained and deployed. During training, the audience persona description 704 serves as training input for the description encoder model.

[0025] The activity encoder model 524 and the description encoder model 522 are jointly trained such that the embeddings generated by the two models for single individual align with each other. In other words, the two models are trained such that the audience description embedding 706 and the audience activity embedding 708 align for a common individual. In some embodiments, the training may be performed using contrastive learning. The audience description embedding 706 and the audience activity embedding 708 in effect operate as augmentations of the audience activity data associated with specific individuals. The augmentations allow the contrasting learning of the activity encoder model 524 and the description encoder model 522. Contrastive learning is a machine learning technique used to learn the general features of a dataset without labels by teaching the model which data points are similar or different. The dataset under consideration for this disclosure is the audience activity database 530 which comprises records of activities of the pool of audiences. The general features relate to the embeddings in the common vector space generated by both the activity encoder model 524 and the description encoder model 522.

[0026] The samples/training dataset used for contrastive learning may be represented in the form of (<candidate audience activity embedding>, <candidate audience description embedding>, <similarity indicator>). The candidate audience activity embedding and candidate audience description embedding values are embedding values generated by an untrained activity encoder model 524 and an untrained description encoder model 522. The similarity metric value is set to 1 for samples where the <candidate audience activity embedding> and <candidate audience description embedding> relate to a common individual. For samples where the <candidate audience activity embedding> and <candidate audience description embedding> relate to different individuals, the similarity metric value is set to 0. Other alternative structures or formats

for arranging the training samples/dataset may be employed by the embodiments to achieve the goal of contrastive learning. Contrastive learning is performed by computing a contrastive loss function and progressively updating the weights of the activity encoder model 524 and the description encoder model 522 to optimize the contrastive loss. Optimization of contrastive loss comprises minimizing a distance metric between <candidate audience activity embedding> and <candidate audience description embedding> for embeddings originating from a common individual. Optimization of contrastive loss also comprises maximizing a distance metric between <candidate audience activity embedding> and <candidate audience description embedding> for embeddings originating from different individuals.

[0027] The disclosed embodiments may be deployed as part of a marketing or advertising management platform that allows the direction of specific marketing content such as web-based marketing content or emails or messages to a selected audience based on natural language input. The embodiments advantageously allow the provision of a more personalized marketing experience for the recipients of the targeted messages. Some embodiments also advantageously allow the combination of natural language text-based audience description with details of a seed audience. The combination of criteria associated with an intended target audience stored in two modalities allows the more accurate identification of the target audience. In addition, the capability of targeting audiences based on natural language text allows the identification of a target audience without the need to fully understand or review the complex or varied database architecture of the audience activity database. The embodiments empower laypersons to perform complex database segmentation or audience identification tasks that would otherwise require advanced database querying expertise. The embodiments also allow iterative refinement of the natural language description of the target audience. The embodiments may present a preview or a summary of target audience results associated with a natural language description. Based on the preview or summary, the natural language description may be refined to optimize the intended target audience.

[0028] The disclosed embodiments can also be advantageously used for clustering audiences based on descriptive text. Large pools of audiences can be segmented into different target groups based on descriptive natural language text. The embodiments allow efficient and intuitive segmentation of audiences into distinct groups. In complex relational or non-relational databases storing audience information, performing segmentation may require a detailed understanding of the underlying structure of the

database. In contrast, the disclosed embodiments allow segmentation of the audience using natural language text that can be written by a layperson with a clear understanding of the underlying nature of the intended audience segments. The embodiments may assist segmentation of the audience for product development or testing purposes such as for A/B testing.

[0029] The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

[0030] Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

[0031] The scope of this disclosure encompasses all changes, substitutions, variations, alterations, and modifications to the example embodiments described or illustrated herein that a person having ordinary skill in the art would comprehend. The scope of this disclosure is not limited to the example embodiments described or illustrated herein. Moreover, although this disclosure describes and illustrates respective embodiments herein as including particular components, elements, feature, functions, operations, or steps, any of these embodiments may include any combination or permutation of any of the components, elements, features, functions, operations, or steps described or illustrated anywhere herein that a person having ordinary skill in the art would comprehend. Although this disclosure describes or illustrates particular embodiments as providing particular advantages, particular embodiments may provide none, some, or all of these advantages.

Claims

1. A system for target audience selection from a pool of prospective audiences, the system comprising:

    one or more processors;

    memory accessible to the one or more processors, the memory comprising instructions executable by the one or more processors to cause the system to:

        receive natural language text comprising a general description of an intended target audience;

        process the natural language model using a description encoder model to obtain a target audience description embedding;

        perform similarity analysis between the target audience description embedding and a database of embedding corresponding to the pool of prospective audiences; and

        identify a target audience subset based on the similarity analysis results.

2. The system of claim 1, wherein the memory comprises further instructions executable by the one or more processors to cause the system to:

    process a plurality of audience activity records using an activity encoder model to generate activity based audience embeddings for each audience in the pool of prospective audiences;

    process the audience activity records using a keyword generation model to generate a plurality of keywords associated with each audience in the pool of prospective audiences;

    generate an audience description for each audience in the pool of prospective audiences by processing the respective plurality of keywords using a text generation model;

process the generated audience descriptions using the description encoder model to obtain a description based audience embedding for each audience in the pool of prospective audiences;

perform contrastive learning to train the description encoder model and the activity encoder model by aligning the activity based audience embeddings and the description based audience embeddings into a common vector space.

3. The system of claim 2, wherein the memory comprises further instructions executable by the one or more processors to cause the system to generate the database of embedding corresponding to the pool of prospective audiences by processing the plurality of historical audience activity records using the activity encoder model.
4. The system of any one of claims 1 to 3, wherein the memory comprises further instructions executable by the one or more processors to cause the system to direct a communication to the identified target audiences.
5. The system of any one of claims 1 to 4, wherein the memory comprises further instructions executable by the one or more processors to cause the system to:

receive identifiers of seed audiences forming part of the intended target audience;

retrieve embeddings corresponding to the seed audiences from the database of embedding corresponding to the pool of prospective audiences;

generate a composite embedding based on the target audience description embedding and the retrieved embeddings corresponding to the seed audiences;

wherein the similarity analysis comprises similarity analysis between the composite embedding and the database of embedding corresponding to the pool of prospective audiences.

6. A system for training a description encoder model for audience targeting, the system comprising:

one or more processors;

memory accessible to the one or more processors, the memory comprising instructions executable by the one or more processors to cause the system to:

process a plurality of historical audience activity records using an activity encoder model to generate activity based audience embeddings for each audience in a pool of prospective audiences;

process the historical audience activity records using a keyword generation model to generate a plurality of keywords associated with each audience in the pool of prospective audiences;

generate an audience description for each audience in the pool of prospective audiences by processing the respective plurality of keywords using a text generation model;

process the generated audience descriptions using the description encoder model to obtain a description based audience embedding for each audience in the pool of prospective audiences;

perform contrastive learning to train the description encoder model and the activity encoder model by aligning the activity based audience embeddings and the description based audience embeddings into a common vector space.

7. The system of any one of claims 2 to 6, wherein the historical audience activity records comprise one or more than one of:

audience transaction records;

audience clickstream records; or

audience travel records.

8. A computer-implemented method for target audience selection from a pool of prospective audiences, the method comprising:

receiving natural language text comprising a general description of an intended target audience;

processing the natural language model using a description encoder model to obtain a target audience description embedding;

performing similarity analysis between the target audience description embedding and a database of embedding corresponding to the pool of prospective audiences; and

identifying a target audience subset based on the similarity analysis results.

9. The method of claim 8, further comprising:

processing a plurality of historical audience activity records using an activity encoder model to generate activity based audience embeddings for each audience in the pool of prospective audiences;

processing the historical audience activity records using a keyword generation model to generate a plurality of keywords associated with each audience in the pool of prospective audiences;

generating an audience description for each audience in the pool of prospective audiences by processing the respective plurality of keywords using a text generation model;

processing the generated audience descriptions using the description encoder model to obtain a description based audience embedding for each audience in the pool of prospective audiences;

performing contrastive learning to train the description encoder model and the activity encoder model by aligning the activity based audience embeddings and the description based audience embeddings into a common vector space.

10. The method of claim 9, further comprising generating the database of embedding corresponding to the pool of prospective audiences by processing the plurality of historical audience activity records using the activity encoder model.
11. The method of any one of claims 8 to 10, further comprising directing a communication to the identified target audiences.
12. The method of any one of claims 8 to 10, further comprising:

receiving identifiers of seed audiences forming part of the intended target audience;

retrieving embeddings corresponding to the seed audiences from the database of embedding corresponding to the pool of prospective audiences;

generating a composite embedding based on the target audience description embedding and the retrieved embeddings corresponding to the seed audiences;

wherein the similarity analysis comprises similarity analysis between the composite embedding and the database of embedding corresponding to the pool of prospective audiences.

13. A computer-implemented method for training a description encoder model for audience targeting, the method comprising:

processing a plurality of historical audience activity records using an activity encoder model to generate activity based audience embeddings for each audience in a pool of prospective audiences;

processing the historical audience activity records using a keyword generation model to generate a plurality of keywords associated with each audience in the pool of prospective audiences;

generating an audience description for each audience in the pool of prospective audiences by processing the respective plurality of keywords using a text generation model;

processing the generated audience descriptions using the description encoder model to obtain a description based audience embedding for each audience in the pool of prospective audiences;

performing contrastive learning to train the description encoder model and the activity encoder model by aligning the activity based audience embeddings and the description based audience embeddings into a common vector space.

14. The method of any one of claims 9 to 13, wherein the historical audience activity records comprise one or more than one of:

audience transaction records;

audience clickstream records; or

audience travel records.

Abstract

System and methods for selection from a pool of prospective audiences by receiving natural language text comprising a general description of an intended target audience; processing the natural language model using a description encoder model to obtain a target audience description embedding; performing similarity analysis between the target audience description embedding and a database of embedding corresponding to the pool of prospective audiences; and identifying a target audience subset based on the similarity analysis results.

Figure 5