(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 June 2016 (02.06.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/086075 A1

(51) International Patent Classification:
G06Q 30/02 (2012.01)

(21) International Application Number:
PCT/US2015/062551

(22) International Filing Date:
24 November 2015 (24.11.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/083,716 24 November 2014 (24.11.2014) US

(71) Applicant: THE NIELSEN COMPANY (US), LLC [US/US]; 85 Broad Street, New York, NY 10004 (US).

(72) Inventors: CUI, Jingsong; 82 Chicory Lane, Pennington, NJ 08534 (US). DOE, Peter, Campbell; 140 Liberty Street, Ridgewood, NJ 07450 (US).

(74) Agent: JACKLIN, William, E.; Hanley, Flight & Zimmerman, LLC, 150 S. Wacker Drive, Suite 2200, Chicago, IL 60606 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,

[Continued on next page]

(54) Title: METHODS AND APPARATUS TO PREDICT TIME-SHIFTED EXPOSURE TO MEDIA

(57) Abstract: Methods, apparatus, systems and articles of manufacture are disclosed to predict timeshifted exposure to media. An example method includes normalizing, with a processor, audience measurement data corresponding to media exposure data and social media activity data. The example method also includes building an estimation model based on a relationship between a first subset of the normalized audience measurement data associated with a characteristic of the media asset and historical rating lift measurements associated with the media asset. The example method also includes estimating, with the processor, current ratings for the media asset based on time-period based ratings and broadcast time-periods. The example method also includes applying data related to the media asset and the estimated current ratings to the estimation model to estimate, with the processor, the ratings lift for the media asset.

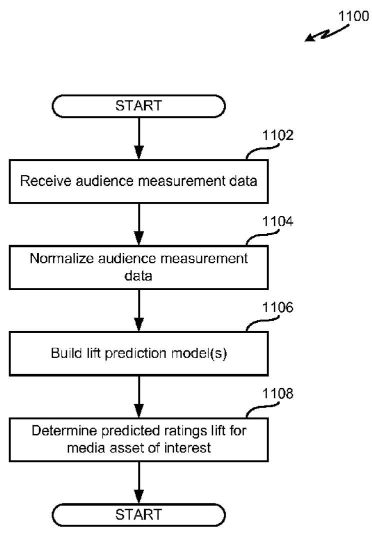


FIG. 11

WO 2016/086075 A1

LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHODS AND APPARATUS TO PREDICT TIME-SHIFTED EXPOSURE TO MEDIA

RELATED APPLICATION

[0001] This patent claims the benefit of, and priority from, U.S. Provisional Patent Application No. 62/083,716, filed November 24, 2014, entitled "Methods and Apparatus to Predict TV Rating Lift." U.S. Provisional Patent Application No. 62/083,716 is hereby incorporated by reference in its entirety.

FIELD OF THE DISCLOSURE

[0002] This disclosure relates generally to audience measurement, and, more particularly, to methods and apparatus to predict time-shifted exposure to media.

BACKGROUND

[0003] Audience measurement of media (e.g., content and/or advertisements presented by any type of medium such as television, in theater movies, radio, Internet, etc.) is typically carried out by monitoring media exposure of panelists that are statistically selected to represent particular demographic groups. Audience measurement companies, such as The Nielsen Company (US), LLC, enroll households and persons to participate in measurement panels. By enrolling in these measurement panels, households and persons agree to allow the corresponding audience measurement company to monitor their exposure to information presentations, such as media output via a television, a radio, a computer, etc. Using various statistical methods, the collected media exposure data is processed to determine the size and/or demographic composition of the audience(s) for media of interest. The audience size and/or demographic information is valuable to, for example, advertisers, broadcasters, content providers, manufacturers, retailers, product developers, and/or other entities. For example, audience size and demographic information is a factor in the placement of advertisements, in valuing commercial time slots during a particular program and/or generating ratings for piece(s) of media.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] FIG. 1 illustrates an example system for audience measurement analysis implemented in accordance with the teachings of this disclosure to predict time-shifted exposure to media.

[0005] FIG. 2 is an example data table that may be used by the example central facility of FIG. 1 to store raw data variables in the example raw data database of FIG. 1.

[0006] FIG. 3 is an example calculation of estimated current television ratings that may be performed by the example program ratings estimator of FIG. 1.

[0007] FIG. 4 is an example block diagram of an example implementation of the data translator of FIG. 1.

[0008] FIG. 5 is an example data table that may be used by the example data translator of FIGS. 1 and/or 4 to translate ratings data variables.

[0009] FIG. 6 is an example calculation of estimated current television ratings that may be performed by the example ratings handler of FIG. 4.

[0010] FIG. 7 is an example data table that may be used by the example data translator of FIGS. 1 and/or 4 to translate program attributes data variables.

[0011] FIG. 8 is an example data table that may be used by the example data translator of FIGS. 1 and/or 4 to translate social media data variables.

[0012] FIG. 9 is an example schema that may be used by the example central facility of FIG. 1 to determine lift prediction models based on season-episode groupings.

[0013] FIG. 10 is an example data table that may be used by the example central facility of FIG. 1 to associate translated data variables to corresponding lift prediction models.

[0014] FIG. 11 is a flowchart representative of example machine-readable instructions that may be executed by the example central facility of FIG. 1 to predict time-shifted exposure to media.

[0015] FIG. 12 is a flowchart representative of example machine-readable instructions that may be executed by the example data translator of FIGS. 1 and/or 4 to normalize raw audience measurement data.

[0016] FIG. 13 is a flowchart representative of example machine-readable instructions that may be executed by the example model builder of FIG. 1 to build a lift prediction model.

[0017] FIG. 14 is a flowchart representative of example machine-readable instructions that may be executed by the example program ratings estimator of FIG. 1 to predict time-shifted exposure to media.

[0018] FIG. 15 is a block diagram of an example processing platform structured to execute the example machine-readable instructions of FIGS. 11-13 and/or 14 to implement the example central facility and/or the example data translator of FIGS. 1 and/or 4.

[0019] Wherever possible, the same reference numbers will be used throughout the drawing(s) and accompanying written description to refer to the same or like parts.

DETAILED DESCRIPTION

[0020] Examples disclosed herein facilitate predicting time-shifted exposure to media. For example, disclosed examples enable prediction of television ratings for households tuned to (or persons exposed to) a program or time-period of a specific station or cable network at the actual time that the program was telecast (e.g., live ratings), and any delayed viewing of the program during a predetermined time period. For example, "live + 7" ratings represent the percentage of an audience that was exposed to media during the original telecast of the media (e.g., the live ratings) and the incremental viewing that takes place during the 7 days following the original telecast.

[0021] Exposure information (e.g., ratings) may be useful for determining a marketing campaign and/or evaluating the effectiveness of a marketing campaign. For example, an advertiser who wants exposure of their asset (e.g., a product, a service, etc.) to reach a specific audience will place advertisements in media (e.g., a television program) whose audience represents the characteristics of the target market. In some examples, networks determine the cost of including an advertisement in their media based on the ratings of the media. For example, a high rating for a television program represents a large number of audience members who tuned to (or were exposed to) the television program. In such instances, the larger the audience of a television program (e.g., a higher rating), the more networks can charge for advertisements during the program.

[0022] Traditionally, live ratings were used as the metric for evaluating the performance of media. However, as the methods for consuming media have evolved, there is an increasing trend in time-shifted viewing of media. For example, it is becoming more common for consumers to watch a television program via a digital video recorder (DVR) or video-on-demand service at a time that is convenient to them. In some instances, time-shifted viewing may account for fifty percent of the viewership of a television program. In such instances, live ratings do not accurately represent the audience size of the television program.

[0023] "Live + 7" ratings more accurately represent how viewers are watching media. "Live + 7" ratings reflect viewing done via DVR services and/or video-on-demand services within seven days (e.g., 168 hours) of the original telecast of the program. In some instances, the additional viewership that is included in "live + 7" ratings can make or break whether a low profile program is renewed or cancelled. However, measuring "live + 7" ratings involves waiting at least seven days to determine the time-shifted viewing.

[0024] Examples disclosed herein facilitate predicting time-shifted viewing of media. Examples disclosed herein collect and/or develop audience measurement data related to media. For example, disclosed examples measure audience exposure to television programs via people meters and/or by monitoring social media messages referencing television programs. Examples disclosed herein use the audience measurement data to build and/or train lift prediction models that may be used to estimate ratings lift. For example, disclosed examples determine a relationship between features of the audience measurement data and historical ratings lifts. As used herein, features of the audience measurement data include, but are not limited to, current and historical ratings-related information, social media indicators, characteristics of the media asset, etc.

[0025] Examples disclosed herein apply data related to a media asset of interest to a lift prediction model to determine a predicted lift for the media asset. In some disclosed examples, different lift prediction models are developed based on characteristics of the media asset. For example, disclosed examples may characterize media assets based on season-episode groupings. In some such examples, the data applied to a lift prediction model to determine estimated ratings lift may depend on the season-episode grouping.

[0026] FIG. 1 is a diagram of an example environment in which a system 100 constructed in accordance with the teachings of this disclosure operates to predict time-shifted exposure to media of interest. The example system 100 of FIG. 1 includes one or more audience measurement system(s) 105 and a central facility 125 to facilitate predicting time-shifted exposure to media of interest in accordance with the teachings of this disclosure. In the illustrated example of FIG. 1, the central facility 125 estimates an increase in viewership (e.g., a "lift") over a period of seven days after the initial (e.g., live) broadcast of media of interest (e.g., a television show). However, any other time-shifted period may additionally or alternatively be used.

[0027] The example system 100 of FIG. 1 includes the one or more audience measurement system(s) 105 to collect audience measurement data 110 from panelists and non-panelists. The example audience measurement system(s) 105 of FIG. 1 collect panelist media measurement data 110A via, for example, people meters operating in statistically-selected households, set-top boxes and/or other media devices (e.g., such as digital video recorders, personal computers, tablet computers, smartphones, etc.) capable of monitoring and returning monitored data for media presentations, etc. The example panelist media measurement data 110A of FIG. 1 includes media exposure data such as live exposure data, delayed exposure data (e.g., relative to time-shifted viewing of media via, for example, a digital video recorder and/or video on-demand), media performance data such as TV ratings (e.g., historical TV ratings and current TV ratings), program characteristics (e.g., attributes) such as broadcast day-of-week information, broadcast time information, originator information (e.g., a network or channel that broadcast the media) and/or genre information, etc. In some examples, the panelist media measurement data 110A is associated with demographic information (e.g., gender, age, income, etc.) of the panelists exposed to the media.

[0028] As used herein, the term "media" includes any type of content and/or advertisement delivered via any type of distribution medium. Thus, media includes television programming or advertisements, radio programming or advertisements, movies, web sites, streaming media, etc.

[0029] Example methods, apparatus, and articles of manufacture disclosed herein monitor media presentations at media devices. Such media devices may include, for example, Internet-enabled televisions, personal computers, Internet-enabled mobile handsets (e.g., a smartphone), video game consoles (e.g., Xbox®, PlayStation®), tablet computers (e.g., an iPad®), digital media players (e.g., a Roku® media player, a Slingbox®, etc.), etc. In some examples, media monitoring information is aggregated to determine ownership and/or usage statistics of media devices, relative rankings of usage and/or ownership of media devices, types of uses of media devices (e.g., whether a device is used for browsing the Internet, streaming media from the Internet, etc.), and/or other types of media device information. In examples disclosed herein, monitoring information includes, but is not limited to, media identifying information (e.g., media-identifying metadata, codes, signatures, watermarks, and/or other information that may be used to identify presented media), application usage information (e.g., an identifier of an application, a time and/or duration of use of the application, a rating of the application, etc.), and/or user-identifying information (e.g., demographic information, a user identifier, a panelist identifier, a username, etc.).

[0030] The example audience measurement system(s) 105 of FIG. 1 also collect social media activity data 110B related to media via, for example, social media servers that provide social media services to users of the social media server. As used herein, the term social media services is defined to be a service provided to users to enable users to share information (e.g., text, images, data, etc.) in a virtual community and/or network. Example social media services may include, for example, Internet forums (e.g., a message board), blogs, micro-blogs (e.g., Twitter®), social networks (e.g., Facebook®, LinkedIn, Instagram, etc.), etc. For example, the audience measurement system(s) 105 may monitor social media messages communicated via social media services and identify media-exposure social media messages (e.g., social media messages that reference at least one media asset (e.g., media and/or a media event)). The example audience measurement system(s) 105 may filter the media-exposure social media messages for media-exposure social media messages of interest (e.g., social media messages that reference media of interest).

[0031] The example social media activity data 110B of FIG. 1 includes one or more of message identifying information (e.g., a message identifier, a message author, etc.), timestamp information indicative of when the social media message was posted and/or viewed, the content of the social media message and an identifier of the media asset referenced in the media-exposure social media message. In some examples, the audience measurement system(s) 105 may process the media-exposure social media messages of interest and aggregate information related to the social media messages. For example, the audience measurement system(s) 105 may determine a count of the media-exposure social media messages of interest, may determine a number of unique authors who posted the media-exposure social media messages of interest and/or may determine a number of impressions of (e.g., exposure to) the media-exposure social media messages of interest.

[0032] In the illustrated example of FIG. 1, the audience measurement system(s) 105 send the audience measurement data 110 to the central facility 125 via an example network 115. The example network 115 of the illustrated example of FIG. 1 is the Internet. However, the example network 115 may be implemented using any suitable wired and/or wireless network(s) including, for example, one or more data buses, one or more Local Area Networks (LANs), one or more wireless LANs, one or more cellular networks, one or more private networks, one or more public networks, etc. The example network 115 enables the central facility 125 to be in communication with the audience measurement system(s) 105. As used herein, the phrase “in communication,” including variances therefore, encompasses direct communication and/or indirect communication through one or more intermediary components and does not require direct physical (e.g., wired) communication and/or constant communication, but rather includes selective communication at periodic or aperiodic intervals, as well as one-time events.

[0033] In the illustrated example, the central facility 125 is operated by an audience measurement entity (AME) 120. The example AME 120 of the illustrated example of FIG. 1 is an entity such as The Nielsen Company (US), LLC that monitors and/or reports exposure to media and operates as a neutral third party. That is, in the illustrated example, the audience measurement entity 120 does not provide media (e.g., content and/or advertisements) to end users. This un-involvement with the media production and/or delivery ensures the neutral status of the audience measurement entity 120 and, thus, enhances the trusted nature of the data the AME 120 collects and processes. The reports generated by the audience measurement entity (sometimes referred to as an "audience analytics entity" (AAE)) may identify aspects of media usage such as the number of people who are watching television programs and characteristics of the audiences (e.g., demographic information of who is watching the television programs, when they are watching the television programs, etc.).

[0034] The example AME 120 of FIG. 1 operates the central facility 125 to predict time-shifted exposure to media such as a lift in ratings associated with a media asset of interest over a course of 7 days (e.g., 168 hours) after the initial broadcast of the media asset. As used herein, a media asset of interest is a particular media program (e.g., identified via an episode number and season number) that is being analyzed (e.g., for a report). For example, a first media asset of interest may be episode 3 of season 2 of a program "Sports Stuff" and a second media asset of interest may be episode 4 of season 2 of the program "Sports Stuff." The central facility 125 of the illustrated example includes a server and/or database that collects and/or receives audience measurement data related to media assets (e.g., media and/or media events) and predicts time-shifted (or delayed) exposure to media assets of interest.

[0035] In some examples, the central facility 125 is implemented using multiple devices and/or the audience measurement system(s) 105 is (are) implemented using multiple devices. For example, the central facility 125 and/or the audience measurement system(s) 105 may include disk arrays and/or multiple workstations (e.g., desktop computers, workstation servers, laptops, etc.) in communication with one another. In the illustrated example, the central facility 125 is in communication with the audience measurement system(s) 105 via one or more wired and/or wireless networks represented by the network 115.

[0036] The example central facility 125 of the illustrated example of FIG. 1 processes the audience measurement data 110 returned by the audience measurement system(s) 105 to predict time-shifted exposure to media. For example, the central facility 125 may process the audience measurement data 110 to determine a relationship between features (sometimes referred to herein as "variables," "predictors" or "factors") of the audience measurement data 110 and measured lift to build one or more prediction models. The example central facility 125 may then apply data associated with a media asset of interest to a prediction model to determine a predicted lift for the media asset.

[0037] In the illustrated example of FIG. 1, the central facility 125 includes an example data interface 130, an example raw data database 135, an example data translator 140, an example translated data database 145, an example model builder 150, an example models data store 155 and an example program ratings estimator 160. In the illustrated example of FIG. 1, the example central facility 125 includes the example data interface 130 to provide an interface between the network 115 and the central facility 125. For example, the data interface 130 may be a wired network interface, a wireless network interface, a Bluetooth® network interface, etc. and may include the associated software and/or libraries needed to facilitate communication between the network 115 and the central facility 125. In the illustrated example of FIG. 1, the data interface 130 receives the audience measurement data 110 returned by the example audience measurement system(s) 105 of FIG. 1. The example data interface 130 records the audience measurement data 110 in the example raw data database 135.

[0038] In the illustrated example of FIG. 1, the example central facility 125 includes the example raw data database 135 to record data (e.g., the example audience measurement data 110) provided by the audience measurement system(s) 105 via the example data interface 130. An example data table 200 of the illustrated example of FIG. 2 illustrates example raw data variables that may be recorded in the example raw data database 135. The example raw data database 135 may be implemented by a volatile memory (e.g., a Synchronous Dynamic Random Access Memory (SDRAM), Dynamic Random Access Memory (DRAM), RAMBUS Dynamic Random Access Memory (RDRAM), etc.) and/or a non-volatile memory (e.g., flash memory). The example raw data database 135 may additionally or alternatively be implemented by one or more double data rate (DDR) memories, such as DDR, DDR2, DDR3, mobile DDR (mDDR), etc. The example raw data database 135 may additionally or alternatively be implemented by one or more mass storage devices such as hard disk drive(s), compact disk drive(s), digital versatile disk drive(s), etc. While in the illustrated example the raw data database 135 is illustrated as a single database, the raw data database 135 may be implemented by any number and/or type(s) of databases.

[0039] The example central facility 125 of the illustrated example of FIG. 1 combines multiple disparate data sets to enable modeling and assessment of multiple inputs simultaneously. As described below, at least some of the variables are translated (e.g., modified and/or manipulated) from their raw form to be more meaningfully handled when building the prediction models and estimating the ratings lift. For example, raw data may be multiplied, aggregated, averaged, etc., and stored as translated data (sometimes referred to herein as “sanitized,” “normalized” or “recoded” data) prior to generating the prediction models used to estimate ratings lift.

[0040] In the illustrated example of FIG. 1, the example central facility 125 includes the example data translator 140 to translate the audience measurement data 110 received from the example audience measurement system(s) 105 into a form more meaningfully handled by the example model builder 150. For example, the data translator 140 of FIG. 1 may retrieve and/or query the audience measurement data 110 recorded in the example raw data database 135 and normalize the disparate data to a common scale. As described in detail below, the example data translator 140 modifies and/or manipulates audience measurement data 110 based on the type of data. For example, the data translator 140 may translate (e.g., map) data that is a string data type (e.g., “Day-of-Week” is “Tuesday”) to a Boolean data type (e.g., “Day Tues” is set to true (e.g., “1”)).

[0041] As described above and in connection with the example data table 200 of FIG. 2, the audience measurement data 110 may be in different data formats and/or different units of measure. For example, program characteristic information such as program title, episode and season identifying information, day of week, broadcast time, broadcast network and genre may be stored as string data types. Current and historical ratings information may be represented via television rating scores (e.g., floating data types). Social media indicators (e.g., message identifiers, message timestamps, message content, message author identifiers, message impression information, etc.) may be represented as string data types. In the illustrated example of FIG. 1, the data translator 140 normalizes the audience measurement data 110 into numerical data types (e.g., Boolean data types, integer data types and/or floating data types). The example data translator 140 of FIG. 1 records translated data in the example translated data database 145.

[0042] In the illustrated example of FIG. 1, the example central facility 125 includes the example translated data database 145 to record translated data provided by the example data translator 140. Example data tables 500, 600, 700 and 800 of the illustrated examples of FIGS. 5, 6, 7 and 8, respectively, illustrate example translated data variables that may be recorded in the example translated data database 145. The example translated data database 145 may be implemented by a volatile memory (e.g., an SDRAM, DRAM, RDRAM, etc.) and/or a non-volatile memory (e.g., flash memory). The example translated data database 145 may additionally or alternatively be implemented by one or more DDR memories, such as DDR, DDR2, DDR3, mDDR, etc. The example translated data database 145 may additionally or alternatively be implemented by one or more mass storage devices such as hard disk drive(s), compact disk drive(s), digital versatile disk drive(s), etc. While in the illustrated example the translated data database 145 is illustrated as a single database, the translated data database 145 may be implemented by any number and/or type(s) of databases.

[0043] In the illustrated example of FIG. 1, the central facility 125 includes the example model builder 150 to build one or more prediction model(s) that may be used to estimate ratings lift for media assets of interest over a time-shifted period (e.g., seven days after the initial broadcast of media assets of interest). In the illustrated example, the model builder 150 determines a relationship between one or more translated data variables retrieved from the example translated data database 145 and ratings lifts to build one or more prediction model(s). Equation 1 below is an example equation representative of a prediction model that may be built by the example model builder 150.

Equation 1: $$\text{lift} = (100\text{-Live Ratings}) * \frac{\exp(\sum X_i * a_i)}{1 + \exp(\sum X_i * a_i)}$$

[0044] In Equation 1 above, the lift (*lift*) for a media asset of interest is calculated using live ratings information (*Live Ratings*) associated with the media asset of interest (e.g., current ratings) and values of translated data variables (X_i) that are modified by coefficients (a_i). In the illustrated example, the model builder 150 applies historical values of lift and live ratings data, and historical values of one or more translated data variables from the translated data database 145 to Equation 1 above to train the model of Equation 1 to determine the value of the coefficients (a_i). The example model builder 150 of FIG. 1 uses logistic regression to determine the value of the coefficients (a_i). However, any other technique may additionally or alternatively be used to determine the value of the coefficients (a_i). For example, a linear equation (represented by Equation 2 below) may additionally or alternatively be used.

Equation 2: $$\text{lift} = (100\text{-Live Ratings}) * \sum (X_i * a_i)$$

[0045] In the illustrated example of FIG. 1, the example model builder 150 builds different models to estimate ratings lift for a media asset of interest based on season and episode number attribute of the media asset of interest. For example, the model builder 150 may apply different sets of the translated data variables to Equation 1 above to determine the coefficient values (a_i) of the translated data variables (X_i) based on season-episode groupings. An example schema 900 of the illustrated example of FIG. 9 illustrates example ratings lift models based on season-episode groupings.

[0046] In the illustrated example of FIG. 1, the example model builder 150 selects the translated data variables (X_i) to apply to Equation 1 above based on season and episode number attributes of media assets. For example, the model builder 150 may build a first lift prediction model by applying data related to episodes one through three of the first season of a media asset to Equation 1 above. The example model builder 150 may build a second lift prediction model by applying data related to episodes four through six of the first season of a media asset and/or data related to episodes one through six of subsequent seasons of the media asset to Equation 1 above. The example model builder 150 may build a third lift prediction model by applying data related to episodes seven and later of any season of a media asset to Equation 1 above. In some such instances, while the general form of Equation 1 is used to build lift prediction models, the translated data variables included in the corresponding models is different and, as a result, the coefficient values (a_i) of the translated data variables (X_i) differ between the three models. While the illustrated example describes three example season-episode groupings, any other number of season-episode groupings may additionally or alternatively be used. The example model builder 150 of FIG. 1 stores the generated lift prediction models in the example models data store 155.

[0047] In the illustrated example of FIG. 1, the example central facility 125 includes the example models data store 155 to store lift prediction models generated by the example model builder 150. The example models data store 155 may be implemented by a volatile memory (e.g., SDRAM, DRAM, RDRAM, etc.) and/or a non-volatile memory (e.g., flash memory). The example models data store 155 may additionally or alternatively be implemented by one or more DDR memories, such as DDR, DDR2, DDR3, mDDR, etc. The example models data store 155 may additionally or alternatively be implemented by one or more mass storage devices such as hard disk drive(s), compact disk drive(s), digital versatile disk drive(s), etc. While in the illustrated example the models data store 155 is illustrated as a single database, the models data store 155 may be implemented by any number and/or type(s) of databases.

[0048] In the illustrated example of FIG. 1, the central facility 125 includes the example program ratings estimator 160 to use the lift prediction models generated by the example model builder 150 to estimate ratings lift for a media asset of interest. For example, the program ratings estimator 160 may apply data related to the media asset of interest to predict a lift in viewership of the media asset of interest over a period (e.g., within seven days of the original broadcast of the media asset of interest). In the illustrated example, the program ratings estimator 160 uses program characteristics of the media asset of interest to select a lift prediction model to apply. For example, the program ratings estimator 160 may determine a lift prediction model based on the season and episode number associated with the media asset of interest.

[0049] In the illustrated example of FIG. 1, in response to selecting the lift prediction model to apply, the example program ratings estimator 160 retrieves data related to the media asset of interest from the translated data database 145. In some examples, the program ratings estimator 160 estimates current ratings information related to the media asset of interest based on real-time ratings information. For example, fast affiliate reports provide first national ratings for a media asset the day after telecast. In some examples, the fast affiliate ratings (sometimes referred to as "overnight ratings") are time-period information based on the scheduled broadcast time of media assets rather than the actual broadcast time of media assets. For example, fast affiliate ratings for a media asset of interest may include stations that did not air the media asset, that cut away from the media asset for local programming (e.g., local breaking news), etc.

[0050] In the illustrated example of FIG. 1, the example program ratings estimator 160 estimates current ratings information based on a fast affiliate reports and actual broadcast information obtained from, for example, the affiliate stations. For example, the program ratings estimator 160 may adjust fast affiliate ratings for a media asset of interest based on actual broadcast information. FIG. 3 illustrates an example calculation 300 of estimated current TV ratings based on fast affiliate reports 305 and actual broadcast information 310. In the illustrated example, the program ratings estimator 160 combines the fast affiliate reports 305 with the actual broadcast information 310 to estimate current TV ratings 320. Although the example fast affiliate reports 305 provide "live" ratings and "live + same day" ratings for thirty-minute periods, any other time period may additionally or alternatively be used. For example, the fast affiliate reports 305 may provide ratings on a minute-by-minute basis to increase the granularity of the ratings information.

[0051] The example program ratings estimator 160 of the illustrated example of FIG. 1 applies the estimated current ratings and data related to a media asset of interest to the generated lift prediction models stored in the example models data store 155 to generate reports 165 estimating the ratings lift for the media asset of interest. For example, the program ratings estimator 160 may estimate the ratings lift for a media asset of interest by applying program attributes information, social media indicators information and/or media performance information to a lift prediction model. As used herein, program attributes information includes genre information of the media asset of interest, day-of-week information related to the media asset of interest, broadcast time related to the media asset of interest, originator (e.g., network or channel) information related to the media asset of interest, etc. As used herein, social media indicators information includes a social media messages count related to the number of media-exposure social media messages of interest, a social media unique authors count related to the number of unique authors who posted media-exposure social media messages of interest, a social media impressions count related to the number of users who were exposed to the media-exposure social media messages of interest, etc. As used herein, media performance information includes current ratings associated with the media asset of interest (e.g., “live” ratings and “live + same day” ratings), historical ratings associated with the media asset of interest (e.g., historical “live” ratings and historical “live + same day” ratings), program lead-in ratings associated with the media asset of interest (e.g., “live” ratings and “live + same day” ratings of the program that was broadcast prior to the media asset of interest), a program loyalty score associated with the media asset of interest (e.g., average number of episodes of the media asset watched by viewers during a time period (e.g., five weeks)), etc.

[0052] FIG. 2 is an example data table 200 that lists raw audience measurement data variables that the example data interface 130 of FIG. 1 may store in the example raw data database 135 of FIG. 1. In the illustrated example of FIG. 2, the raw audience measurement data variables represent the data collected and/or provided by the audience measurement system(s) 105 of FIG. 1. For example, the raw audience measurement data variables may include the panelist media measurement data 110A collected via, for example, people meters operating in statistically-selected households, set-top boxes and/or other media devices (e.g., such as digital video recorders, personal computers, tablet computers, smartphones, etc.) capable of monitoring and returning monitored data for media presentations, etc. The example raw audience measurement data variables included in the data table 200 may also include the social media activity data 110B associated with media of interest referenced by social media messages collected via, for example, social media servers that provide social media services to users of the social media server.

[0053] The example data table 200 of the illustrated example of FIG. 2 includes a variable name identifier column 205, a variable data type identifier column 210 and a variable meaning identifier column 215. The example variable name identifier column 205 indicates example variables that may be associated with a telecast and/or useful for predicting time-shifted exposure to media. The example variable data type identifier column 210 indicates a data type of the corresponding variable. The example variable meaning identifier column 215 provides a brief description of the value associated with the corresponding variable. While three example variable identifier columns are represented in the example data table 200 of FIG. 2, more or fewer variable identifier columns may be represented in the example data table 200. For example, the example data table 200 may additionally or alternatively include a variable identifier column indicative of the source of the corresponding data (e.g., the example panelist media measurement data 110A, the example social media activity data 110B, etc.).

[0054] The example data table 200 of the illustrated example of FIG. 2 includes eighteen example rows corresponding to example raw audience measurement data variables. The first example block of rows 250 identifies attributes and/or characteristics of a media asset and is stored as strings. For example, the “Title” variable identifies the name of the media asset (e.g., “Sports Stuff”), the “Episode Identifier” variable identifies the season and episode number associated with the media asset (e.g., Season 2, Episode 4), the “Day of Week” variable identifies the day of the week that the media asset was broadcast (e.g., “Tuesday”), the “Broadcast Time” variable identifies the time during which the media asset was broadcast (e.g., “20:00-20:30”), the “Network” variable identifies on which network the media asset was broadcast (e.g., Channel “ABC”), and the “Genre” variable identifies the genre that the media asset is classified (e.g., a “comedy”).

[0055] In the example data table 200 of FIG. 2, the second example block of rows 255 identifies ratings information associated with a media asset and the corresponding information is stored as floating type data. For example, the “Ratings Live” variable identifies the program ratings during the original broadcast of the program (e.g., “1.01”). The example “Ratings LiveSameDay” variable and example “Ratings LivePlus7” variable identify the program ratings during the original broadcast of the program and also include time-shifted incremental viewing that takes place via, for example, a DVR or video-on-demand (VOD) service. For example, the “Ratings LiveSameDay” variable represents the number of people who viewed the media asset during its original broadcast time and/or during the same day as the original broadcast (e.g., “Live + Same Day” ratings). The example "Ratings LivePlus7" variable represents the number of people who viewed the media asset during its original broadcast time and/or within seven days following the original telecast (e.g., "Live + 7" or "L7" ratings). Although the example data table 200 includes three example ratings (e.g., "Live" ratings, "Live + Same Day" ratings and "Live + 7" ratings), in other examples, the ratings information may be associated with more or fewer time-shifted periods.

[0056] In the illustrated example of FIG. 2, the example row 260 indicates the "Lead-In Ratings" variable is stored as a floating data type and corresponds to the program ratings for a media asset preceding the media asset of interest. For example, when processing the "Sports Stuff" program, the "lead-in ratings" information corresponds to the media asset that is broadcast on "Tuesday" at "19:30-20:00" on Channel "ABC."

[0057] In the illustrated example of FIG. 2, the example row 265 indicates the "Panelist ID" variable is stored as a string and uniquely identifies the panelist who provided the viewership information. For example, panelists who are provided people meters may be assigned a panelist identifier to monitor the media exposure of the panelist. In the illustrated example, the panelist identifier is an obfuscated alphanumeric string to protect the identity of the panelist. In some examples, the panelist identifier is obfuscated in a manner so that the same obfuscated panelist identifier information corresponds to the same panelist. In this manner, user activities may be monitored for particular users without exposing sensitive information regarding the panelist. However, any other approach to protecting the privacy of a panelist may additionally or alternatively be used. In some examples, the panelist identifier is used to identify demographic information associated with the panelist. For example, the panelist identifier "0123" may link to demographic information indicating the panelist is a male, age 18-49.

[0058] In the example data table 200 of FIG. 2, the third example block of rows 270 identifies information regarding social media messages. For example, the "Message ID" variable is stored as a string and is a unique identifier of a social media message. In the illustrated example, the example "Message Timestamp" variable is stored as a string data type and identifies the date and/or time when the corresponding social media message was posted. In the illustrated example, the example "Message Content" variable is stored as a string data type and identifies the content of corresponding social media message. In the illustrated example, the example "Message Author" variable is stored as a string data type and identifies the author of the corresponding social media message.

[0059] In the example data table 200 of FIG. 2, the fourth example block of rows 275 represents overnight time-period ratings that may be used to estimate current ratings for a media asset. In the illustrated example, the example "Overnight Live" variable and the example "Overnight LiveSameDay" variable are stored as floating data types and represent "Live" ratings and "Live + Same Day" ratings based on time-periods. For example, the "overnight live" ratings for the media asset "Sports Stuff" may represent the number of viewers who were tuned to Channel "ABC" between "20:00-20:30" on "Tuesday" (e.g., the broadcast time and network when "Sports Stuff" is scheduled to air). However, in certain markets, the media asset may not air at the scheduled time. For example, due to a football game that went into overtime, the media asset "Sports Stuff" may have actually aired between "20:08-20:35." The example "Actual Broadcast" variable 280 is stored as a string and identifies the actual broadcast time of the media asset (e.g., "20:08 – 20:35").

[0060] While eighteen example raw data variables are represented in the example data table 200 of FIG. 2, more or fewer raw data variables may be represented in the example data table 200 corresponding to the many raw audience measurement data variables that may be collected and/or provided by the audience measurement system(s) 105 of FIG. 1.

[0061] FIG. 4 is a block diagram of an example implementation of the data translator 140 of FIG. 1 that may facilitate manipulating and/or modifying raw audience measurement data 110 retrieved from the example raw data database 135. As described above, the example data translator 140 translates the raw audience measurement data 110 to a form that may be meaningfully handled by the example model builder 150 to generate one or more lift prediction models. The example data translator 140 of FIG. 4 includes an example ratings handler 405, an example attributes handler 410 and an example social media handler 415. In the illustrated example, the ratings handler 405, the attributes handler 410 and the social media handler 415 record the translated information in the example translated data database 145 of FIG. 1.

[0062] In the illustrated example of FIG. 4, the example data translator 140 includes the example ratings handler 405 to process ratings-related information representative of media assets. For example, the ratings handler 405 may query and/or retrieve ratings-related information from the raw data database 135 (e.g., current ratings information, historical ratings information, etc.) to translate into a form meaningfully handled by the example model builder 150.

[0063] An example data table 500 of the illustrated example of FIG. 5 illustrates example translated ratings data variables that may be recorded by the ratings handler 405 in the example translated data database 145. The example data table 500 of the illustrated example of FIG. 5 includes a variable name identifier column 505, a variable data type identifier column 510 and a variable meaning identifier column 515. The example variable name identifier column 505 indicates example variables that may be associated with a media asset broadcast and/or useful for predicting time-shifted exposure to the media asset. The example variable data type identifier column 510 indicates a data type of the corresponding variable. The example variable meaning identifier column 515 provides a brief description of the value associated with the corresponding variable. While three example variable identifier columns are represented in the example data table 500 of FIG. 5, more or fewer variable identifier columns may be represented in the example data table 500.

[0064] The example data table 500 of the illustrated example of FIG. 5 includes five example rows corresponding to example translated ratings data variables. The first example row 550 indicates the ratings handler 405 of FIG. 4 stores the “Same Day Lift” variable as a floating data type. In the illustrated example, the ratings handler 405 determines a “Same Day Lift” value associated with a media asset of interest based on the “Ratings LiveSameDay” variable and the “Ratings Live” variable retrieved from the example raw data database 135. For example, the ratings handler 405 may query the raw data database 135 for the “Ratings LiveSameDay” rating and the “Ratings Live” rating related to a media asset of interest (e.g., season 2, episode 4 of “Sports Stuff”). In the illustrated example, the ratings handler 405 calculates the difference between the “Ratings LiveSameDay” rating and the “Ratings Live” rating and records the logarithm transformation of the calculated difference as the “Same Day Lift” of the media asset of interest in the example translated data database 145.

[0065] In the illustrated example, the second example row 555 indicates the ratings handler 405 determines a “Program Lead-In” value associated with a media asset of interest based on the “Lead-In Ratings” variable retrieved from the example raw data database 135. For example, the ratings handler 405 may query the raw data database 135 for the “Lead-In Ratings” rating related to a media asset of interest (e.g., season 2, episode 4 of “Sports Stuff”). In the illustrated example, the ratings handler 405 records the logarithm transformation of the returned rating as the “Program Lead-In” of the media asset of interest in the example translated data database 145.

[0066] In the illustrated example, the third example row 560 indicates the ratings handler 405 determines an “L7 Lift” value based on the “Ratings LivePlus7” variable and the “Ratings Live” variable retrieved from the example raw data database 135. For example, the ratings handler 405 may query the raw data database 135 for the “Ratings LivePlus7” rating and the “Ratings Live” rating related to a media asset of interest (e.g., season 2, episode 4 of “Sports Stuff”). In the illustrated example, the ratings handler 405 calculates the difference between the “Ratings LivePlus7” rating and the “Ratings Live” rating and records the logarithm transformation of the calculated difference as the “L7 Lift” of the media asset of interest in the example translated data database 145.

[0067] In the illustrated example, the fourth example row 565 indicates the ratings handler 405 determines a “Historical Ratings Lift” value based on the “Ratings LivePlus7” variable and the “Ratings Live” variable retrieved from the example raw data database 135. For example, the ratings handler 405 may query the raw data database 135 for the “Ratings LivePlus7” rating and the “Ratings Live” rating related to a media asset of interest. In the illustrated example, the ratings handler 405 calculates the “historical ratings lift” based on an average of “L7 Lift” values over, for example, two previous weeks (or episodes). For example, to determine the “historical ratings lift” for a media asset of interest (e.g., season 2, episode 4 of “Sports Stuff”), the example ratings handler 405 retrieves the Ratings LivePlus7” rating and the “Ratings Live” rating for season 2, episodes 2 and 3 of “Sports Stuff.” In the illustrated example, the ratings handler 405 calculates an average of the “L7 Lift” value for episode 2 of season 2 and the “L7 Lift” value for episode 3 of season 2. The example ratings handler 405 records the logarithm transformation of the average of the “L7 Lift” values as the “Historical Ratings Lift” of the media asset of interest in the example translated data database 145.

[0068] In the illustrated example, the fifth example row 570 indicates the ratings handler 405 determines a “Program Loyalty” score for a media asset based on an average number of episodes of the media viewed per user over a period (e.g., five weeks). In the illustrated example, the ratings handler 405 queries the raw data database 135 for panelist identifiers associated with panelists who viewed previously aired episodes of the media asset. For example, the ratings handler 405 may determine the “program loyalty” score for episode 7 of season 1 of “Sports Stuff” by (1) identifying panelists who viewed at least one of episodes 2-6 of season 1 of “Sports Stuff,” (2) determining the total number of episodes 2-6 of season 1 that were viewed by each of the panelists and (3) calculating a ratio of the total number of episodes 2-6 of season 1 that were watched by each of the panelists and the total number of identified panelists. In the illustrated example, the ratings handler 405 records the calculated ratio as the “program loyalty” score of the media asset of interest (e.g., episode 7 of season 1 of “Sports Stuff”) in the example translated data database 145.

[0069] In the illustrated example of FIG. 5, the example ratings handler 405 applies a logarithm transformation to the "Same Day Lift" variable (row 550), the "Program Lead-In" variable (row 555), the "L7 Lift" variable (row 560) and the example "Historical Ratings Lift" variable (row 565). However, the ratings handler 405 may additionally or alternatively apply different transformations, including no transformations, to the variables.

[0070] FIG. 6 illustrates an example program loyalty score calculation. In the illustrated example of FIG. 6, a "program loyalty" score 600 for a media asset is calculated based on the number of weeks a panelist viewed media over a five week period. In the illustrated example, three example panelists 605, 610, 615 are identified as panelists who viewed at least one episode of a program over a period of five weeks. For example, the first example panelist 605 viewed four total episodes (e.g., watched during weeks 1-3 and week 5), the second example panelist 610 viewed three total episodes (e.g., watched during weeks 1, 2 and 5), and the third example panelist 615 viewed one total episode (e.g., watched during week 1). In the illustrated example, the program loyalty score 600 (e.g., 2.67) is calculated as a ratio of the total number of episodes 1-5 that were viewed by each of panelists (e.g., 4 + 3 + 1 = 8) and the total number of panelists (e.g., 3).

[0071] In the illustrated example, a five-week period is used to determine a program loyalty score. However, any other period over which program loyalty is assessed may additionally or alternatively be used. Furthermore, any other technique for calculating program loyalty scores may additionally or alternatively be used. In some examples, raw data related to a media asset may not be immediately available. For example, panelist identifiers associated with panelists who viewed an episode may not be available, for example, for two weeks after the media asset is broadcast. In some such examples, the example ratings handler 405 may not determine a program loyalty score for a media asset that is earlier than, for example, episode seven of a season.

[0072] In the illustrated example of FIG. 4, the example data translator 140 includes the example attributes handler 410 to process attributes and/or characteristics representative of media assets. For example, the attributes handler 410 may query and/or retrieve program attributes information from the raw data database 135 (e.g., genre-identifying information, day-of-week information, broadcast time-identifying information, etc.) to translate into a form meaningfully handled by the example model builder 150.

[0073] An example data table 700 of the illustrated example of FIG. 7 illustrates example translated program attributes data variables that may be recorded by the attributes handler 410 in the example translated data database 145. The example data table 700 of the illustrated example of FIG. 7 includes a variable name identifier column 705, a variable data type identifier column 710 and a variable meaning identifier column 715. The example variable name identifier column 705 indicates example variables that may be associated with a media asset and/or useful for predicting time-shifted exposure to the media asset. The example variable data type identifier column 710 indicates a data type of the corresponding variable. The example variable meaning identifier column 715 provides a brief description of the value associated with the corresponding variable. While three example variable identifier columns are represented in the example data table 700 of FIG. 7, more or fewer variable identifier columns may be represented in the example data table 700.

[0074] The example data table 700 of the illustrated example of FIG. 7 includes thirteen example rows corresponding to example translated program attributes data variables. In the illustrated example, the example translated program attributes data variables of the data table 700 represent four example characteristics of a media asset. The first example block of rows 750 indicates that the example attributes handler 410 stores day-of-week information as Boolean variables. In the illustrated example, the attributes handler 410 translates day-of-week information that is stored as a string data type at the raw data database 135 to one or more day-of-week Boolean variables. For example, the attributes handler 410 may retrieve day-of-week information related to a media asset indicating the date of the week that the media asset is broadcast (e.g., “Tuesday”) and set the corresponding day-of-week Boolean variable to true (e.g., “1”) and set (or reset) other day-of-week Boolean variables to false (e.g., “0”). In the illustrated example, in response to determining that the raw day-of-week information indicates the media asset is broadcast on a “Tuesday,” the example attributes handler 410 sets the value of the corresponding “Day Tues” variable to true (e.g., “1”) and sets (or resets) the values of the other day-of-week Boolean variables (e.g., “Day Mon,” “Day Wed,” “Day Thurs,” “Day Fri” and “Day SatSun”) to false (e.g., “0”). Although the example day-of-week information is represented as six example Boolean variables in the example data table 700 of FIG. 7, any other number of Boolean variables may additionally or alternatively be used. For example, the attributes handler 410 may group the days-of-week information into weekday or weekend Boolean variables.

[0075] The second example block of rows 755 of the data table 700 of FIG. 7 indicates that the example attributes handler 410 stores genre-identifying information as Boolean variables. In the illustrated example, the attributes handler 410 translates genre-identifying information that is stored as a string data type at the raw data database 135 to one or more genre-related Boolean variables. For example, the attributes handler 410 may retrieve genre-identifying information indicative of the genre classification of a media asset (e.g., a documentary, drama, variety, comedy, etc.) and set the corresponding genre-related Boolean variable to true (e.g., “1”) and set (or reset) other genre-related Boolean variables to false (e.g., “0”). In the illustrated example, in response to determining that retrieved raw genre-identifying information indicates the corresponding media asset is a “comedy,” the example attributes handler 410 sets the value of the corresponding “Genre Comedy” variable to true (e.g., “1”) and sets (or resets) the values of the other genre-related Boolean variables (e.g., “Genre Documentary,” “Genre Drama” and “Genre Variety”) to false (e.g., “0”). Although the example genre-identifying information is represented as four example genre-related Boolean variables in the example data table 700 of FIG. 7, any other number of Boolean variables representative of the genre of a media asset may additionally or alternatively be used.

[0076] The third example block of rows 760 of the data table 700 of FIG. 7 indicates that the example attributes handler 410 stores originator-identifying information as Boolean variables. In the illustrated example, the attributes handler 410 translates originator-identifying information that is stored as a string data type at the raw data database 135 to one or more originator-related Boolean variables. For example, the attributes handler 410 may retrieve originator-identifying information indicative of the network (or channel) that broadcasts a media asset (e.g., channel “ABC,” channel “XYZ,” etc.) and set the corresponding originator-related Boolean variable to true (e.g., “1”) and set (or reset) other originator-related Boolean variables to false (e.g., “0”). In the illustrated example, in response to determining that retrieved raw originator-identifying information indicates the corresponding media asset is broadcast on channel “ABC,” the example attributes handler 410 sets the value of the corresponding “Originator ABC” variable to true (e.g., “1”) and sets (or resets) the values of the other originator-related Boolean variables (e.g., “Originator XYZ”) to false (e.g., “0”). While two example originators are represented in the example data table 700 of FIG. 7, more or fewer originators may be represented in the example data table 700 corresponding to the many broadcast networks and cable networks that broadcast media assets.

[0077] The example row 765 of the data table 700 of FIG. 7 indicates that the example attributes handler 410 stores broadcast time-identifying information as an integer data type. In the illustrated example, the attributes handler 410 maps broadcast time-identifying information that is stored as a string data type at the raw data database 135 to an integer. For example, the attributes handler 410 may retrieve broadcast time-identifying information indicative of when a media asset is broadcast (e.g., "00:00 – 00:30," "00:30 – 01:00," … "23:30 – 00:00") and set the "HH Block" variable value based on a corresponding half-hour block. For example, the attributes handler 410 may map the broadcast time "00:00 – 00:00" to half-hour block "0," may map the broadcast time "00:30 – 01:00" to half-hour block "1," etc. Although the example broadcast time-identifying information is represented as half-hour blocks, any other granularity may additionally or alternatively be used. For example, the broadcast times may be based on quarter-hours, full hours, etc.

[0078] While the example data table 700 of FIG. 7 includes four example program attributes related to a media asset (e.g., day-of-week, genre, originator and broadcast time), any other number of program attributes may additionally or alternatively be used.

[0079] In the illustrated example of FIG. 4, the example data translator 140 includes the example social media handler 415 to process social media messages representative of media assets. For example, the social media handler 415 may query and/or retrieve social media messages and/or social media messages-related information from the raw data database 135 (e.g., message identifiers, message timestamps, message content, message authors, etc.) to translate into a form meaningfully handled by the example model builder 150.

[0080] An example data table 800 of the illustrated example of FIG. 8 illustrates example translated social media data variables that may be recorded by the social media handler 415 in the example translated data database 145. The example data table 800 of the illustrated example of FIG. 8 includes a variable name identifier column 805, a variable data type identifier column 810 and a variable meaning identifier column 815. The example variable name identifier column 805 indicates example variables that may be associated with a media asset broadcast and/or information useful for predicting time-shifted exposure to the media asset. The example variable data type identifier column 810 indicates a data type of the corresponding variable. The example variable meaning identifier column 815 provides a brief description of the value associated with the corresponding variable. While three example variable identifier columns are represented in the example data table 800 of FIG. 8, more or fewer variable identifier columns may be represented in the example data table 800.

[0081] The example data table 800 of the illustrated example of FIG. 8 includes three example rows corresponding to example translated social media data variables. The first example row 850 indicates the social media handler 415 of FIG. 4 stores a "SM Count" variable as a floating data type in the example translated data database 145. In the illustrated example, the social media handler 415 determines a "SM Count" value associated with a media asset of interest based on a number of posted social media messages of interest. For example, the social media handler 415 may inspect the social media messages returned by the raw data database 135 for social media messages that indicate exposure to a media asset. For example, a media asset may be “Sports Stuff.” In such instances, a social media message of interest may include the text “Jon is my favorite character on Sports Stuff!” and may include a message timestamp indicating that the social media message was posted by the message author during broadcast of the media asset. In the illustrated example, the social media handler 415 may count the number of social media messages identified as of interest and record a logarithm transformation of the number of social media messages of interest (e.g., the social media messages that indicate exposure to a media asset) as the “SM Count” corresponding to the media asset of interest in the example translated data database 145.

[0082] The second example row 855 of the data table 800 of FIG. 8 indicates that the example social media handler 415 stores a value related to the number of unique authors who posted social media messages of interest as a floating data type in the example translated data database 145. The second example row 855 of the data table 800 of FIG. 8 indicates the social media handler 415 of FIG. 4 stores a “SM UAuthors” variable as a floating data type in the example translated data database 145. In the illustrated example, the social media handler 415 determines a “SM UAuthors” value associated with a media asset of interest based on a number of unique authors who posted social media messages of interest. For example, the social media handler 415 may inspect the social media messages returned by the raw data database 135 for social media messages that indicate exposure to a media asset. In the illustrated example, the social media handler 415 may count the number of unique authors who posted the social media messages identified as of interest and record a logarithm transformation of the number of unique authors as the “SM UAuthors” corresponding to the media asset of interest in the example translated data database 145.

[0083] The third example row 860 of the data table 800 of FIG. 8 indicates the social media handler 415 of FIG. 4 stores a “SM Impressions” variable as a floating data type in the example translated data database 145. In the illustrated example, the social media handler 415 determines a “SM Impressions” value associated with a media asset of interest based on impression information associated with social media messages of interest. For example, the social media handler 415 may count the number of times social media messages identified as of interest were viewed by users of the social media service and record a logarithm transformation of the number of impressions of social media messages of interest as the “SM Impressions” corresponding to the media asset of interest in the example translated data database 145.

[0084] In the illustrated example of FIG. 8, the example social media handler 415 applies a logarithm transformation to the "SM Count" variable (row 850), the "SM UAuthors" variable (row 855) and the example "SM Impressions" variable (row 860). However, the social media handler 415 may additionally or alternatively apply different transformations, including no transformations, to the variables.

[0085] In the illustrated example, the example social media handler 415 inspects social media messages and/or social media messages-related information retrieved from the raw data database 135 to translate into a form meaningfully handled by the example model builder 150. In some examples, the raw audience measurement data 110 may be provided as aggregated data. For example, rather than providing social media messages and/or social media messages-related information, the example audience measurement system(s) 105 of FIG. 1 may count the number of posted social media messages related to media assets of interest, may count the number of unique authors who posted social media messages related to media assets of interest and may count the number of impressions associated with posted social media messages related to media assets of interest, and provide the respective counts to the example central facility 125. In some such instances, the example social media handler 415 may retrieve the respective counts and store the logarithm transformation of the corresponding numbers as the respective social media-related variables. However, any other technique may be used to determine the number of posted social media messages related to media assets of interest, the number of unique authors who posted social media messages related to media assets of interest and the number of impressions associated with posted social media messages related to media assets of interest.

[0086] An example schema 900 of the illustrated example of FIG. 9 illustrates example ratings lift models based on season-episode groupings that may be used by the example model builder 150 when building lift prediction models and/or used by the example program ratings estimator 160 when applying data to a lift prediction model to predict a ratings list for a media asset of interest. The example schema 900 of the illustrated example of FIG. 9 includes a season identifier column 905 to identify a season number of a media asset, an episode number identifier column 910 to identify an episode number of a media asset and a lift prediction model identifier column 915 to identify a lift prediction model applicable to the corresponding season-episode grouping.

[0087] The example schema 900 of FIG. 9 indicates that, in the illustrated example, during the first season of a media asset (corresponding section 930 of the schema 900), three different lift prediction models may be used. For example, the first three episodes are associated with a first lift prediction model (Model 1). The example schema 900 also indicates that the next three episodes (e.g., episodes four through six of season one) are associated with a second lift prediction model (Model 2). The example schema 900 also indicates that episode seven and subsequent episodes of season one are associated with a third lift prediction model (Model 3).

[0088] In the illustrated example of FIG. 9, the example schema 900 indicates that during the second and subsequent seasons (corresponding to section 935 of schema 900), two different lift prediction models may be used. For example, the first six episodes of the second season are associated with the second lift prediction model (Model 2) while episodes seven and later are associated with the third lift prediction model (Model 3). While three example lift prediction models are represented in the example schema 900 of FIG. 9, more or fewer lift prediction models may be represented in the example schema 900 based on different season-episode groupings.

[0089] An example data table 1000 of the illustrated example of FIG. 10 illustrates example sets of translated data variables that are used when generating (e.g., training) lift prediction models and/or that are applied to a lift prediction model when estimating ratings lift for media assets of interest. The example data table 1000 of the illustrated example of FIG. 10 includes a variable identifier column 1005, a first lift prediction model identifier column 1010, a second lift prediction model identifier column 1015 and a third lift prediction model identifier column 1020. In the illustrated example, the three example lift prediction models of FIG. 10 correspond to the three example lift prediction models of FIG. 9.

[0090] The example data table 1000 of the illustrated example of FIG. 10 includes example rows corresponding to example translated ratings data variables (e.g., features) that are applied to the respective lift prediction models. In the illustrated example, the first example row 1050 corresponds to different demographic groupings (e.g., households, persons age two and up, persons between the ages of 25 and 54 and persons between the ages of 19 and 49). In some instances, it may be beneficial to develop lift prediction models for different demographic groupings since members of the respective groupings may influence ratings lift predictions differently. For example, social media indicators may not be a predictive variable for media assets whose audience(s) tends to not use social media services. Although the example data table 1000 indicates that twelve lift prediction models are generated and/or applied (e.g., four demographic groupings for each of the three lift prediction models), any other number of lift prediction models may additionally or alternatively be used. For example, the number of demographic groupings for one or more of the example lift prediction models vary based on, for example, client preference.

[0091] The example data table 1000 of the illustrated example of FIG. 10 indicates that current TV ratings 1055 (e.g., “live + same day” ratings and “live” ratings), program lead-in 1060 (e.g., program lead-in ratings), program characteristics 1065 (e.g., day-of-week, genre, originator and half-hour block broadcast time) and social media indicators 1070 are used when training each of the lift prediction models and/or when applying each of lift prediction models. As described above, the current TV ratings 1055 correspond to the example audience measurement data 110 provided from the audience measurement system(s) 105 and used when building and/or training the lift prediction models. When applying the lift prediction models to estimate ratings lift for a media asset of interest, the current TV ratings 1055 correspond to the estimated current TV ratings determined by the example program ratings estimator 160 of FIG. 1.

[0092] In the illustrated example, historical TV ratings data 1075 is not used when generating and/or applying the first example lift predictor model, but is used with respect to the second and third example lift predictor models. Excluding the historical TV ratings data may be beneficial in connection with the first example lift predictor model because, as described above, the historical TV ratings data includes determining an average of previous TV ratings of the media asset. In some such instances, because the media asset is a new program (e.g., one of the first three episodes of the program), previous TV ratings may not be available (e.g., when the media asset of interest is the pilot) and/or may not be a reliable predictor of ratings lift. For example, consumers may switch between different pilots and/or episodes as they evaluate whether the program is worth their time, resulting in skewed TV ratings for the first and/or second episodes.

[0093] In the illustrated example, program loyalty scores 1080 are excluded from the first and second example lift prediction models, but are included with respect to the third example lift prediction model. In the illustrated example, excluding program loyalty scores may be beneficial in connection with the first and second example lift prediction models because, as described above, the program loyalty scores are determined based on five previous episodes and, in some examples, may also include a two-week lag. In some such instances, because seven episodes-worth of audience measurement data is needed to develop the program loyalty scores, including such scores in the first and/or second example lift prediction models would not be useful in estimating ratings lift for the media asset.

[0094] While an example manner of implementing the central facility 125 of FIG. 1 is illustrated in FIG. 1, one or more of the elements, processes and/or devices illustrated in FIG. 1 may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way. Further, the example data interface 130, the example raw data database 135, the example data translator 140, the example translated data database 145, the example model builder 150, the example models database 155, the example program ratings estimator 160 and/or, more generally, the example central facility 125 of FIG. 1 may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware. Thus, for example, any of the example data interface 130, the example raw data database 135, the example data translator 140, the example translated data database 145, the example model builder 150, the example models database 155, the example program ratings estimator 160 and/or, more generally, the example central facility 125 of FIG. 1 could be implemented by one or more analog or digital circuit(s), logic circuits, programmable processor(s), application specific integrated circuit(s) (ASIC(s)), programmable logic device(s) (PLD(s)) and/or field programmable logic device(s) (FPLD(s)). When reading any of the apparatus or system claims of this patent to cover a purely software and/or firmware implementation, at least one of the example data interface 130, the example raw data database 135, the example data translator 140, the example translated data database 145, the example model builder 150, the example models database 155, the example program ratings estimator 160 and/or, more generally, the example central facility 125 of FIG. 1 is/are hereby expressly defined to include a tangible computer readable storage device or storage disk such as a memory, a digital versatile disk (DVD), a compact disk (CD), a Blu-ray disk, etc. storing the software and/or firmware. Further still, the example central facility 125 of FIG. 1 may include one or more elements, processes and/or devices in addition to, or instead of, those illustrated in FIG. 1, and/or may include more than one of any or all of the illustrated elements, processes and devices.

[0095] While an example manner of implementing the data translator 140 of FIG. 1 is illustrated in FIG. 4, one or more of the elements, processes and/or devices illustrated in FIG. 4 may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way. Further, the example ratings handler 405, the example attributes handler 410, the example social media handler 415 and/or, more generally, the example data translator 140 of FIG. 4 may be implemented by hardware, software, firmware and/or any combination of hardware, software and/or firmware. Thus, for example, any of the example ratings handler 405, the example attributes handler 410, the example social media handler 415 and/or, more generally, the example data translator 140 of FIG. 4 could be implemented by one or more analog or digital circuit(s), logic circuits, programmable processor(s), application specific integrated circuit(s) (ASIC(s)), programmable logic device(s) (PLD(s)) and/or field programmable logic device(s) (FPLD(s)). When reading any of the apparatus or system claims of this patent to cover a purely software and/or firmware implementation, at least one of the example ratings handler 405, the example attributes handler 410, the example social media handler 415 and/or, more generally, the example data translator 140 of FIG. 4 is/are hereby expressly defined to include a tangible computer readable storage device or storage disk such as a memory, a digital versatile disk (DVD), a compact disk (CD), a Blu-ray disk, etc. storing the software and/or firmware. Further still, the example data translator 140 of FIG. 1 may include one or more elements, processes and/or devices in addition to, or instead of, those illustrated in FIG. 4, and/or may include more than one of any or all of the illustrated elements, processes and devices.

[0096] Flowcharts representative of example machine readable instructions for implementing the example central facility of FIG. 1 are shown in FIGS. 11-13 and/or 14. In this example, the machine readable instructions comprise a program for execution by a processor such as the processor 1512 shown in the example processor platform 1500 discussed below in connection with FIG. 15. The program may be embodied in software stored on a tangible computer readable storage medium such as a CD-ROM, a floppy disk, a hard drive, a digital versatile disk (DVD), a Blu-ray disk, or a memory associated with the processor 1512, but the entire program and/or parts thereof could alternatively be executed by a device other than the processor 1512 and/or embodied in firmware or dedicated hardware. Further, although the example program is described with reference to the flowcharts illustrated in FIGS. 11-13 and/or 14, many other methods of implementing the example central facility 125 may alternatively be used. For example, the order of execution of the blocks may be changed, and/or some of the blocks described may be changed, eliminated, or combined.

[0097] As mentioned above, the example processes of FIGS. 11-13 and/or 14 may be implemented using coded instructions (e.g., computer and/or machine readable instructions) stored on a tangible computer readable storage medium such as a hard disk drive, a flash memory, a read-only memory (ROM), a compact disk (CD), a digital versatile disk (DVD), a cache, a random-access memory (RAM) and/or any other storage device or storage disk in which information is stored for any duration (e.g., for extended time periods, permanently, for brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term tangible computer readable storage medium is expressly defined to include any type of computer readable storage device and/or storage disk and to exclude propagating signals and to exclude transmission media. As used herein, "tangible computer readable storage medium" and "tangible machine readable storage medium" are used interchangeably. Additionally or alternatively, the example processes of FIGS. 11-13 and/or 14 may be implemented using coded instructions (e.g., computer and/or machine readable instructions) stored on a non-transitory computer and/or machine readable medium such as a hard disk drive, a flash memory, a read-only memory, a compact disk, a digital versatile disk, a cache, a random-access memory and/or any other storage device or storage disk in which information is stored for any duration (e.g., for extended time periods, permanently, for brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term non-transitory computer readable medium is expressly defined to include any type of computer readable storage device and/or storage disk and to exclude propagating signals and to exclude transmission media. As used herein, when the phrase "at least" is used as the transition term in a preamble of a claim, it is open-ended in the same manner as the term "comprising" is open ended. “Comprising” and all other variants of “comprise” are expressly defined to be open-ended terms. “Including” and all other variants of “include” are also defined to be open-ended terms. In contrast, the term “consisting” and/or other forms of “consist” are defined to be close-ended terms.

[0098] FIG. 11 is a flowchart representative of example machine-readable instructions 1100 that may be executed by the example central facility 125 of FIG. 1 to predict time-shifted exposure to media. The example instructions 1100 of FIG. 11 begin at block 1102 when the example central facility 125 receives audience measurement data 110 from the example audience measurement system(s) of FIG. 1. For example, the example data interface 130 (FIG. 1) may periodically obtain and/or retrieve example panelist media measurement data 110A and/or example social media activity data 110B. In some examples, the data interface 130 may obtain and/or retrieve the example audience measurement data 110 aperiodically and/or as a one-time event. The example data interface 130 stores the audience measurement data 110 in the example raw data database 135 (FIG. 1).

[0099] At block 1104, the example central facility 125 normalizes the audience measurement data 110. For example, the example data translator 140 (FIG. 1) may manipulate and/or modify features of the audience measurement data 110 from their raw form to be more meaningfully handled when building the prediction models and estimating the ratings lift. The example data translator 140 stores the translated data variables in the example translated data database 145 (FIG. 1). An example approach to normalize audience measurement data 110 is described below in connection with FIG. 12.

[00100] At block 1106, the example central facility 125 builds a lift prediction model. For example, the example model builder 150 (FIG. 1) determines a relationship between translated data variables stored in the example translated data database 145 and ratings lifts. The example model builder 150 stores the generated model(s) in the example models data store 155 (FIG. 1). An example approach to build a lift prediction model is described below in connection with FIG. 13.

[00101] At block 1108, the example central facility 125 determines an estimated ratings lift for a media asset of interest. For example, the example program ratings estimator 160 (FIG. 1) may apply data related to the media asset of interest to a lift prediction model to estimate a ratings lift for the media asset. An example approach to determine an estimated ratings lift is described below in connection with FIG. 14. The example process 1100 of FIG. 11 ends.

[00102] While in the illustrated example, the example instructions 1100 of FIG. 11 represent a single iteration of predicting time-shifted exposure to media, in practice, the example instructions 1100 of the illustrated example of FIG. 11 may be executed in parallel (e.g., in separate threads) to allow the central facility 125 to handle multiple requests for ratings lift estimations at a time.

[00103] FIG. 12 is a flowchart representative of example machine-readable instructions 1200 that may be executed by the example data translator 140 of FIGS. 1 and/or 4 to normalize raw audience measurement data. The example process 1200 of the illustrated example of FIG. 12 begins at block 1202 when the example data translator 140 obtains ratings-related information associated with media assets. For example, the data translator 140 may retrieve and/or query “live + same day” ratings, “live” ratings, “live + 7” ratings and/or information representative of whether a panelist viewed a particular episode of a media asset from the example raw data database 135. At block 1204, the example data translator 140 translates the ratings-related information to ratings data variables for use by the example model builder 150 and/or the example program ratings estimator 160. In the illustrated example, the data translator translates the ratings-related information in accordance with the example translated ratings data variables table 500 of FIG. 5. At block 1206, the example data translator 140 determines whether there is additional ratings-related information to translate. If, at block 1206, the data translator 140 determined that there is additional ratings-related information to translate, control returns to block 1202.

[00104] If, at block 1206, the data translator 140 determined that there is not additional ratings-related information to translate, then, at block 1208, the example data translator 140 obtains program attributes information associated with media assets. For example, the data translator 140 may retrieve and/or query the example raw data database 135 for day-of-week information, genre information, originator information and/or broadcast time information. At block 1210, the example data translator 140 translates the program attributes information to program attributes data variables for use by the example model builder 150 and/or the example program ratings estimator 160. In the illustrated example, the data translator translates the program attributes information in accordance with the example translated program attributes data variables table 700 of FIG. 7. At block 1212, the example data translator 140 determines whether there is additional program attributes information to translate. If, at block 1212, the data translator 140 determined that there is additional program attributes information to translate, control returns to block 1208.

[00105] If, at block 1212, the data translator 140 determined that there is not additional program attributes information to translate, then, at block 1214, the example data translator 140 obtains social media messages-related information associated with media assets. For example, the data translator 140 may retrieve and/or query the example raw data database 135 for a number of posted social media messages of interest, a number of unique authors who posted social media messages of interest and/or a number of impressions of social media messages of interest. At block 1216, the example data translator 140 translates the social media messages-related information to social media data variables for use by the example model builder 150 and/or the example program ratings estimator 160. In the illustrated example, the data translator translates the social media messages-related information in accordance with the example translated social media data variables table 800 of FIG. 8. At block 1218, the example data translator 140 determines whether there is additional social media messages-related information to translate. If, at block 1218, the data translator 140 determined that there is additional social media messages-related information to translate, control returns to block 1214.

[00106] If, at block 1218, the data translator 140 determined that there is not additional social media messages-related information to translate, then, at block 1220, the example data translator 140 determines whether to continue normalizing audience measurement data. If, at block 1220, the example data translator 140 determined to continue normalizing audience measurement data, control returns to block 1202 to wait to obtain ratings-related information for translating.

[00107] If, at block 1220, the example data translator 140 determined not to continue normalizing audience measurement data, the example process 1200 of FIG. 12 ends.

[00108] While in the illustrated example, the example instructions 1200 of FIG. 12 represent a single iteration of normalizing audience measurement data, in practice, the example instructions 1200 of the illustrated example of FIG. 12 may be executed in parallel (e.g., in separate threads) to allow the central facility 125 to handle multiple requests for normalizing audience measurement data at a time.

[00109] FIG. 13 is a flowchart representative of example machine-readable instructions 1300 that may be executed by the example model builder 150 of FIG. 1 to build a lift prediction model used to estimate ratings lift for media assets of interest. The example instructions 1300 of FIG. 13 begin at block 1302 when the example model builder 150 selects a lift prediction model to build based on the translated data variables stored in the example translated data database 145 of FIG. 1. For example, the model builder 150 may select a lift prediction model based on season-episode groupings of a particular media asset of interest for which ratings list is to be predicted. At block 1304, the example model builder 150 selects a demographic grouping associated with the lift prediction model. For example, the model builder 150 may generate a plurality of lift prediction models corresponding to different demographic segments.

[00110] At block 1306, the example model builder 150 obtains translated data variables based on the selected lift prediction model and the selected demographic grouping. For example, depending on the selected lift prediction model, the model builder 150 may obtain program characteristics, social media indicators, current ratings information, historical ratings information and/or program loyalty scores for different media assets. In some examples, the model builder 150 may consult the example data table 1000 of the illustrated example of FIG. 10 to determine the translated data variables associated with the selected lift prediction model.

[00111] At block 1308, the example model builder 150 determines a relationship between the obtained translated data variables and rating lifts. For example, the model builder 150 may represent the lift prediction model via an equation (e.g., the example Equation 1 described above) and by applying logistic regression to the translated data variables, the model builder 150 may solve for coefficients of the equation.

[00112] At block 1310, the example model builder 150 records the generated lift prediction model. For example, the model builder 150 may store the determined coefficients in the example models data store 155 of FIG. 1. At block 1312, the example model builder

150 determines whether there is another demographic grouping to process for the selected lift prediction model. If, at 1312, the model builder 150 determined that there is another demographic grouping to process, then control returns to block 1304 to select a demographic grouping to process.

[00113] If, at block 1312, the model builder 150 determined that there is not another demographic grouping to process, then, at block 1314, the example model builder 150 determines whether there is another lift prediction model to build. For example, the model builder 150 may determine whether to build lift prediction models for a different season-episode grouping of a different media asset for which ratings lift is to be predicted. If, at block 1314, the model builder 150 determined that there is another lift prediction model to build, then control returns to block 1302 to select a lift prediction model.

[00114] If, at block 1314, the model builder 150 determined that there is not another lift prediction model to build, the example process 1300 of FIG. 13 ends.

[00115] FIG. 14 is a flowchart representative of example machine-readable instructions 1400 that may be executed by the example program ratings estimator 160 of FIG. 1 to predict time-shifted exposure to media. The example instructions 1400 of FIG. 14 begin at block 1402 when the example program ratings estimator 160 selects a lift prediction model to use to estimate ratings lift for a media asset of interest. For example, the example program ratings estimator 160 may select a lift prediction model based on season and episode number information associated with the media asset of interest. In some examples, the program ratings estimator 160 may consult the example schema 900 of the illustrated example of FIG. 9 to identify a lift prediction model based on season-episode groupings.

[00116] At block 1404, the example program ratings estimator 160 obtains translated data variables based on the selected lift prediction model. For example, depending on the selected lift prediction model, the program ratings estimator 160 may obtain, for the media asset of interest, program characteristics, social media indicators, historical ratings information and/or program loyalty scores. In some examples, the program ratings estimator 160 may consult the example data table 1000 of the illustrated example of FIG. 10 to determine the translated data variables associated with the selected lift prediction model.

[00117] At block 1406, the example program ratings estimator 160 determines estimated current ratings information for the media asset of interest. For example, the program ratings estimator 160 may process time-period ratings (e.g., overnight ratings provided via a fast affiliates report) and actual broadcast information to estimate current ratings (e.g., "live" ratings and "live + same day" ratings) associated with the media asset of interest to determine estimated current ratings for the media asset. In some examples, the program ratings estimator 160 may further process such information to estimate program lead-in ratings associated with the program that precedes the media asset of interest.

[00118] At block 1408, the example program ratings estimator 160 applies the translated data variables and the estimated current ratings to the selected lift prediction model to determine the estimated ratings lift for the media asset of interest. At block 1410, the example program ratings estimator 160 determines whether there is another media asset of interest to process. If, at block 1410, the example program ratings estimator 160 determined that there is another media asset of interest to process, then control returns to block 1402 to select a lift prediction model.

[00119] If, at block 1410, the example program ratings estimator 160 determined that there is not another media asset of interest to process, the example process 1400 of FIG. 14 ends.

[00120] FIG. 15 is a block diagram of an example processor platform 1500 capable of executing the instructions of FIGS. 11-13 and/or 14 to implement the central facility 125 of FIG. 1 and/or the data translator 140 of FIGS. 1 and/or 4. The processor platform 1500 can be, for example, a server, a personal computer, or any other type of computing device.

[00121] The processor platform 1500 of the illustrated example includes a processor 1512. The processor 1512 of the illustrated example is hardware. For example, the processor 1512 can be implemented by one or more integrated circuits, logic circuits, microprocessors or controllers from any desired family or manufacturer.

[00122] The processor 1512 of the illustrated example includes a local memory 1513 (e.g., a cache). The processor 1512 of the illustrated example executes the instructions to implement the example data interface 130, the example raw data database 135, the example data translator 140, the example translated data database 145, the example model builder 150, the example models database 155, the example program ratings estimator 160, the example ratings handler 405, the example attributes handler 410 and the example social media handler 415. The processor 1512 of the illustrated example is in communication with a main memory including a volatile memory 1514 and a non-volatile memory 1516 via a bus 1518. The volatile memory 1514 may be implemented by Synchronous Dynamic Random Access Memory (SDRAM), Dynamic Random Access Memory (DRAM), RAMBUS Dynamic Random Access Memory (RDRAM) and/or any other type of random access memory device. The non-volatile memory 1516 may be implemented by flash memory and/or any other desired type of memory device. Access to the main memory 1514, 1516 is controlled by a memory controller.

[00123] The processor platform 1500 of the illustrated example also includes an interface circuit 1520. The interface circuit 1520 may be implemented by any type of interface standard, such as an Ethernet interface, a universal serial bus (USB), and/or a PCI express interface.

[00124] In the illustrated example, one or more input devices 1522 are connected to the interface circuit 1520. The input device(s) 1522 permit(s) a user to enter data and commands into the processor 1512. The input device(s) can be implemented by, for example, an audio sensor, a microphone, a camera (still or video), a keyboard, a button, a mouse, a touchscreen, a track-pad, a trackball, isopoint and/or a voice recognition system.

[00125] One or more output devices 1524 are also connected to the interface circuit 1520 of the illustrated example. The output devices 1524 can be implemented, for example, by display devices (e.g., a light emitting diode (LED), an organic light emitting diode (OLED), a liquid crystal display, a cathode ray tube display (CRT), a touchscreen, a tactile output device, a printer and/or speakers). The interface circuit 1520 of the illustrated example, thus, typically includes a graphics driver card, a graphics driver chip or a graphics driver processor.

[00126] The interface circuit 1520 of the illustrated example also includes a communication device such as a transmitter, a receiver, a transceiver, a modem and/or network interface card to facilitate exchange of data with external machines (e.g., computing devices of any kind) via a network 1526 (e.g., an Ethernet connection, a digital subscriber line (DSL), a telephone line, coaxial cable, a cellular telephone system, etc.).

[00127] The processor platform 1500 of the illustrated example also includes one or more mass storage devices 1528 for storing software and/or data. Examples of such mass storage devices 1528 include floppy disk drives, hard drive disks, compact disk drives, Blu-ray disk drives, RAID systems, and digital versatile disk (DVD) drives. The example mass storage 1528 implements the example raw data database 135, the example translated data database 145 and the example models data store 155.

[00128] The coded instructions 1532 of FIGS. 11-13 and/or 14 may be stored in the mass storage device 1528, in the volatile memory 1514, in the non-volatile memory 1516, and/or on a removable tangible computer readable storage medium such as a CD or DVD.

[00129] From the foregoing, it will appreciate that the above disclosed methods, apparatus and articles of manufacture facilitate predicting time-shifted viewing of media. For example, disclosed examples include building a lift prediction model based on historical audience measurement data. Examples disclosed herein may then apply data related to the media to the lift prediction model to predict time-shifted viewing of the media. In some examples,

real-time ratings information (e.g., overnight ratings) is also applied to the lift prediction model to predict the time-shifted viewing of the media.

[00130] Although certain example methods, apparatus and articles of manufacture have been disclosed herein, the scope of coverage of this patent is not limited thereto. On the contrary, this patent covers all methods, apparatus and articles of manufacture fairly falling within the scope of the claims of this patent.

What Is Claimed Is:

1. A method to estimate ratings lift for a media asset, the method comprising:

normalizing, with a processor, audience measurement data corresponding to media exposure data and social media activity data;

building an estimation model based on a relationship between a first subset of the normalized audience measurement data associated with a characteristic of the media asset and historical rating lift measurements associated with the media asset;

estimating, with the processor, current ratings for the media asset based on time-period based ratings and broadcast time-periods; and

applying data related to the media asset and the estimated current ratings to the estimation model to estimate, with the processor, the ratings lift for the media asset.

2. The method as defined in claim 1, wherein the normalizing of the audience measurement data includes:

transforming ratings-related information included in the audience measurement data to a first common scale;

transforming program attributes-related information included in the audience measurement data to a second common scale; and

transforming social media-related information included in the audience measurement data to a third common scale.

3. The method as defined in claim 1, wherein the normalizing of the audience measurement data includes transforming the audience measurement data from a first data type to a second data type.

4. The method as defined in claim 1, wherein the characteristic of the media asset corresponds to a season-episode grouping.

5. The method as defined in claim 1, wherein the estimation model corresponds to an equation including coefficients to be applied to the data related to the media asset and the estimated current ratings, and the building of the estimation model includes determining values of the coefficients based on the first subset of the normalized audience measurement data.

7. The method as defined in claim 1, wherein the determining of the estimated current ratings for the media asset includes:

determining a telecast time of the media asset based on the broadcast time-periods; and

mapping the time-period based ratings to the telecast time.

8. An apparatus to estimate ratings lift for a media asset, the apparatus comprising:

a data translator to normalize audience measurement data corresponding to media exposure data and social media activity data;

a model builder to build an estimation model based on a relationship between a first subset of the normalized audience measurement data associated with a characteristic of the media asset and historical rating lift measurements associated with the media asset; and

a ratings estimator to:

estimate current ratings for the media asset based on time-period based ratings and broadcast time-periods; and

apply data related to the media asset and the estimated current ratings to the estimation model to estimate the rating lift for the media asset.

9. The apparatus as defined in claim 8, wherein the data translator includes:

a ratings handler to normalize ratings-related information included in the audience measurement data to a first common scale;

an attributes handler to normalize program attributes-related information included in the audience measurement data to a second common scale; and

a social media handler to normalize social media-related information included in the audience measurement data to a third common scale.

10. The apparatus as defined in claim 8, wherein the data translator transforms the audience measurement data from a first data type to a second data type.

11. The apparatus as defined in claim 8, wherein the characteristic of the media asset corresponds to a season-episode grouping.

12. The apparatus as defined in claim 8, wherein the estimation model is represented by an equation including coefficients to be applied to the data related to the media asset and the estimated current ratings.

13. The apparatus as defined in claim 12, wherein the model builder is to build the estimation model by determining values of the coefficients based on the first subset of the normalized audience measurement data.

14. The apparatus as defined in claim 8, wherein the ratings estimator is to:

determine a telecast time of the media asset based on the broadcast time-periods; and

map the time-period based ratings to the telecast time to determine the estimated current ratings for the media asset.

15. A tangible machine-readable storage medium comprising instructions that, when executed, cause a processor to at least:

normalize audience measurement data corresponding to media exposure data and social media activity data;

build an estimation model based on a relationship between a first subset of the normalized audience measurement data associated with a characteristic of the media asset and historical rating lift measurements associated with the media asset;

estimate current ratings for the media asset based on time-period based ratings and broadcast time-periods; and

apply data related to the media asset and the estimated current ratings to the estimation model to estimate ratings lift for the media asset.

16. The tangible machine-readable storage medium as defined in claim 15, wherein the instructions further cause the processor to normalize the audience measurement data by:

transforming ratings-related information included in the audience measurement data to a first common scale;

transforming program attributes-related information included in the audience measurement data to a second common scale; and

transforming social media-related information included in the audience measurement data to a third common scale.

17. The tangible machine-readable storage medium as defined in claim 15, wherein the instructions further cause the processor to normalize the audience measurement data by transforming the audience measurement data from a first data type to a second data type.

18. The tangible machine-readable storage medium as defined in claim 15, wherein the characteristic of the media asset corresponds to a season-episode grouping.

19. The tangible machine-readable storage medium as defined in claim 15, wherein the estimation model corresponds to an equation including coefficients to be applied to the data related to the media asset and the estimated current ratings, and wherein the instructions further cause the processor to build the estimation model by determining values of the coefficients based on the first subset of the normalized audience measurement data.

20. The tangible machine-readable storage medium as defined in claim 15, wherein the instructions further cause the processor to determine the estimated current ratings for the media asset by:

determining a telecast time of the media asset based on the broadcast time-periods; and

mapping the time-period based ratings to the telecast time.

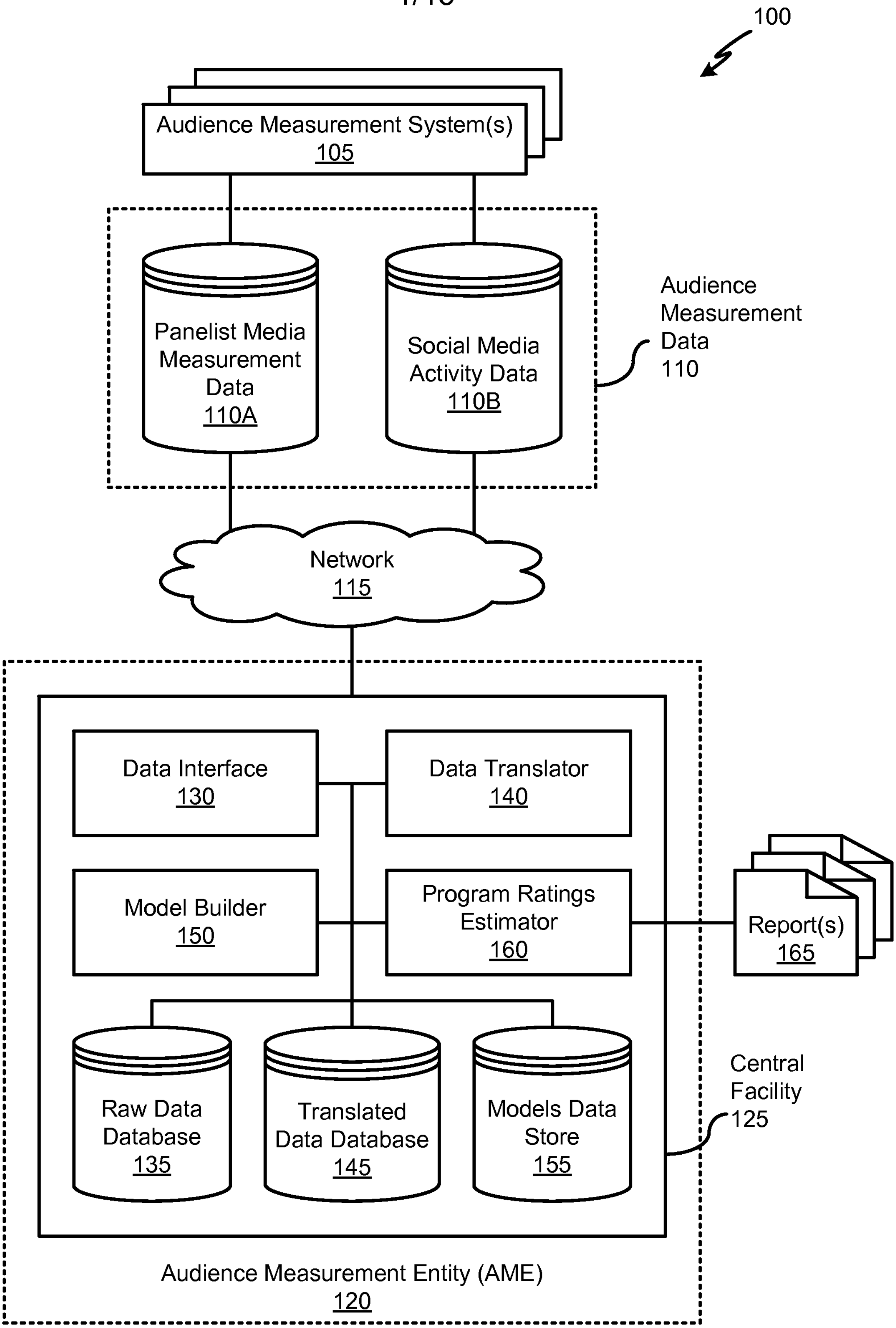

100
Audience Measurement System(s)
105
Panelist Media Measurement Data
110A
Social Media Activity Data
110B
Audience Measurement Data
110
Network
115
Data Interface
130
Data Translator
140
Model Builder
150
Program Ratings Estimator
160
Report(s)
165
Raw Data Database
135
Translated Data Database
145
Models Data Store
155
Central Facility
125
Audience Measurement Entity (AME)
120

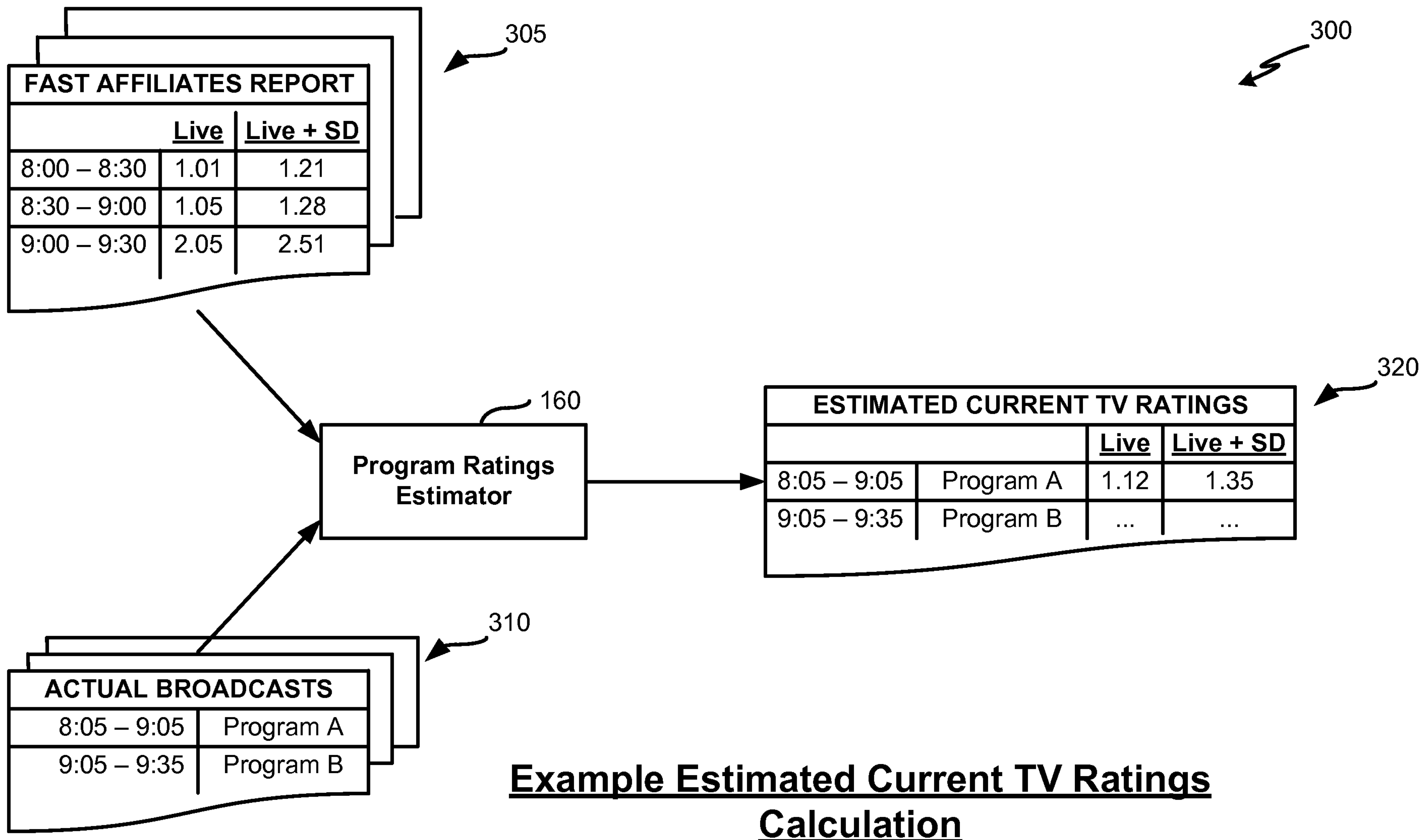
FIG. 1

200

	Name 205	Type 210	Meaning 215
250	Title	String	Name of program
	Episode Identifier	String	Season and episode number of media asset
	Day of Week	String	Day of week of broadcast of media asset
	Broadcast Time	String	Time during media asset program was broadcast (e.g., 20:00-20:30)
	Network	String	Name of network that broadcast the media asset
	Genre	String	Media asset Genre (e.g., documentary, drama, variety, comedy, etc.)
255	Ratings Live	Floating	Program ratings during original telecast
	Ratings LiveSameDay	Floating	Program ratings include viewing during the same broadcast day as the original telecast
	Ratings LivePlus7	Floating	Program ratings include incremental viewing that takes place during the 7 days following the original telecast
260	Lead-In Ratings	Floating	Program ratings for program preceding the media asset
265	Panelist ID	String	Unique identifiers for panelists who viewed the media asset
270	Message ID	String	Unique identifier of social media message
	Message Timestamp	String	Posting time of social media message
	Message Content	String	Content of social media message
	Message Author	String	Unique identifier of posting user
275	Overnight Live	Floating	Program ratings based on scheduled broadcast time (e.g., 20:00-20:30)
	Overnight LiveSameDay	Floating	Program ratings include viewing during the same broadcast day as the scheduled broadcast time
280	Actual Broadcast	String	Actual broadcast times of media asset (e.g., 20:08-20:35)

Raw Audience Measurement Data Variables Table

FIG. 2



Example Estimated Current TV Ratings Calculation

FIG. 3

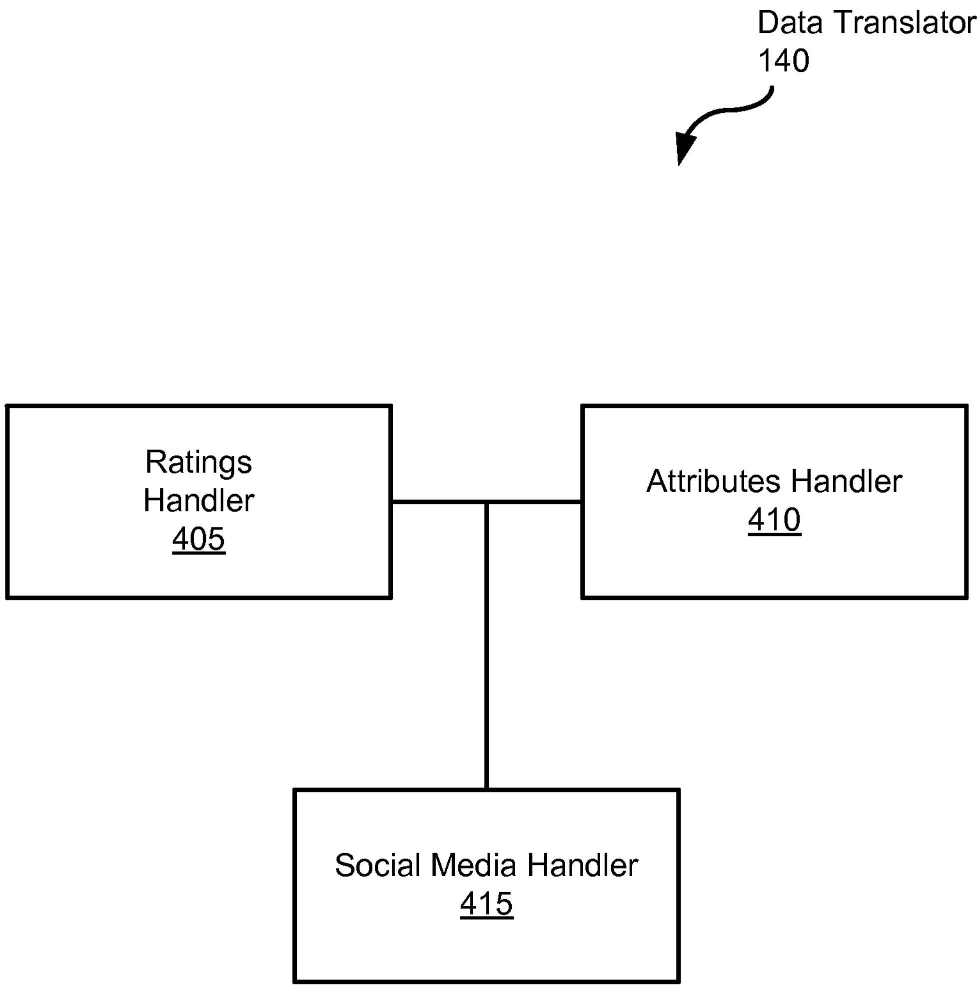

Data Translator
140
Ratings
Handler
405
Attributes Handler
410
Social Media Handler
415

FIG. 4

500

	505 Name	510 Type	515 Meaning
550	Same Day Lift	Floating	Log(LiveSameDay – Live)
555	Program Lead-In	Floating	Log(Lead-In Live Rating)
560	L7 Lift	Floating	Log(LivePlus7 – Live)
565	Historical Ratings Lift	Floating	Log(LivePlus7 – Live); 2 week average
570	Program Loyalty	Floating	Average episodes per viewer; 2 weeks lag

Translated Ratings Data Variables Table

FIG. 5

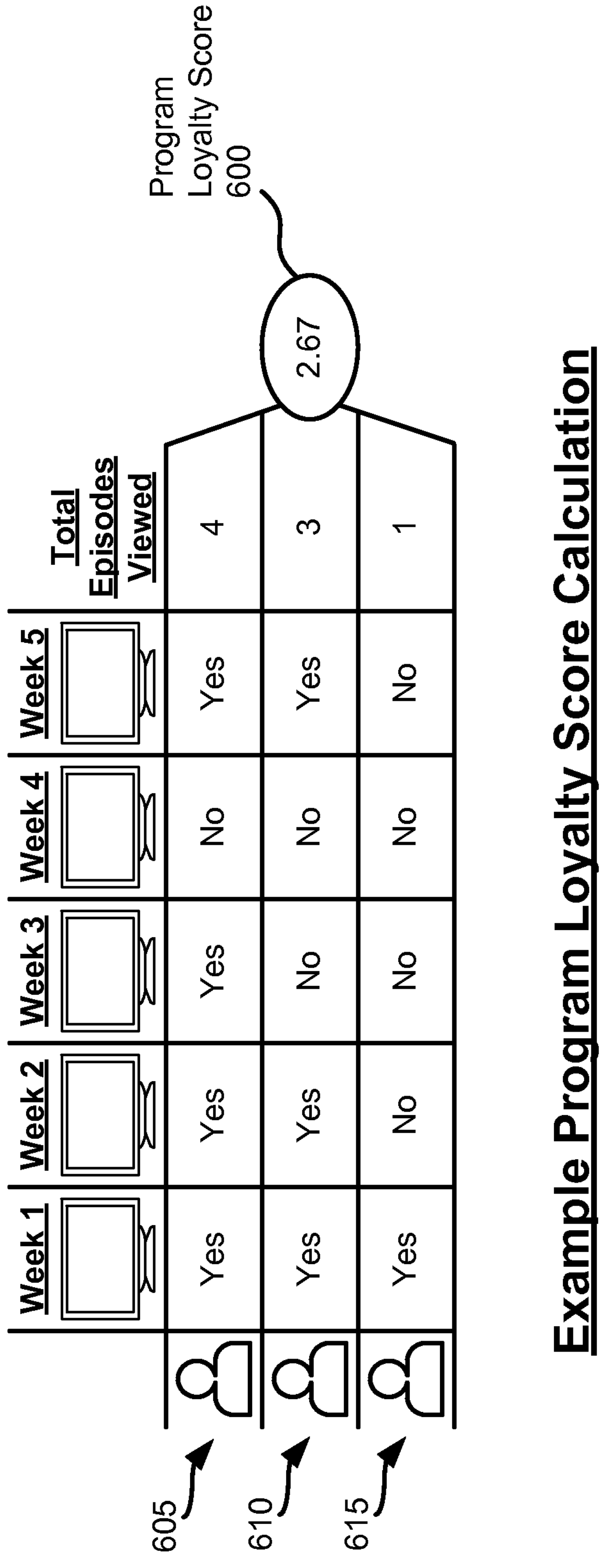

Week 1
Week 2
Week 3
Week 4
Week 5
Total Episodes Viewed
Yes
Yes
Yes
No
Yes
4
Yes
Yes
No
No
Yes
3
Yes
No
No
No
No
1
605
610
615
2.67
Program Loyalty Score 600
Example Program Loyalty Score Calculation

FIG. 6

700

Name	Type	Meaning
Day Mon	Boolean	1 = day of week is Monday
Day Tues	Boolean	1 = day of week is Tuesday
Day Wed	Boolean	1 = day of week is Wednesday
Day Thurs	Boolean	1 = day of week is Thursday
Day Fri	Boolean	1 = day of week is Friday
Day SatSun	Boolean	1 = day of week is Saturday or Sunday
Genre Documentary	Boolean	1 = program genre is documentary
Genre Drama	Boolean	1 = program genre is drama
Genre Variety	Boolean	1 = program genre is variety
Genre Comedy	Boolean	1 = program genre is comedy
Originator ABC	Boolean	1 = telecast is on network ABC
Originator XYZ	Boolean	1 = telecast is on network XYZ
HH Block	Int	0 = 00:00 – 00:30; …; 3 = 01:30 – 02:00; …; 47 = 23:30 – 00:00

705 710 715 750 755 760 765

Translated Program Attributes Data Variables Table

FIG. 7

800

805	810	815

	Name	Type	Meaning
850	SM Count	Floating	Log(Number of posted social media messages of interest)
855	SM UAuthors	Floating	Log(Number of unique authors who posted social media messages of interest)
860	SM Impressions	Floating	Log(Number of impressions of social media messages of interest)

Translated Social Media Data Variables Table

FIG. 8

900

905 910 915

930 935

Season	Episode	Model
Season 1	Pilot	Model 1
	Episode 2	
	Episode 3	
	Episode 4	Model 2
	Episode 5	
	Episode 6	
	Episode 7	Model 3
	Episode 8	
	…	
	Finale	
Season 2+	Premiere	Model 2
	Episode 2	
	Episode 3	
	Episode 4	
	Episode 5	
	Episode 6	
	Episode 7	Model 3
	Episode 8	
	…	
	Finale	

Time-Shifted Ratings Prediction Models
By TV Seasons and Episode Groups

FIG. 9

1000

	1005 Variable	1010 Model 1	1015 Model 2	1020 Model 3
1050	DEMOGRAPHICS	HH | P2+ | P25-54 | P19-49	HH | P2+ | P25-54 | P19-49	HH | P2+ | P25-54 | P19-49
1055	CURRENT TV RATINGS			
	Live Same Day	Yes	Yes	Yes
	Live	Yes	Yes	Yes
1060	PROGRAM LEAD-IN	Yes	Yes	Yes
1065	PROGRAM CHARACTERISTICS			
	Day of Week	Yes	Yes	Yes
	Genre	Yes	Yes	Yes
	Originator	Yes	Yes	Yes
	Half-Hour Block	Yes	Yes	Yes
1070	SOCIAL MEDIA			
	Social Media Messages	Yes	Yes	Yes
	Unique Authors	Yes	Yes	Yes
	Impressions Count	Yes	Yes	Yes
1075	HISTORICAL TV RATINGS			
	Live 7	No	Yes	Yes
	Live	No	Yes	Yes
1080	PROGRAM LOYALTY SCORE	No	No	Yes

Time-Shifted Ratings Prediction Model Features

FIG. 10

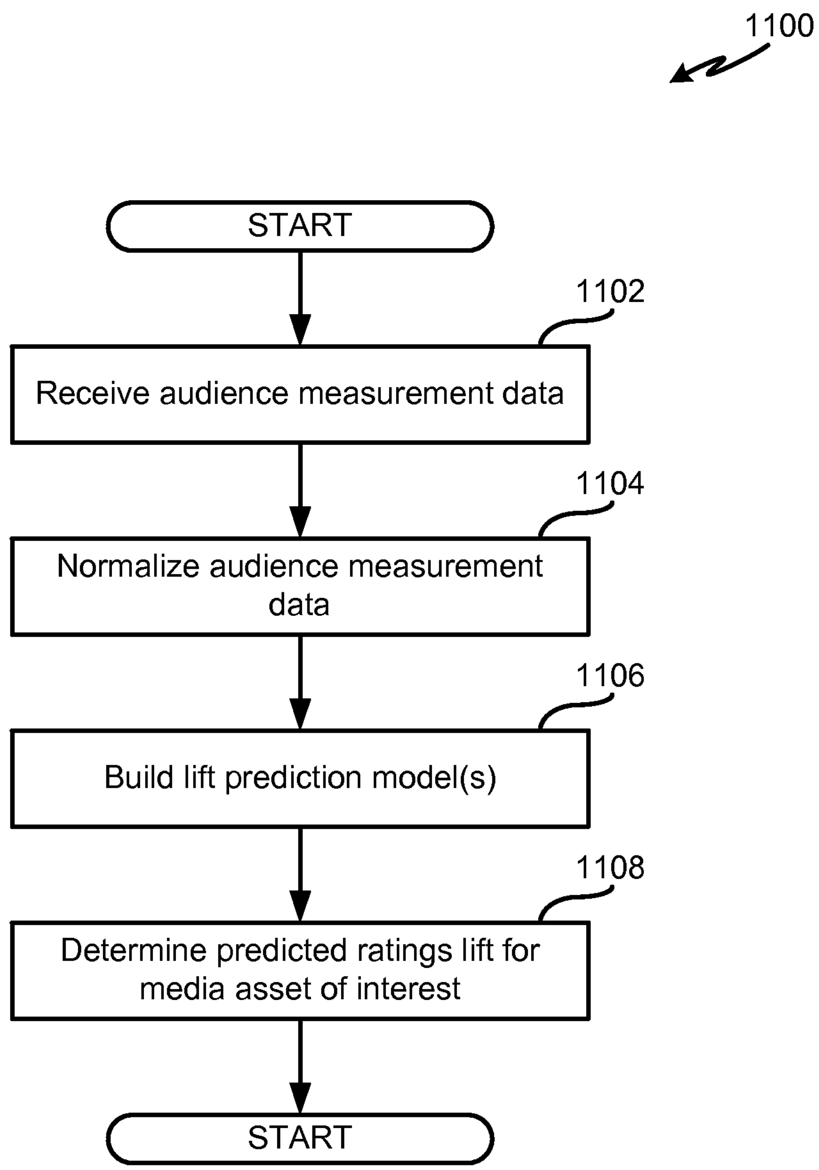

1100
START
1102
Receive audience measurement data
1104
Normalize audience measurement data
1106
Build lift prediction model(s)
1108
Determine predicted ratings lift for media asset of interest
START

FIG. 11

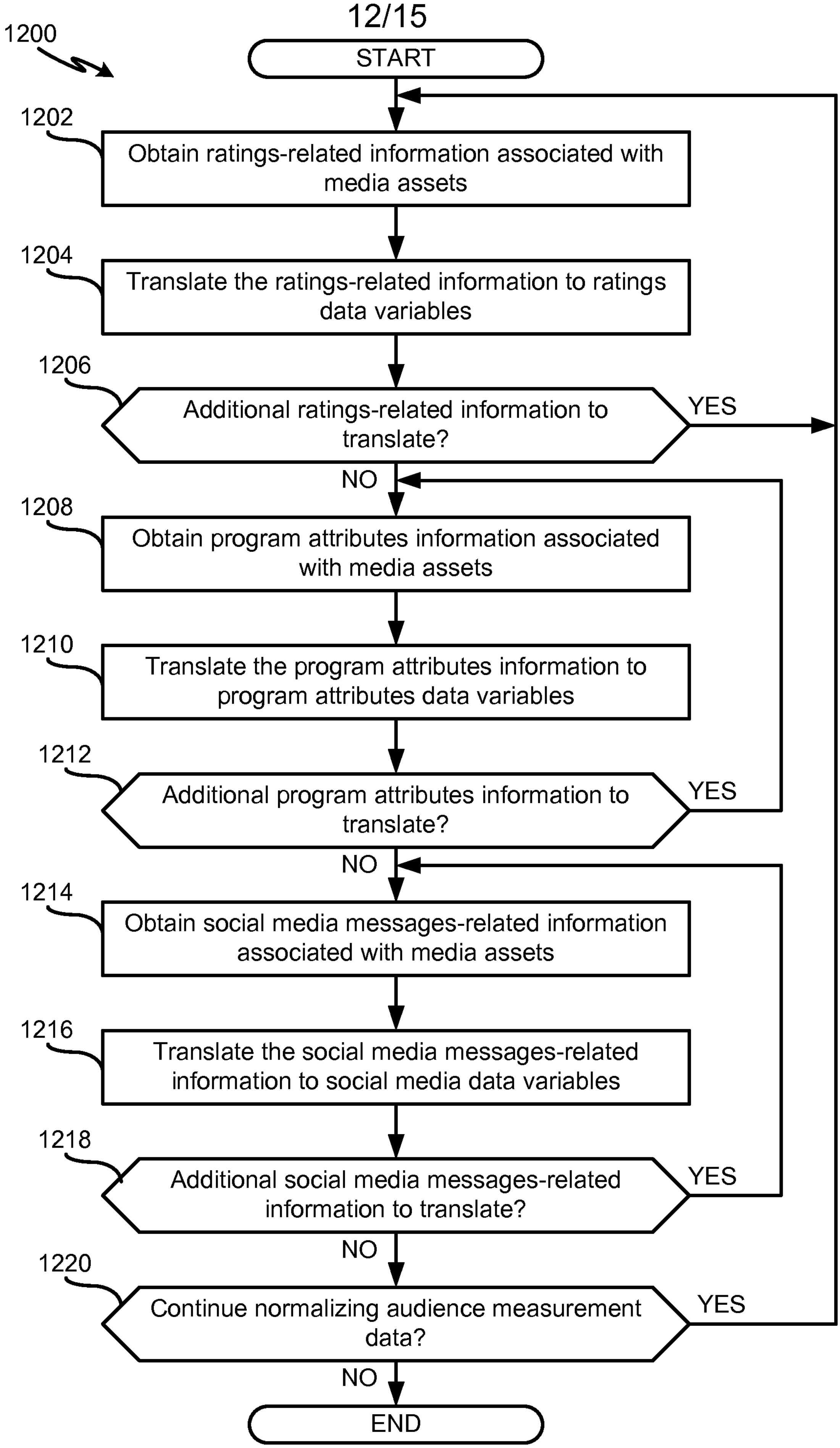

1200
START
1202
Obtain ratings-related information associated with media assets
1204
Translate the ratings-related information to ratings data variables
1206
Additional ratings-related information to translate?
YES
NO
1208
Obtain program attributes information associated with media assets
1210
Translate the program attributes information to program attributes data variables
1212
Additional program attributes information to translate?
YES
NO
1214
Obtain social media messages-related information associated with media assets
1216
Translate the social media messages-related information to social media data variables
1218
Additional social media messages-related information to translate?
YES
NO
1220
Continue normalizing audience measurement data?
YES
NO
END

FIG. 12

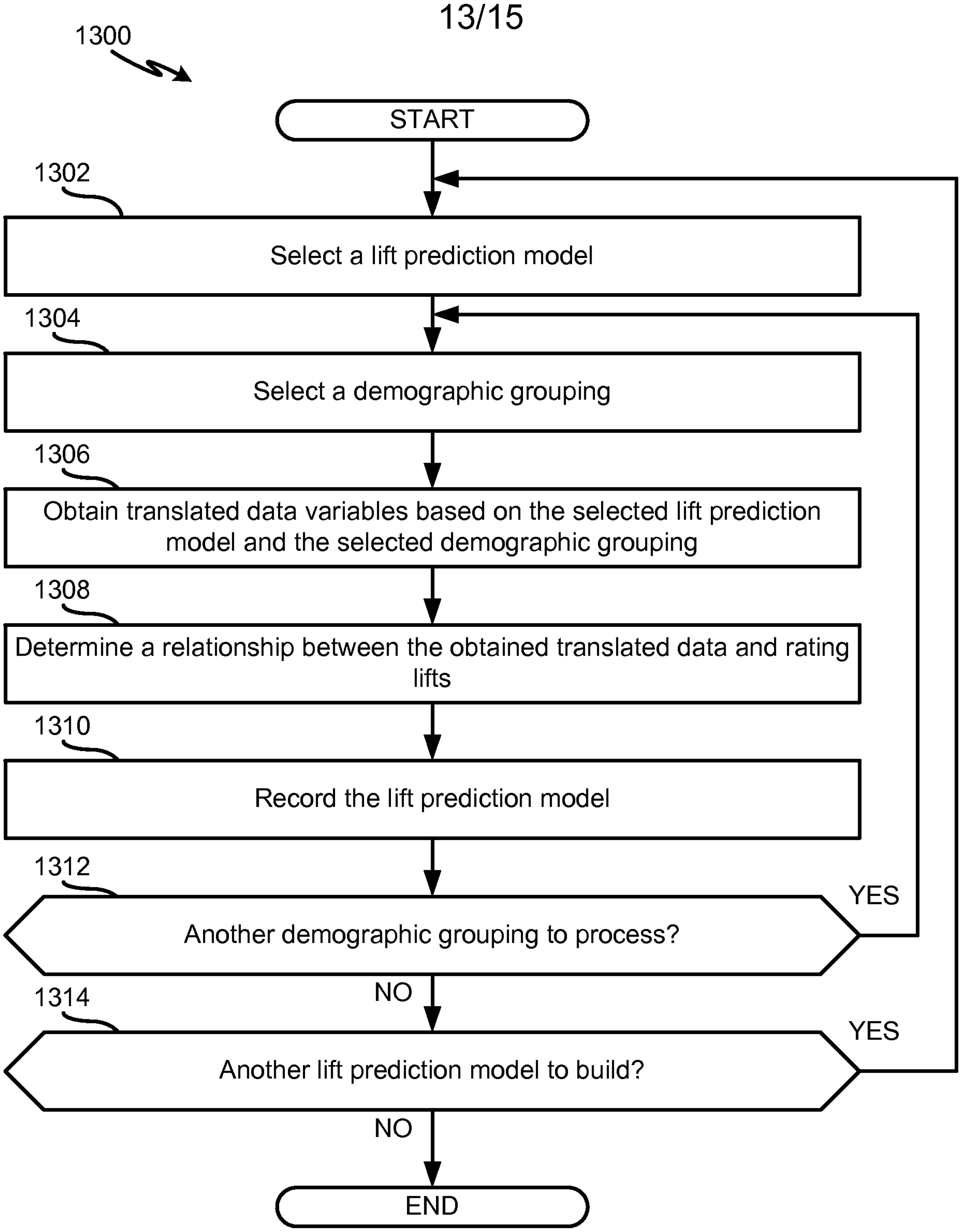

1300
START
1302
Select a lift prediction model
1304
Select a demographic grouping
1306
Obtain translated data variables based on the selected lift prediction model and the selected demographic grouping
1308
Determine a relationship between the obtained translated data and rating lifts
1310
Record the lift prediction model
1312
Another demographic grouping to process?
YES
NO
1314
Another lift prediction model to build?
YES
NO
END

FIG. 13

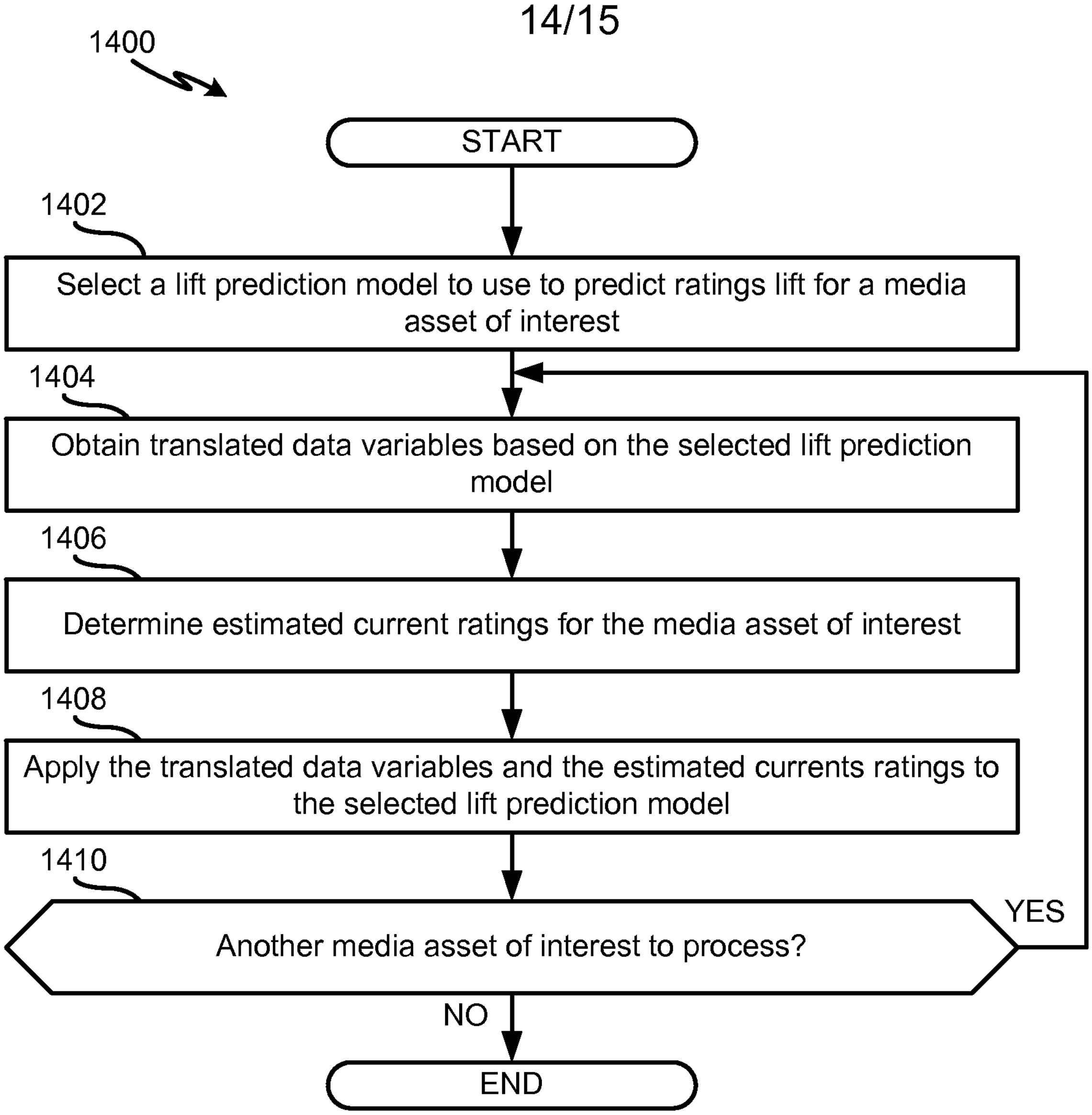

1400
START
1402
Select a lift prediction model to use to predict ratings lift for a media asset of interest
1404
Obtain translated data variables based on the selected lift prediction model
1406
Determine estimated current ratings for the media asset of interest
1408
Apply the translated data variables and the estimated currents ratings to the selected lift prediction model
1410
Another media asset of interest to process?
YES
NO
END

FIG. 14

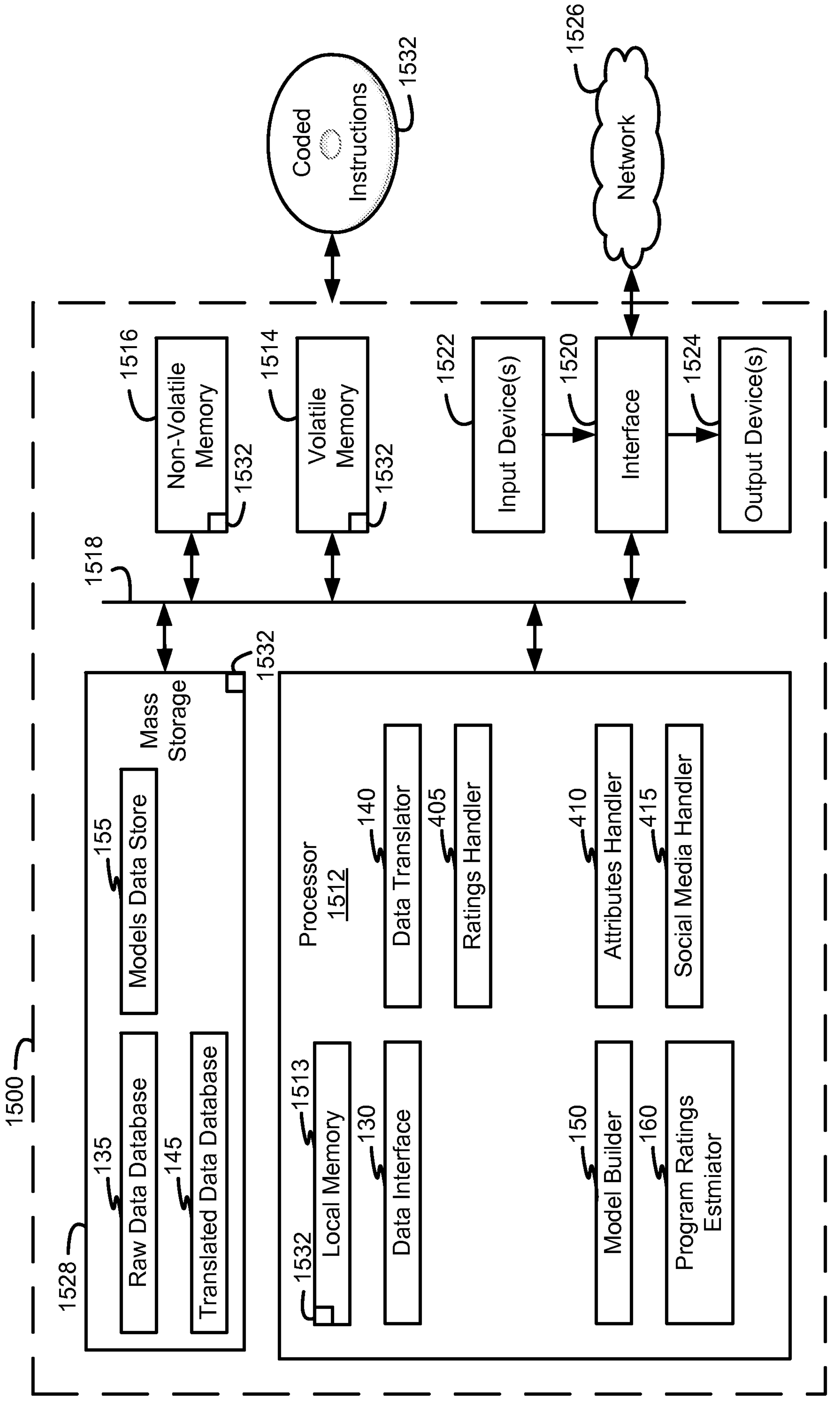

Coded Instructions
1532
Network
1526
1500
1516
Non-Volatile Memory
1532
1514
Volatile Memory
1532
1522
Input Device(s)
1520
Interface
1524
Output Device(s)
1518
1532
Mass Storage
1528
135
Raw Data Database
145
Translated Data Database
155
Models Data Store
Processor
1512
1513
Local Memory
1532
130
Data Interface
140
Data Translator
405
Ratings Handler
150
Model Builder
160
Program Ratings Estmiator
410
Attributes Handler
415
Social Media Handler

FIG. 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/062551

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 30/02(2012.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06Q 30/02; H04L 29/08; H04H 60/32; H04H 60/33; H04H 60/29

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: predict, time, shift, exposure, data, media

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2013-0346154 A1 (JOSEPHINE HOLZ et al.) 26 December 2013 See paragraphs [0033],[0053],[0069],[0071] and claim 1.	1-5,7-20
Y	US 2012-0254911 A1 (PETER CAMPBELL DOE) 04 October 2012 See abstract and claim 3.	1-5,7-20
A	US 2010-0242061 A1 (GUTMAN LEVITAN) 23 September 2010 See abstract and claims 1-2.	1-5,7-20
A	US 2009-0158309 A1 (HANKYU MOON et al.) 18 June 2009 See abstract, claims 1-3 and figures 12-13.	1-5,7-20
A	US 2014-0075018 A1 (UMBEL CORPORATION) 13 March 2014 See abstract, claims 1-3 and figures 2-3.	1-5,7-20

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 March 2016 (24.03.2016)	**28 March 2016 (28.03.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
COMMISSIONER
Telephone No. +82-42-481-8112

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 6
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claim 6 is missing in this application.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required addtional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of any additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/062551

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2013-0346154 A1	26/12/2013	None	
US 2012-0254911 A1	04/10/2012	AU 2012-201860 A1	18/10/2012
		AU 2012-201860 B2	18/09/2014
		CA 2774848 A1	01/10/2012
		CN 102740140 A	17/10/2012
		CN 102740140 B	19/08/2015
		CN 104954822 A	30/09/2015
US 2010-0242061 A1	23/09/2010	US 07882054 B2	01/02/2011
US 2009-0158309 A1	18/06/2009	US 09161084 B1	13/10/2015
US 2014-0075018 A1	13/03/2014	None	