US 20220122098A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2022/0122098 A1**

ABERMAN (43) **Pub. Date: Apr. 21, 2022**

(54) **METHODS AND SYSTEMS FOR DETERMINING REACH INFORMATION**

(71) Applicant: **SINTEC MEDIA LTD.**, Jerusalem (IL)

(72) Inventor: **SAM ABERMAN**, LONDON (GB)

(73) Assignee: **SINTEC MEDIA LTD.**, JERUSALEM (IL)

(21) Appl. No.: **17/565,782**

(22) Filed: **Dec. 30, 2021**

Related U.S. Application Data

(63) Continuation of application No. 16/554,934, filed on Aug. 29, 2019, now Pat. No. 11,244,327.

(60) Provisional application No. 62/724,447, filed on Aug. 29, 2018.

Publication Classification

(51) **Int. Cl.**

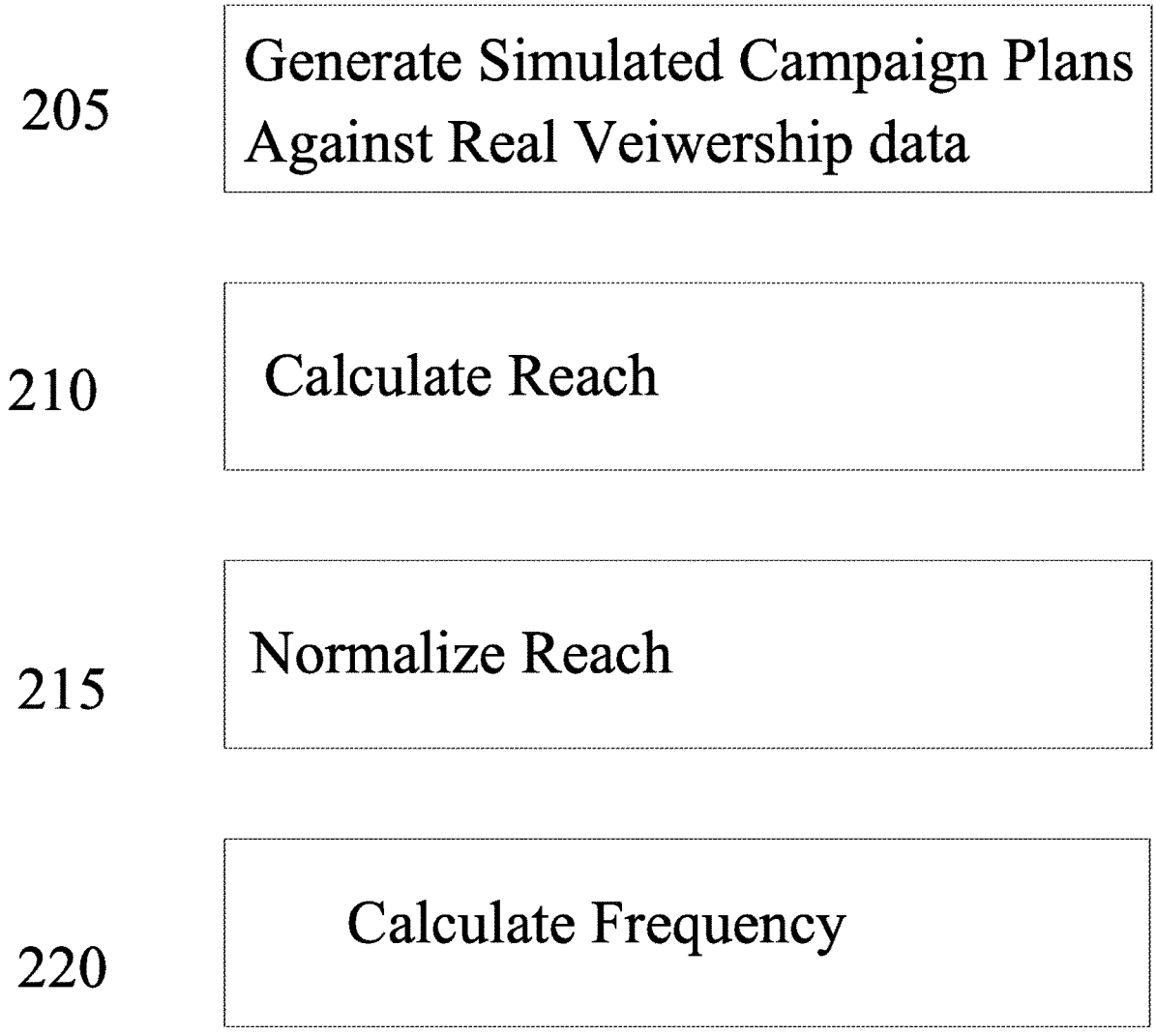
G06Q 30/02	(2006.01)
G06N 20/00	(2006.01)
G06N 5/04	(2006.01)

(52) **U.S. Cl.**

CPC ***G06Q 30/0201*** (2013.01); ***G06N 20/00*** (2019.01); ***G06Q 30/0255*** (2013.01); ***G06N 5/04*** (2013.01); ***G06Q 30/0204*** (2013.01); ***G06Q 30/0254*** (2013.01); ***G06Q 30/0264*** (2013.01); ***G06Q 30/0246*** (2013.01)

(57) **ABSTRACT**

Methods and systems for determining reach information. Watching habits of viewers in a designated market can be analyzed using machine learning algorithms to obtain campaign spot plans. The campaign spot plans can be applied to single viewer data to calculate campaign spot plan Television Average Ratings Points (TARP) pattern information. The TARP pattern information can be translated into reach information determining how many people were uniquely exposed to each campaign spot.



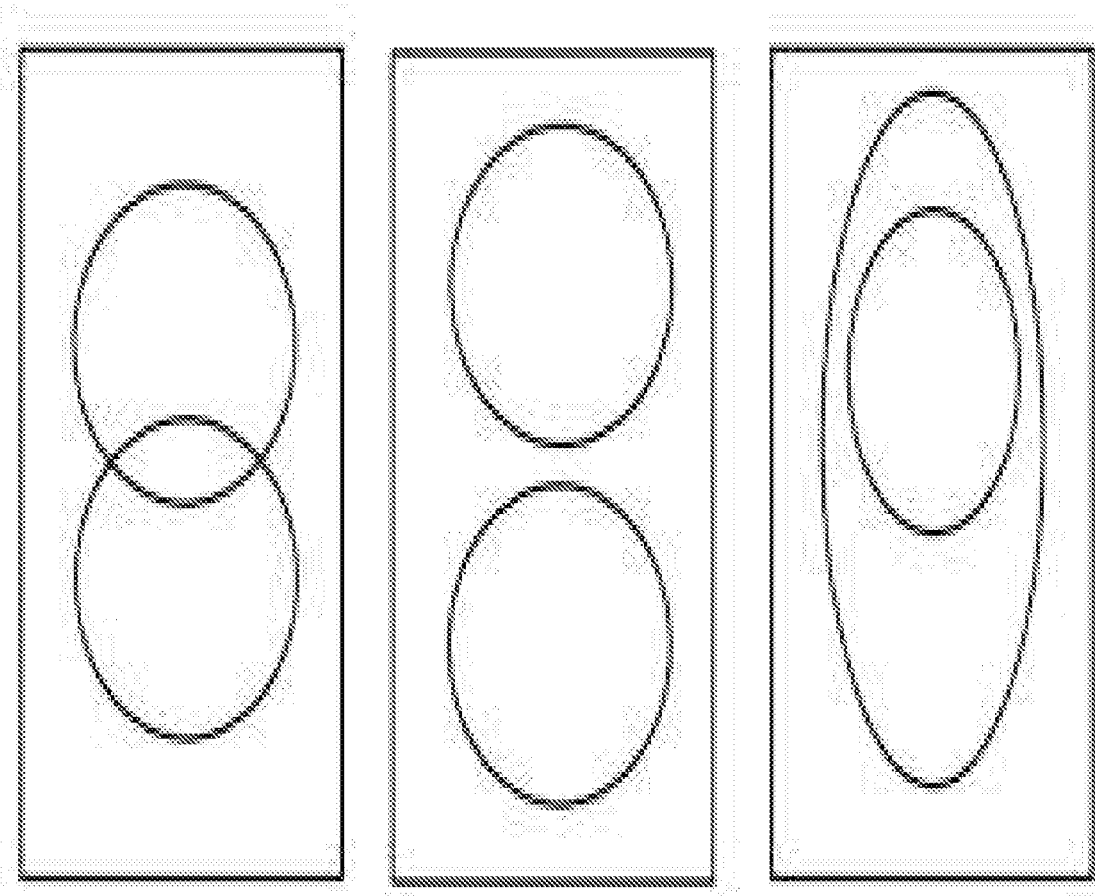
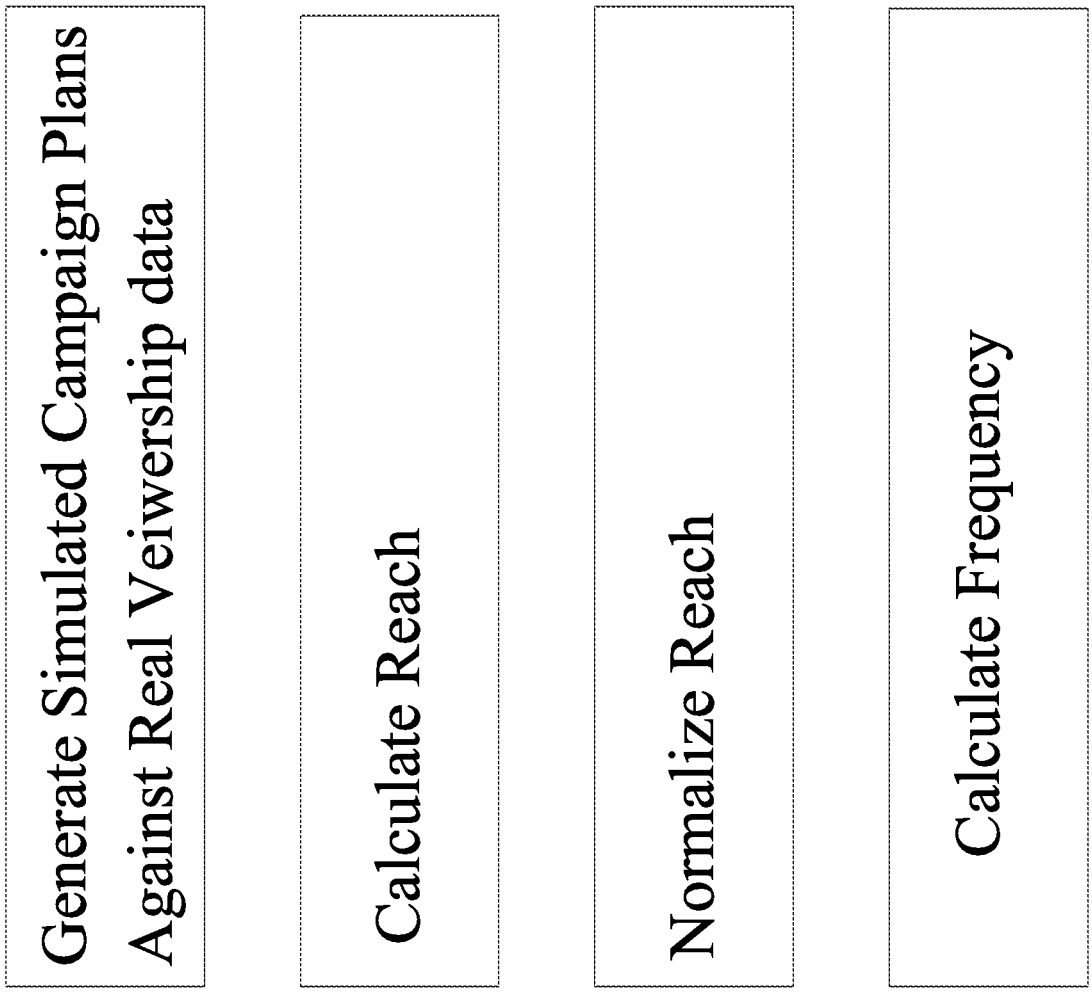


FIG.1A

FIG.1B

FIG.1C

FIG. 2


Generate Simulated Campaign Plans Against Real Veiwership data
Calculate Reach
Normalize Reach
Calculate Frequency
205
210
215
220

NNBC IA25-54 RNNBC IA25-54 RF testing.xls testing

Campaign NNBC IA25-54 R&F testing
Market(s) LOS ANGELES
Agency
Advertiser
Spot Length(s) :30
Campaign Dates 1/1/2018-3/25/2018
Book(s) LPJan/2018LS

AE/SR: Barbara Potasky
Email: barbara.potasky@nbcuni.com
Web:
Exported: 3/19/2018

Station	Program Time	Book	000 P25-54	Spots Len	1/1/2018	1/8/2018	1/15/2018	1/22/2018	1/29/2018	2/5/2018	2/12/2018	2/19/2018	2/26/2018	3/5/2018	3/12/2018	3/19/2018	Total Spots Total Cost
NNBC	FRASIER Mo-Fr 8:00a-9:00a	LPJan/2018LS	0.5	Spots :30	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	12 $300.00
NNBC	WYATT EARP Mo-Su 6:00a-7:00a	LPJan/2018LS	0.3	Spots :30	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	12 $300.00
NNBC	THE REAL MCCOYS MTWRF,U 7:00a-8:00a	LPJan/2018LS	0.6	Spots :30	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	24 $600.00
NNBC	THE VIRGINIAN Sat 7:00a-10:00a	LPJan/2018LS	0.2	Spots :30	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	1 $25.00	12 $300.00
NNBC	WILL & GRACE Mo-Fr 9:00a-10:00a	LPJan/2018LS	0.8	Spots :30	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	2 $25.00	24 $600.00
NNBC	THE NANNY Mo-Fr 10:00a-11:00a	LPJan/2018LS	2.1	Spots :30	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	24 $1,200.00
NNBC	LITTLE HOUSE 11A Mo-Fr 11:00a-12:00p	LPJan/2018LS	2.1	Spots :30	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	12 $600.00
NNBC	ADAM-12 Mo-Fr 12:00p-1:00p	LPJan/2018LS	8.8	Spots :30	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	24 $1,200.00
NNBC	EMERGENCY Mo-Fr 1:00p-2:00p	LPJan/2018LS	5.9	Spots :30	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	3 $50.00	36 $1,800.00
NNBC	QUINCY Mo-Fr 2:00p-3:00p	LPJan/2018LS	7.8	Spots :30	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	12 $600.00
NNBC	MURDER SHE WROTE 3P Mo-Fr 3:00p-4:00p	LPJan/2018LS	2.2	Spots :30	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	12 $600.00
NNBC	MCMILLAN & WIFE Sat 11:30a-1:00p	LPJan/2018LS	1.7	Spots :30	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	12 $600.00
NNBC	A-TEAM Sat 1:00p-3:00p	LPJan/2018LS	2.8	Spots :30	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	12 $600.00
NNBC	MIAMI VICE Sat 3:00p-5:00p	LPJan/2018LS	3.4	Spots :30	1 $50.00	0 $50.00	1 $50.00	0 $50.00	1 $50.00	0 $50.00	1 $50.00	0 $50.00	1 $50.00	0 $50.00	1 $50.00	0 $50.00	6 $300.00
NNBC	ROCKFORD FILES Sun 4:00p-6:00p	LPJan/2018LS	4.3	Spots :30	1 $50.00	2 $50.00	1 $50.00	1 $50.00	2 $50.00	1 $50.00	1 $50.00	2 $50.00	1 $50.00	1 $50.00	2 $50.00	1 $50.00	16 $800.00
NNBC	LITTLE HOUSE 5P Mo-Fr 5:00p-6:00p	LPJan/2018LS	1.5	Spots :30	2 $50.00	1 $50.00	2 $50.00	1 $50.00	2 $50.00	1 $50.00	2 $50.00	1 $50.00	2 $50.00	1 $50.00	2 $50.00	1 $50.00	18 $900.00
NNBC	LITTLE HOUSE 6P Mo-Fr 6:00p-7:00p	LPJan/2018LS	2.6	Spots :30	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	2 $50.00	24 $1,200.00
NNBC	THE NANNY Mo-Fr 7:00p-8:00p	LPJan/2018LS	2.8	Spots :30	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	2 $75.00	24 $1,800.00
NNBC	KOJAK Sat 5:00p-7:00p	LPJan/2018LS	1.0	Spots :30	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	12 $600.00
NNBC	BIONIC WOMAN Sun 6:00p-7:00p	LPJan/2018LS	2.4	Spots :30	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	1 $50.00	12 $600.00
NNBC	SIX MILLION DOLLAR MAN Sun 7:00p-8:00p	LPJan/2018LS	0.5	Spots :30	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	1 $75.00	12 $900.00
NNBC	FRASIER Mo-Fr 8:00p-10:00p	LPJan/2018LS	7.8	Spots :30	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	24 $2,400.00
NNBC	WILL & GRACE Mo-Fr 10:00p-11:00p	LPJan/2018LS	7.2	Spots :30	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	2 $100.00	24 $2,400.00
NNBC	COLUMBO SA Sat 7:00p-10:00p	LPJan/2018LS	3.6	Spots :30	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	12 $1,200.00
NNBC	ROCKFORD FILES SA Sat 10:00p-12:00a	LPJan/2018LS	5.5	Spots :30	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	12 $1,200.00
NNBC	MUNSTERS Sun 8:00p-10:00p	LPJan/2018LS	3.8	Spots :30	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	1 $100.00	12 $1,200.00
NNBC	WILL & GRACE Mo-Fr 11:00p-12:00a	LPJan/2018LS	6.7	Spots :30	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	3 $75.00	36 $2,700.00
NNBC	THE NANNY Mo-Fr 12:00a-1:00a	LPJan/2018LS	5.3	Spots :30	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	12 $120.00
NNBC	QUANTUM LEAP Mo-Fr 2:00a-3:00a	LPJan/2018LS	2.5	Spots :30	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	12 $120.00
NNBC	MAKE ROOM FOR DADDY Mo-Fr 3:00a-4:00a	LPJan/2018LS	1.6	Spots :30	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	1 $10.00	12 $120.00
NNBC	MAGNUM, PI Sun 10:00p-12:00a	LPJan/2018LS	3.8	Spots :30	0 $100.00	1 $100.00	0 $100.00	0 $100.00	1 $100.00	0 $100.00	0 $100.00	0 $100.00	1 $100.00	0 $100.00	0 $100.00	2 $100.00	5 $500.00

	Demo	Len	Totals
Total Imp(000)	P25-54	:30	1,899.2
Total Avg CPM		:30	$14.93
Total Spots			513
Total Cost			$28,360.00

FIG. 3

FIG. 4

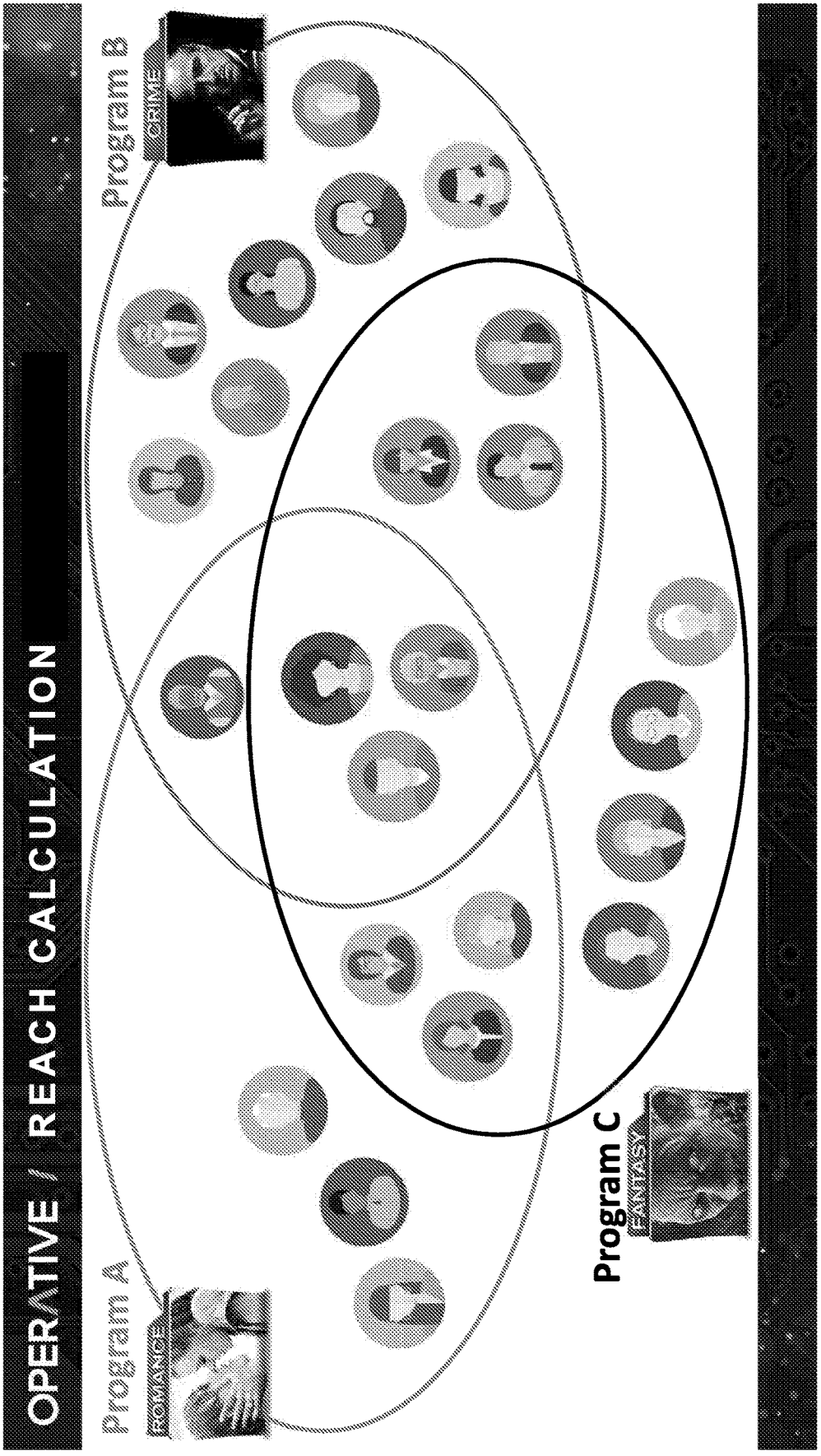

OPERATIVE / REACH CALCULATION
Program A
ROMANCE
Program B
CRIME
Program C
FANTASY

FIG. 5

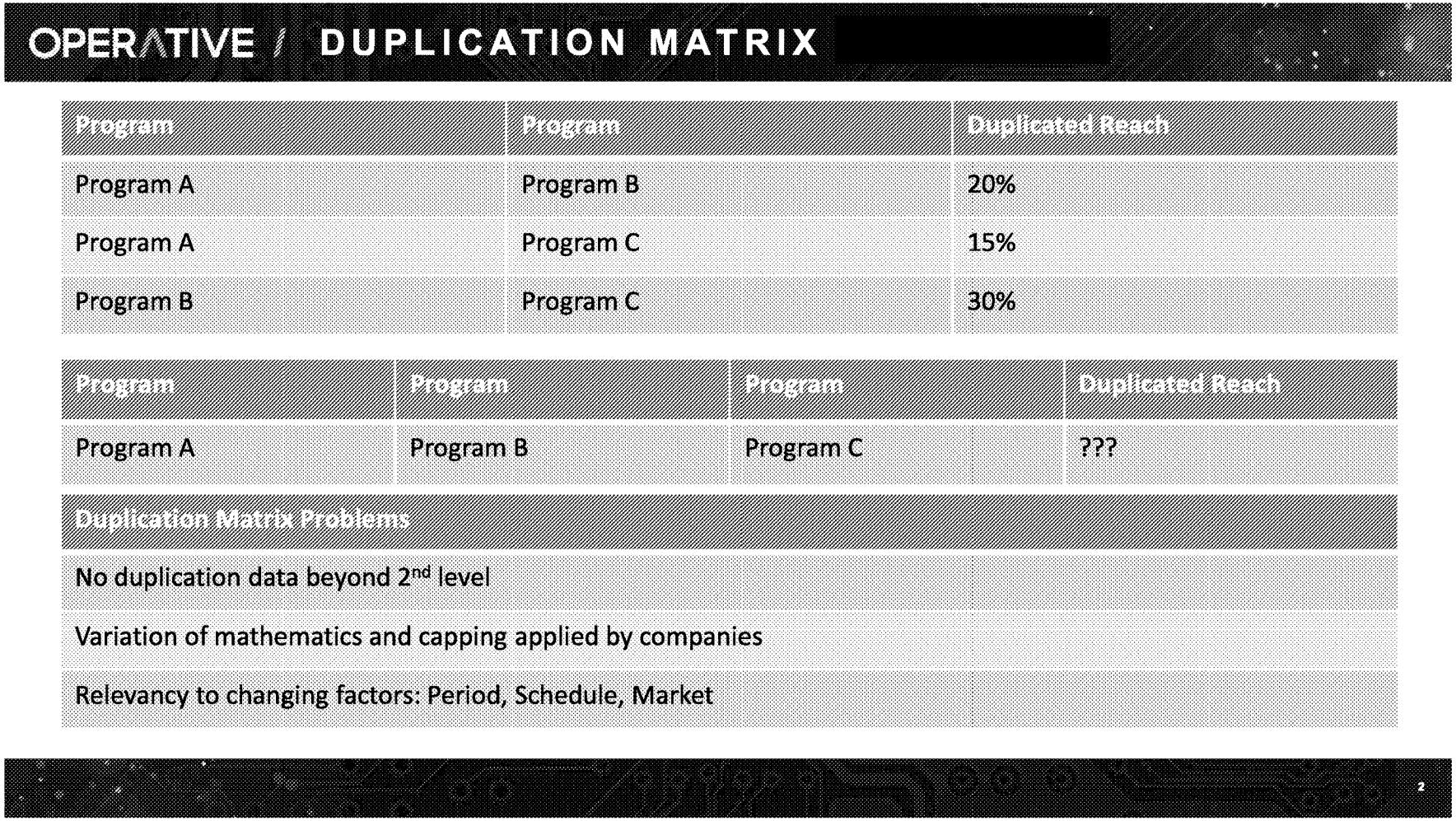

OPERATIVE / DUPLICATION MATRIX
Program
Program
Duplicated Reach
Program A
Program B
20%
Program A
Program C
15%
Program B
Program C
30%
Program
Program
Program
Duplicated Reach
Program A
Program B
Program C
???
Duplication Matrix Problems
No duplication data beyond 2nd level
Variation of mathematics and capping applied by companies
Relevancy to changing factors: Period, Schedule, Market
2

FIG. 6

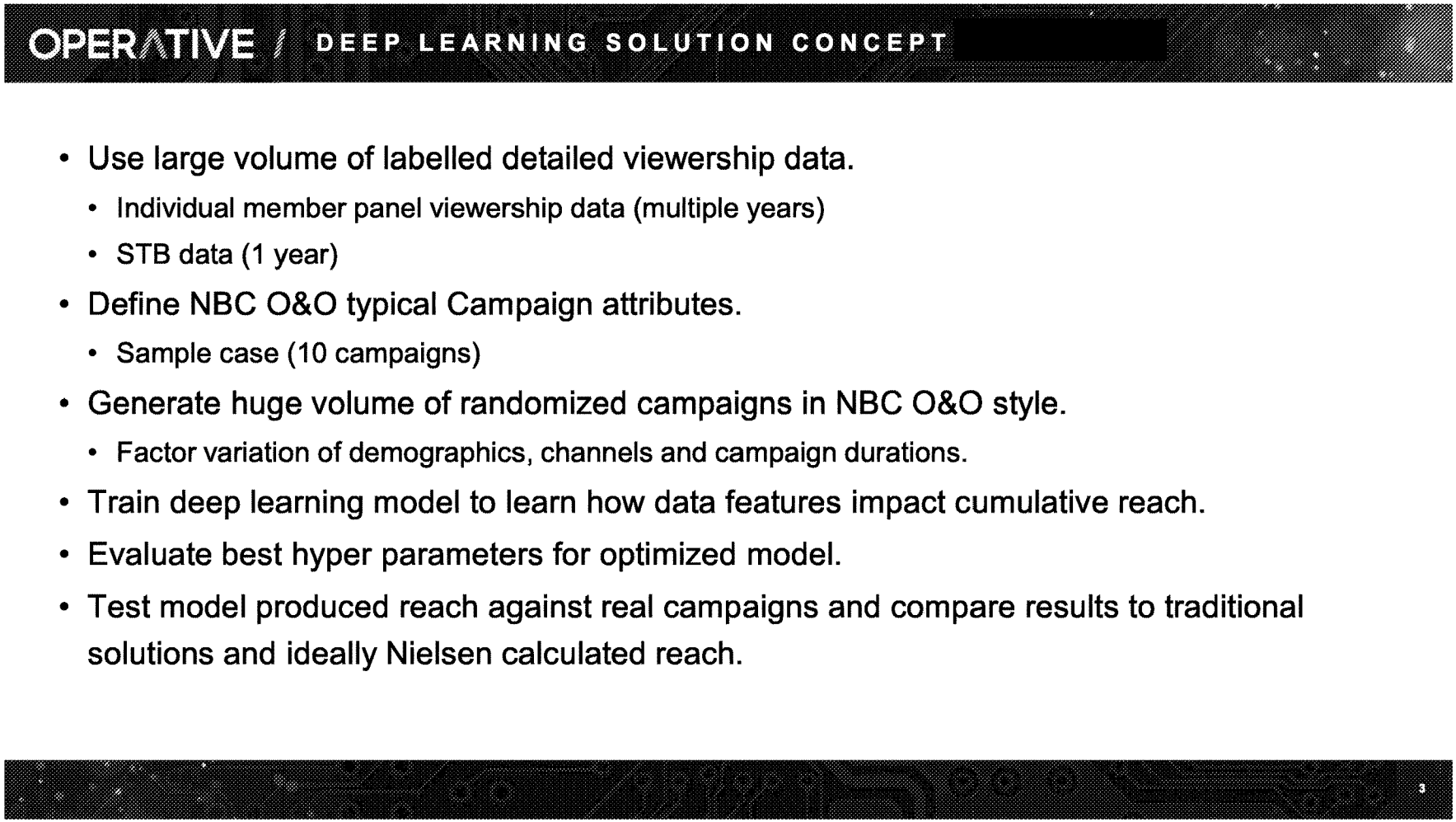

OPERATIVE / DEEP LEARNING SOLUTION CONCEPT
• Use large volume of labelled detailed viewership data.
• Individual member panel viewership data (multiple years)
• STB data (1 year)
• Define NBC O&O typical Campaign attributes.
• Sample case (10 campaigns)
• Generate huge volume of randomized campaigns in NBC O&O style.
• Factor variation of demographics, channels and campaign durations.
• Train deep learning model to learn how data features impact cumulative reach.
• Evaluate best hyper parameters for optimized model.
• Test model produced reach against real campaigns and compare results to traditional solutions and ideally Nielsen calculated reach.
3

FIG. 7

OPERATIVE / LEARNING VIEWING PATTERN [redacted]

Parameter	Program	Program	Program
Content	Program A	Program B	Program C
Channel/Station	Station A	Station A	Station B
Date	Jan 10th	Jan 10th	Jan 11th
Day of Week	Friday	Friday	Saturday
Weekday/Weekend	Weekday	Weekday	Weekend
Start Time	20:00	22:00	21:00
Day Part	Primetime	Primetime	Primetime
Rating By Demo	R1	R2	R3

FIG. 8

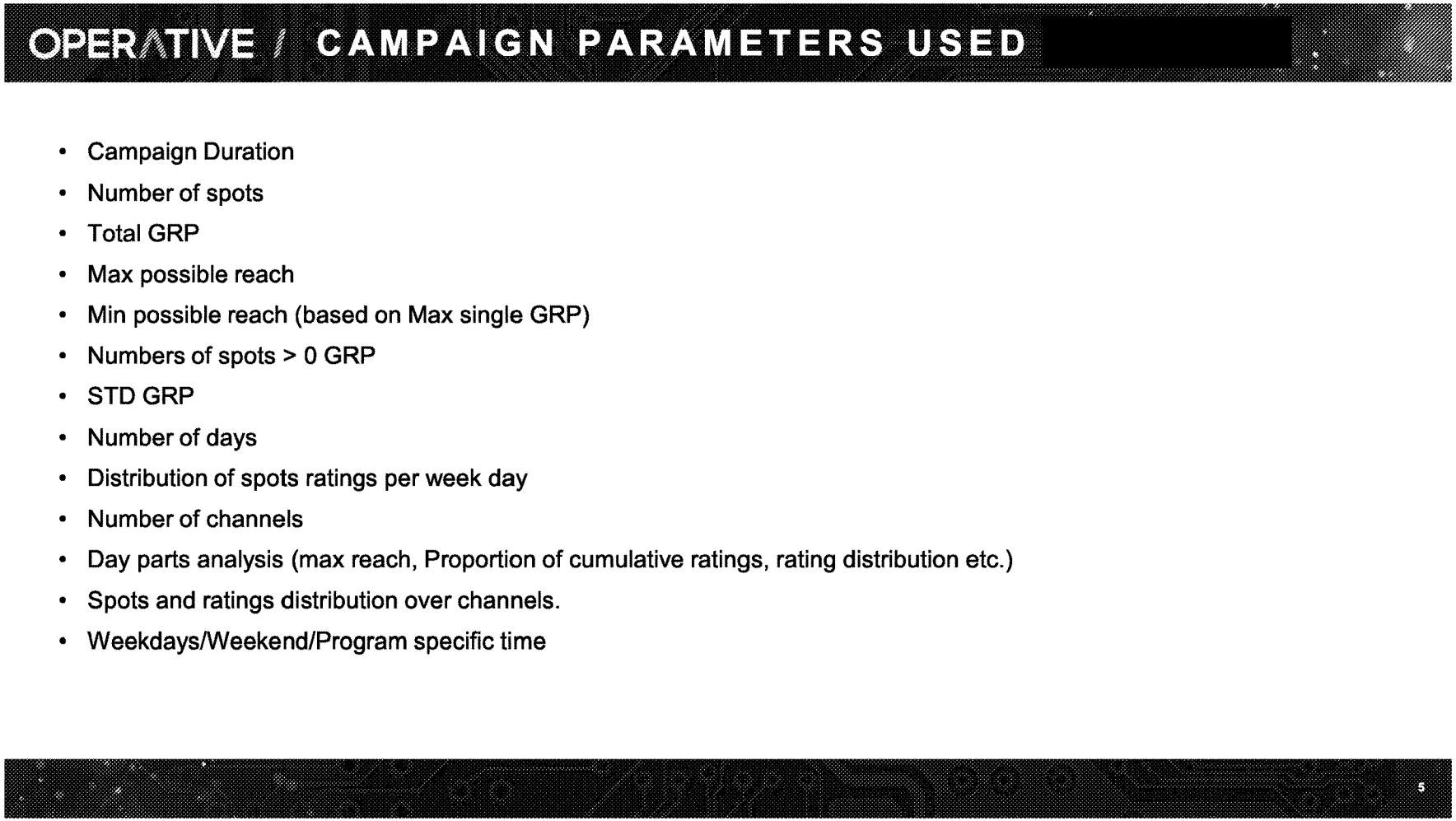
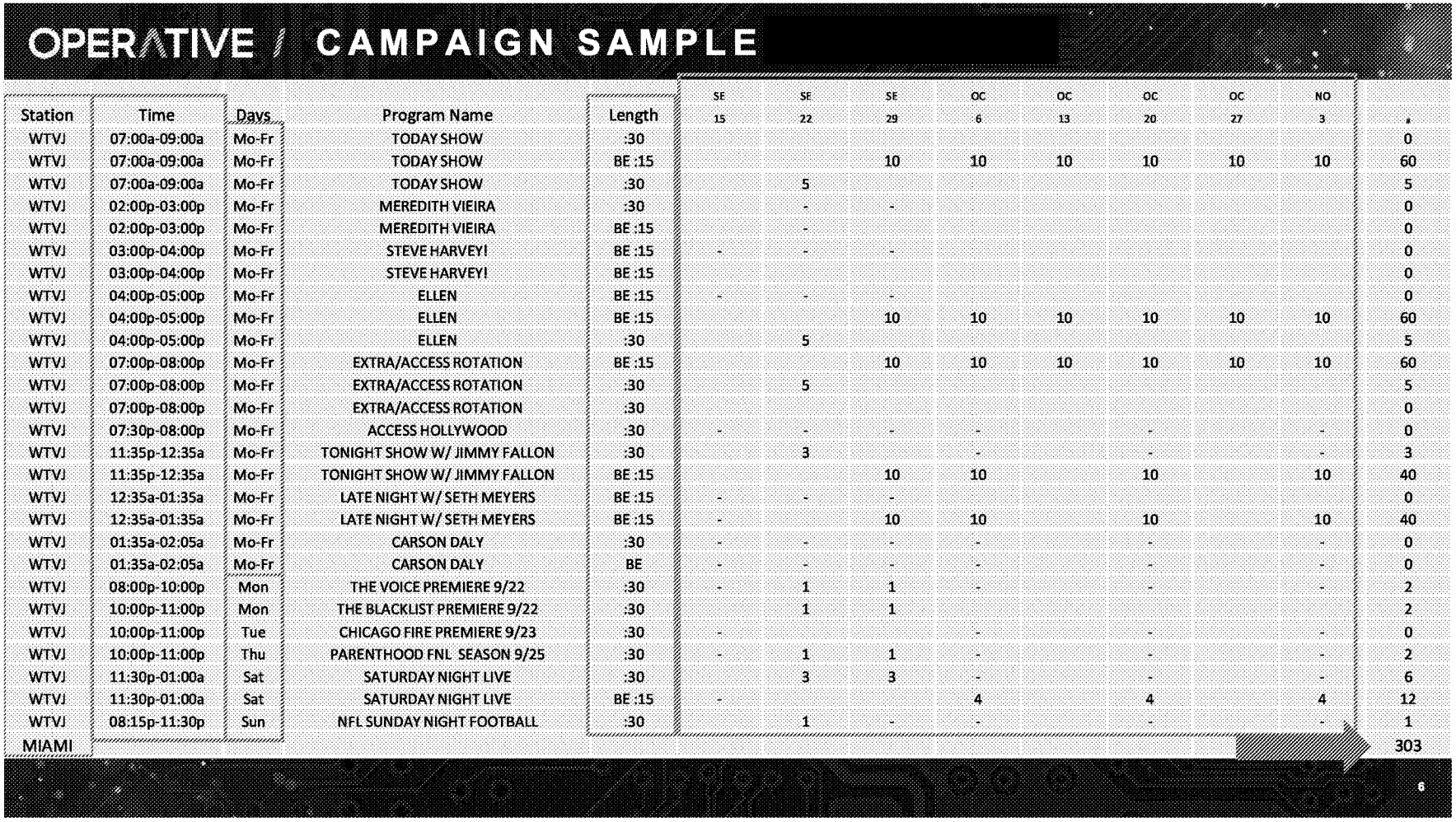

OPERATIVE / CAMPAIGN PARAMETERS USED
• Campaign Duration
• Number of spots
• Total GRP
• Max possible reach
• Min possible reach (based on Max single GRP)
• Numbers of spots > 0 GRP
• STD GRP
• Number of days
• Distribution of spots ratings per week day
• Number of channels
• Day parts analysis (max reach, Proportion of cumulative ratings, rating distribution etc.)
• Spots and ratings distribution over channels.
• Weekdays/Weekend/Program specific time
5

FIG. 9



Station	Time	Days	Program Name	Length	SE 15	SE 22	SE 29	OC 6	OC 13	OC 20	OC 27	NO 3	#
WTVJ	07:00a-09:00a	Mo-Fr	TODAY SHOW	:30									0
WTVJ	07:00a-09:00a	Mo-Fr	TODAY SHOW	BE :15			10	10	10	10	10	10	60
WTVJ	07:00a-09:00a	Mo-Fr	TODAY SHOW	:30		5							5
WTVJ	02:00p-03:00p	Mo-Fr	MEREDITH VIEIRA	:30		-	-						0
WTVJ	02:00p-03:00p	Mo-Fr	MEREDITH VIEIRA	BE :15		-							0
WTVJ	03:00p-04:00p	Mo-Fr	STEVE HARVEY!	BE :15	-	-	-						0
WTVJ	03:00p-04:00p	Mo-Fr	STEVE HARVEY!	BE :15									0
WTVJ	04:00p-05:00p	Mo-Fr	ELLEN	BE :15	-	-	-						0
WTVJ	04:00p-05:00p	Mo-Fr	ELLEN	BE :15			10	10	10	10	10	10	60
WTVJ	04:00p-05:00p	Mo-Fr	ELLEN	:30		5							5
WTVJ	07:00p-08:00p	Mo-Fr	EXTRA/ACCESS ROTATION	BE :15			10	10	10	10	10	10	60
WTVJ	07:00p-08:00p	Mo-Fr	EXTRA/ACCESS ROTATION	:30		5							5
WTVJ	07:00p-08:00p	Mo-Fr	EXTRA/ACCESS ROTATION	:30									0
WTVJ	07:30p-08:00p	Mo-Fr	ACCESS HOLLYWOOD	:30	-	-	-	-		-		-	0
WTVJ	11:35p-12:35a	Mo-Fr	TONIGHT SHOW W/ JIMMY FALLON	:30		3		-		-		-	3
WTVJ	11:35p-12:35a	Mo-Fr	TONIGHT SHOW W/ JIMMY FALLON	BE :15			10	10		10		10	40
WTVJ	12:35a-01:35a	Mo-Fr	LATE NIGHT W/ SETH MEYERS	BE :15	-	-	-						0
WTVJ	12:35a-01:35a	Mo-Fr	LATE NIGHT W/ SETH MEYERS	BE :15	-		10	10		10		10	40
WTVJ	01:35a-02:05a	Mo-Fr	CARSON DALY	:30	-	-	-	-		-		-	0
WTVJ	01:35a-02:05a	Mo-Fr	CARSON DALY	BE	-	-	-	-		-		-	0
WTVJ	08:00p-10:00p	Mon	THE VOICE PREMIERE 9/22	:30	-	1	1	-		-		-	2
WTVJ	10:00p-11:00p	Mon	THE BLACKLIST PREMIERE 9/22	:30		1	1						2
WTVJ	10:00p-11:00p	Tue	CHICAGO FIRE PREMIERE 9/23	:30	-			-		-		-	0
WTVJ	10:00p-11:00p	Thu	PARENTHOOD FNL SEASON 9/25	:30	-	1	1	-		-		-	2
WTVJ	11:30p-01:00a	Sat	SATURDAY NIGHT LIVE	:30		3	3	-		-		-	6
WTVJ	11:30p-01:00a	Sat	SATURDAY NIGHT LIVE	BE :15	-			4		4		4	12
WTVJ	08:15p-11:30p	Sun	NFL SUNDAY NIGHT FOOTBALL	:30		1	-	-		-		-	1
MIAMI													303

FIG. 10

OPERATIVE / DEEP LEARNING MODEL

- AWS
- 256Gb RAM
- 32 Cores
- Parallel Computing
- Deep Learning Model
- Tested Configuration: Best of ≈ 300 hyper parameters combinations (Grid search with cross validation)

- Training time: ≈ 1 Week

FIG. 11

OPERATIVE / FEATURE IMPORTANCE

Feature	Importance
Spot (Frequency)	122,411
Minimal possible Reach	97,107
Maximal possible reach	79,341
TARP (Mean)	72,607
Day Number	71,525
Midweek Identifier	67,570
Number of zero rating spots	58,210
TARP (Standard Div)	55,764
TARP (Tuesday Value)	53,851
Weekend Identifier	52,387
TARP (Saturday Value)	51,900

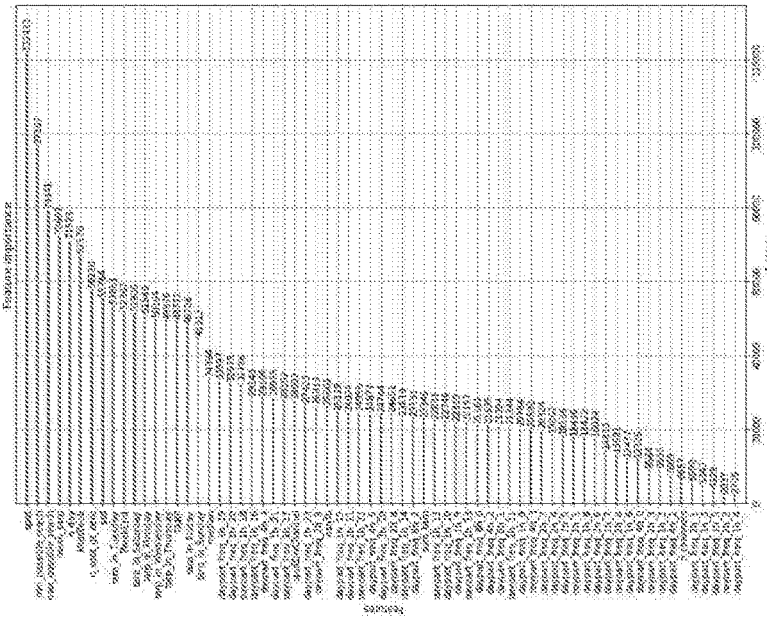


FIG. 12

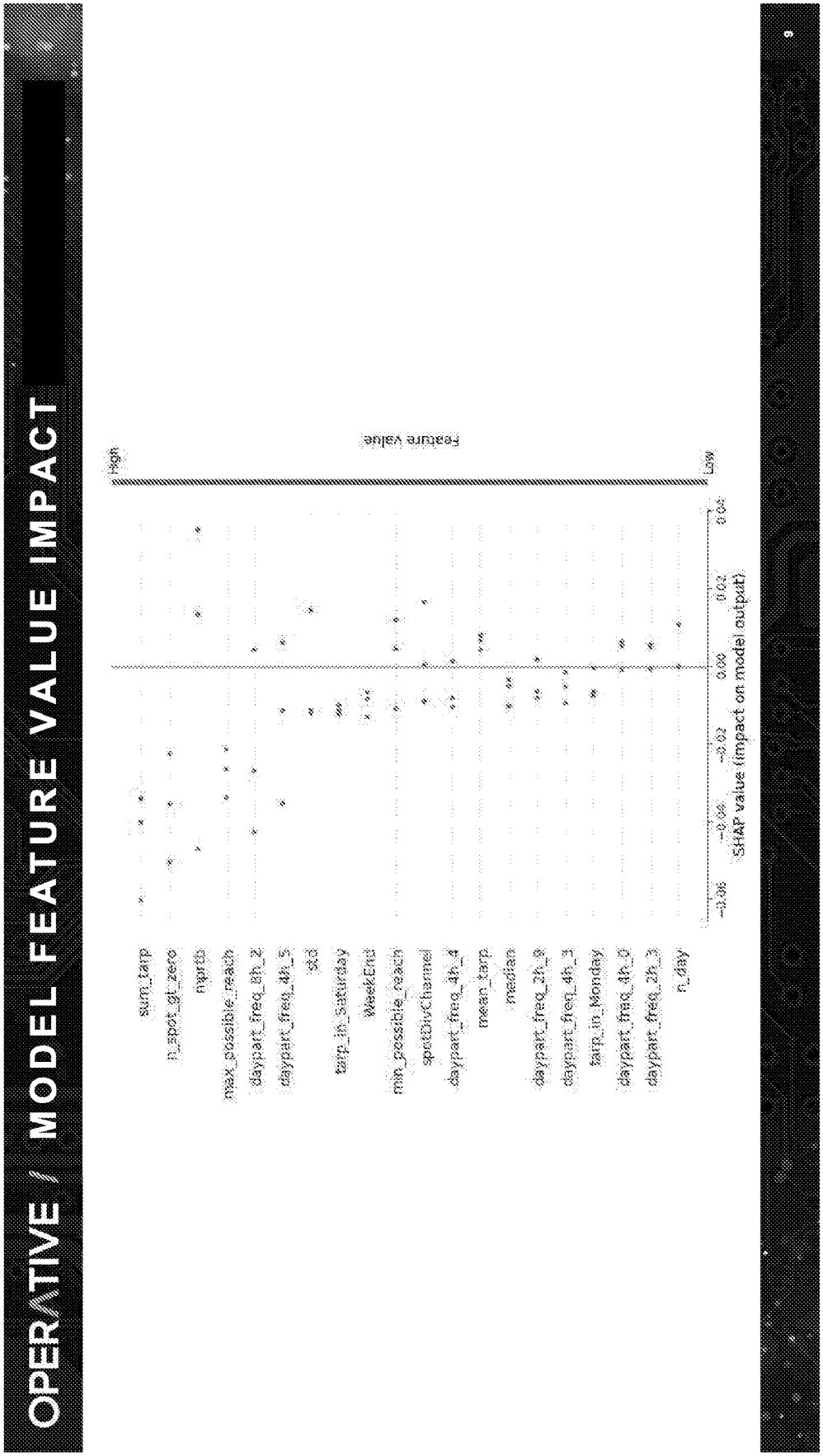
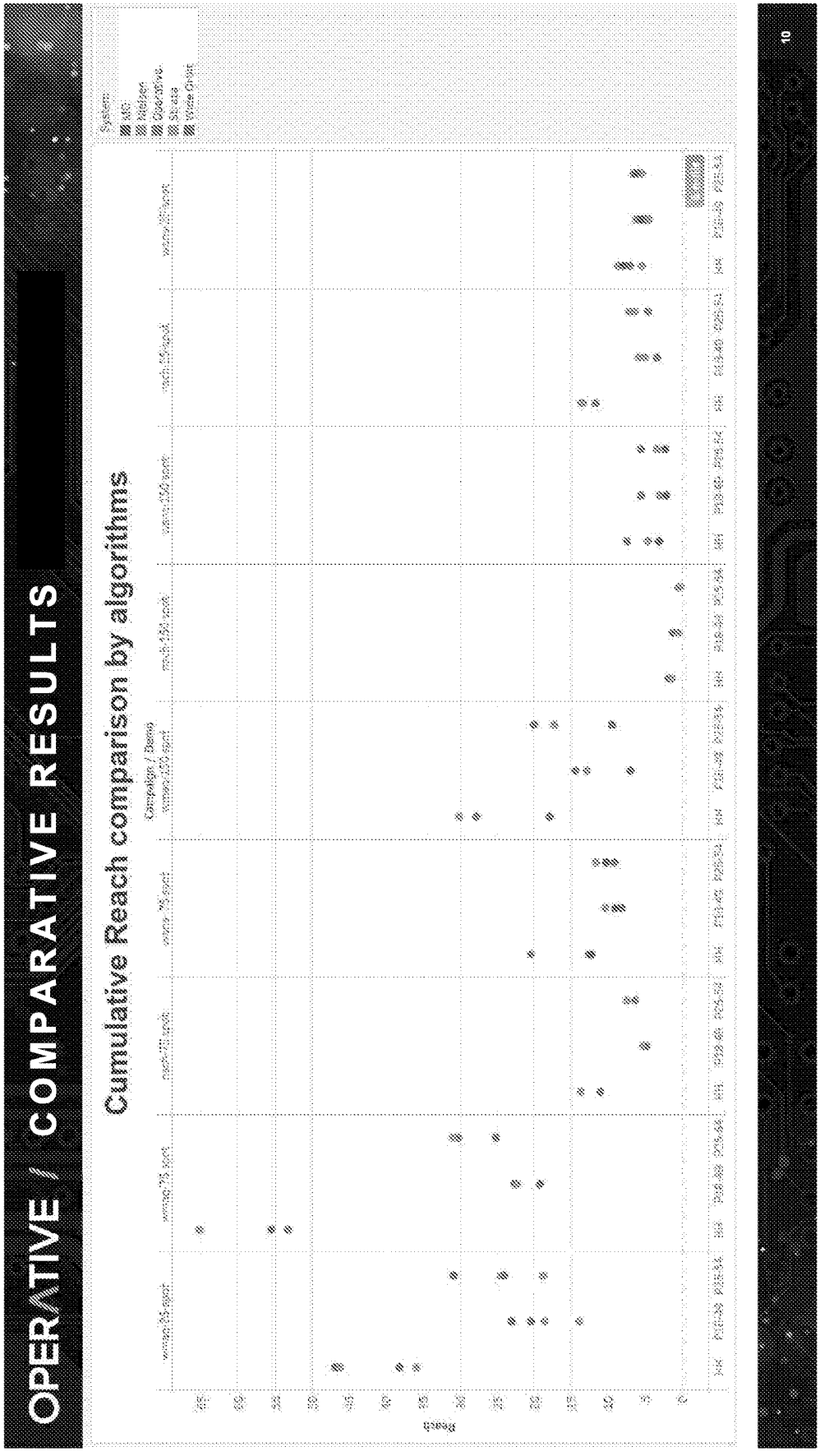

OPERATIVE // MODEL FEATURE VALUE IMPACT
sum_tarp
n_spot_gt_zero
nyrtb
max_possible_reach
daypart_freq_8h_2
daypart_freq_4h_5
std
tarp_in_Saturday
WeekEnd
min_possible_reach
spotDayChannel
daypart_freq_4h_4
mean_tarp
median
daypart_freq_2h_9
daypart_freq_4h_3
tarp_in_Monday
daypart_freq_4h_0
daypart_freq_2h_3
n_day
High
Low
Feature value
-0.06
-0.04
-0.02
0.00
0.02
0.04
SHAP value (impact on model output)

FIG. 13


OPERATIVE / COMPARATIVE RESULTS
Cumulative Reach comparison by algorithms
Reach
System
KO
Meltwi
Operative.
Skada
Vista Grint

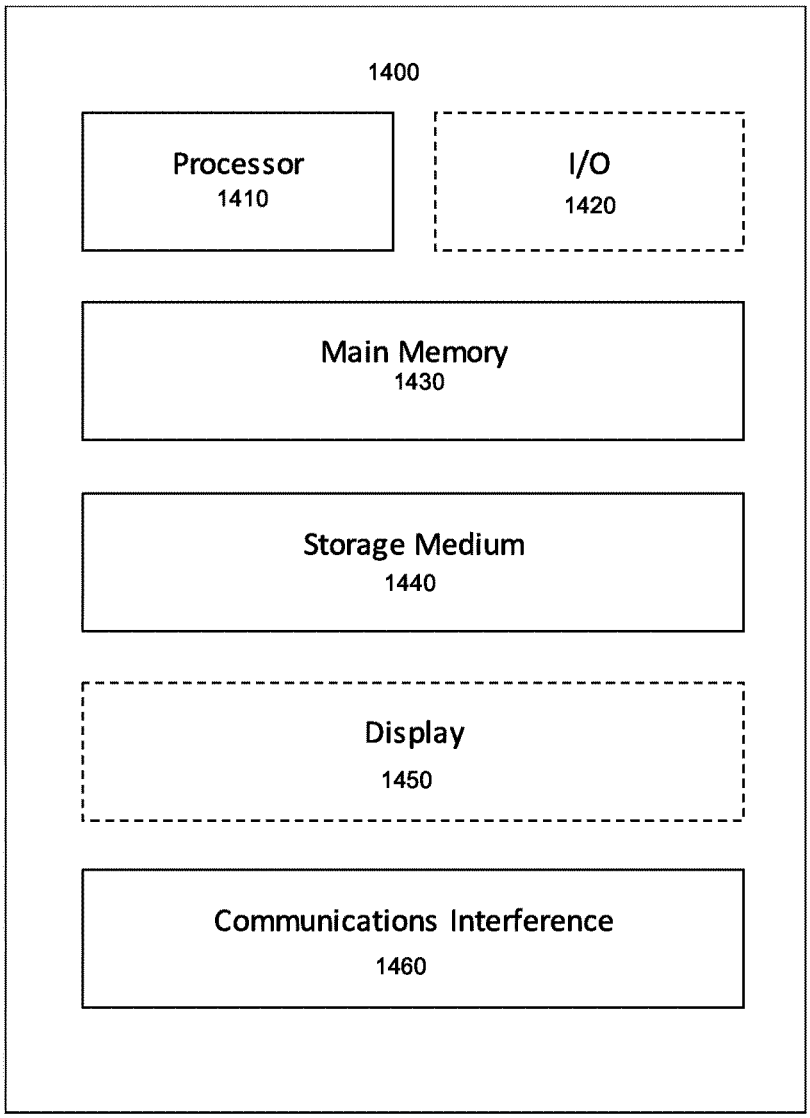

1400
Processor
1410
I/O
1420
Main Memory
1430
Storage Medium
1440
Display
1450
Communications Interference
1460

FIG. 14

METHODS AND SYSTEMS FOR DETERMINING REACH INFORMATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation application to U.S. application Ser. No. 16/554,934, filed Aug. 29, 2019, which claims the benefit of U.S. Provisional Application No. 62/724,447, filed Aug. 29, 2018. These applications are incorporated by reference in their entireties.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] FIGS. **1**A-**1**C illustrate example reach calculations that can be used in embodiments of the invention.

[0003] FIG. **2** illustrates example panel data that can be used against simulated campaign plans in embodiments of the invention.

[0004] FIG. **3** illustrates a simulated campaign plan construct sample that can be used in embodiments of the invention.

[0005] FIG. **4** illustrates a chart where an intersection of people can be seen that can be used in embodiments of the invention.

[0006] FIG. **5** illustrates a chart showing example data that can be used in embodiments of the invention.

[0007] FIG. **6** illustrates an example Reach and Frequency process that can be used in embodiments of the invention.

[0008] FIG. **7**-**8** illustrate example parameters that can be used in embodiments of the invention.

[0009] FIG. **9** illustrates example data that can be used in embodiments of the invention to generate campaign statistics that can be used to train a Reach and Frequency model.

[0010] FIG. **10** illustrates example features of a system that can be used to generate the reach and frequency calculations in some embodiments of the invention.

[0011] FIG. **11** illustrates example features that can be used in some embodiments of the invention.

[0012] FIG. **12** illustrates an example model feature value impact, according to embodiments of the invention.

[0013] FIG. **13** illustrates example comparative results, according to embodiments of the invention.

[0014] FIG. **14** illustrates an example computer that can be used in embodiments of the invention.

DETAILED DESCRIPTION OF EMBODIMENTS

[0015] Systems and methods are disclosed herein for performing advertising Reach and Frequency calculations to determine how many people were uniquely exposed to a broadcaster message (e.g., an advertising campaign).

[0016] Mathematical models based on duplication matrices (e.g., how many people watch a cross of two or more programs) can be used. Machine learning can be utilized along with panel data. Panel data can be data learned from an individual level analysis of viewing patterns (e.g., Nielsen Panels). A group of people who are monitored individually can be selected and weighed to be used in panel data. In some embodiments, the full panel data is utilized with the machine learning. In other embodiments, only a subset of the full panel data is utilized with the machine learning.

[0017] For example, summary reporting of people who watched the different programs where a broadcaster message was shown can be used. The Reach and Frequency calculations can be based on complex training on mass data volume, utilizing machine learning algorithms, and costume feature creation. In some embodiments, numerous (e.g., more than 300,000) advertising campaigns can be processed and analyzed using advanced machine-learning and deep-learning tools in order to predict the Reach and Frequency for a given advertising campaign.

[0018] In some embodiments, the given advertising campaign can be one structured using typical panel data information provided in a particular local market. In some embodiments, the Reach can be for a specific target demographic, using, for example, age & gender constructs. (Note that any type of demographic may be used.)

[0019] In some embodiments, the data that is available to help determine the prediction can be a detailed campaign spot proposal plan, which can contain the following attributes:

[0020] Market

[0021] Station

[0022] Daypart

[0023] Time

[0024] Day, or range of days (Typically Weekdays or Weekend)

[0025] Program Name

[0026] Length (Spot Duration)

[0027] Rate

[0028] Ratings

[0029] The Reach that can be determined can comprise the distinct number of viewers for the campaign (e.g., how many different people saw the campaign at least once). In some situations, mathematical models based on duplication matrix (e.g., how many different people watch a cross of two programs) can be used. However, it becomes complex to use these models when additional programs are added (e.g., how many different people watch programs A+B+C etc.)

[0030] FIGS. **1**A-**1**C illustrate a very simple example. If the campaign contains only two spots, the Reach can vary between the sum of the two ratings (e.g., the maximum possible Reach (e.g., FIG. **1**B) of rating A and rating B, where there is no intersection between the viewers of spot A and the viewers of spot B, and the minimal possible Reach (e.g., FIG. **1**C), where all the viewers of one spot are contained within the viewers of the other spot). FIG. **1**A illustrates a medium case.

[0031] In FIGS. **1**A-**1**C, the Reach is illustrated as the total area of the circles, which can have an inverse correlation to the size of the intersection. (FIG. **4** illustrates a similar chart where the intersection of the people can be seen. FIG. **5** illustrates a chart showing example data.) For example, a small intersection could indicate a large Reach, and a large intersection could indicate a small Reach. The Reach can be calculated using an inclusion-exclusion principle. Thus, for the example above, using the inclusion-exclusion principle, the Reach can be:

$$\text{Reach}=|A|+|B|-|A\cap B|$$

[0032] As another example, for a campaign with three spots, the Reach can be:

$$\text{Reach}=|A|+|B|+|C|-|A\cap C|-|B\cap C|+|A\cap B\cap C|$$

[0033] Similarly, the formula can become exponentially larger with each spot that is added. For a very large campaign, computing the Reach based on this formula is difficult for at least some of the following reasons:

[0034] As campaign spots are added, the amount of intersections that need to be calculated can grow exponentially—e.g. for a campaign with 100 spots, the amount of terms that needed to be computed is 2^100−1. (Note that for an average campaign, the number of spots based on a sample set is approximately 250.)

[0035] It can be difficult to estimate the intersection between n-spots where n is large enough because each such calculation can add noise to the final result. This noise can add up to a significant error on the final Reach calculation.

[0036] We resolved this problem by letting machine learning analyse duplicated watching habits (watching the same time of day, watching the same type of content, watching the same channels etc. etc.)

[0037] In some embodiments, in order to overcome the above issues as well as other issues, methods and systems can be improved by utilizing a machine learning to analyze duplicated watching habits (e.g., watching the same time of day, watching the same type of content, watching the same channels, etc.) to determine Reach patterns.

[0038] FIG. **6** illustrates an example of a process that can be used. As set forth in FIG. **6**, a large volume of labelled detailed viewership data can be utilized (e.g., individual member panel viewership data over years, one year STB data). Campaign attributes (e.g., such as those used by NBC O&O) can be defined for a sample case (e.g., 10 campaigns A large volume of randomized campaigns can be generated in the NBC O&O style. A variation of demographics, channels and campaign durations can be factored in. The deep learning model can be trained to learn how data features impact cumulative reach. The best hyper-parameters can be evaluated for an optimized model. The test model produced reach can then be compared against real campaigns and the results can be compared to traditional solutions.

[0039] FIG. **7** and FIG. **8** illustrate example parameters that can be used.

[0040] For example, as shown in FIG. **2**, in **205**, panel data can be used against simulated campaign plans. The simulated campaign plans should be based on a typical customer campaign structure (e.g., such as one expected to be used by NBC O&O in a particular market).

[0041] Using campaign structure attributes, similar to provided samples, hundreds of thousands of synthetic campaigns (e.g., >300,000 campaigns) can be auto-generated. Variations of number of networks, demographics, campaign durations, spot number in day parts, etc. can be mixed in order to maximize the variety of samples to be used for training.

[0042] Training a deep learning model can involve fine-tuning of hyper parameters in order to optimize the performance of the model (e.g., learning rate, number of iterations, etc.).

[0043] Grid Search can be done with cross validation: For each combination of hyper parameters, the data can be split into train and validation sets. The model can be trained against the train data set and evaluated against the validation data set. After testing each combination, the hyper-parameters combination with the best score on the validation set can be selected as the final model.

[0044] Example pseudocode is below:

```
Previous_score = -inf
For p1 in [param_11, param_12 ,..., param_1A]
  For p2 in [param_21, param_22 ,..., param_2B]
    ...
    For pM in [param_m1, param_M2, param_MC]
      Current_Model=train(campaigns_train_Set)
      Predictions=current_model.predict(campaigns_validation_set)
      Current_Score= compare(predictions, validation_set_values)
      If current_score>previous_score
        Chosen_model=current_model
```

[0045] The construct of each campaign plan can be defined according to expected customer campaign variation needs. For example, a specific day can be used for 40% of the campaign spots, and a range of days (for example Monday-Friday) can be used for the remaining 60% of the spots. (This can be similar to the sample set provided by NBC O&O and sold by them in the local US market.)

[0046] Sample US markets can include some or all of the following in some embodiments:

[0047] LOS ANGELES

[0048] CHICAGO

[0049] DALLAS-FT. WORTH

[0050] HARTFORD & NEW HAVEN

[0051] MIAMI-FT. LAUDERDALE

[0052] NEW YORK

[0053] SAN DIEGO

[0054] SAN FRANCISCO-OAK-SA

[0055] WASHINGTON, DC

[0056] PHILADELPHIA

[0057] In some embodiments, the simulated campaign plans can be preprocessed and additional features can be created for each campaign that helps the algorithm obtain a better prediction. Examples of features that can be added in some embodiments comprise any combination of the following:

[0058] Maximum possible Reach,

[0059] Minimum possible Reach,

[0060] Total Television Average Ratings Points (TARP) for each day,

[0061] Total TARP for each daypart

[0062] In **210**, the final model of the simulated campaign plans data (e.g., the campaign spot plans) can be applied to the known single viewer panel data in order to calculate each spot's TARP patterns, and to "translate" those TARP patterns into a cumulative Reach number for every step of the campaign (e.g., we analyze the campaign after 1 spot, 2 spots, etc.).

[0063] Example pseudocode is below:

```
Campaigns_feature_vectors=[ ]
Labels=[ ]
For c in all_campaigns:
  Current_feature_vector=create_feature_vector(c)
  Campaigns_feature_vectors.append(current_feature_vector)
  Labels.append(reach_of_c)
TRAIN(campaigns_feature_vector=>labels)
```

[0064] In some embodiments, grid-searches can be used in order to find the best hyper-parameters for the model, using cross-validation in order to achieve optimized results (and for example to avoid problems such as overfitting).

[0065] A hyperparameter can be a parameter whose value is set before the learning process begins. By contrast, the values of other parameters can be derived via training. More information on hyper-parameters can be found at Wikipedia (e.g., at the following site, which information is herein incorporated by reference):

[0066] https://en.wikipedia.org/wiki/Hyperparameter_(machine_learning)

[0067] Cross validation (e.g., such as k-fold) can be a model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set. More information on cross validation can be found at Wikipedia (e.g., at the following site, which information is herein incorporated by reference):

[0068] https://en.wikipedia.org/wiki/Cross-validation_(statistics)

[0069] Overfitting can be the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably. More information on overfitting can be found at Wikipedia (e.g., at the following site, which information is herein incorporated by reference):

[0070] https://en.wikipedia.org/wiki/Overfitting

[0071] In **215**, each Reach number can be normalized and converted to a number between 0 and 1, where 0 is the minimum possible Reach and 1 is the maximum possible Reach, according to the following formula:

$$\text{Reach\_normalized} = \frac{\text{Reach\_actual} - \text{Reach\_min}}{\text{Reach\_max} - \text{Reach\_min}}$$

[0072] So, for example, if in a given campaign the maximal possible Reach is 7%, the minimal possible Reach is 3% and the actual Reach is 6%, the normalized Reach would be:

$$\frac{6-3}{7-3} = \frac{3}{4} = 0.75$$

[0073] The normalization Reach can be used so that the algorithm can train on more similar campaigns and improve the learning of the campaign patterns.

[0074] After we get the normalized reach, we can restore the actual reach using the following formula:

$$\text{REACH\_ACTUAL} = \text{NORMALIZED\_REACH} * (\text{REACH\_MAX} - \text{REACH\_MIN}) + \text{REACH\_MIN}$$

[0075] In **220**, the Frequency can be calculated by the formula:

$$\text{Frequency} = \frac{\text{TOTAL\_TARP}}{\text{REACH\_ACTUAL}}$$

[0076] FIG. **3** illustrates a simulated campaign plan construct sample, according to an embodiment. In this example, the campaign is NNBC IA 25-54 R&F testing. The market is Los Angeles. The station column lists the stations. The program and time slot column lists the program names and the time slots involved. The spot length is 30 seconds. The campaign dates are Jan. 1, 2018-Mar. 25, 2018. The book column indicates the book information. The AE/SR (e.g., Barbara Potasky) and her contact information are listed. The date the information was exported is Mar. 19, 2018. The 000 P25-54 column lists the rating per spot for people aged 25-54. The spots length column lists how long the spots are. The columns with dates on them list the prices for various dates. The total spots total cost column lists the total amount of all the spots for a particular row.

[0077] FIG. **9** illustrates other example data that can be used to generate, for example, campaign statistics that can be used to train the Reach and Frequency model:

[0078] Number of Campaigns:¶

[0079] 380,506

[0080] Campaign spots number:

[0081] Max number of spots in campaign: 4,588

[0082] Min number of spots in campaign: 1

[0083] Average number of spots in campaign: 63.49

[0084] Median number of spots in campaign: 20.0

[0085] Std Div of number of spots in campaign: 145.37

[0086] Demographic:

[0087] Number of demographics: 63

[0088] Max number of campaigns in one campaign demographic: 6,099

[0089] Min number of campaigns in one campaign demographic: 5,994

[0090] Average of campaigns per demographic: 6039.77

[0091] Median of campaigns per demographic: 6041.0

[0092] Std Div of campaigns per demographic: 20.67

[0093] Stations:

[0094] Number of stations: 20

[0095] Max of campaigns in one campaign station: 20163

[0096] Min of campaigns in one campaign station: 19620

[0097] Average of campaigns per station: 19897.4

[0098] Median of campaigns per station: 19900.5

[0099] Std Div of campaigns per demographic: 138.631

[0100] Campaign duration (in days):

[0101] Max days of campaigns duration: 365

[0102] Min days of campaigns duration: 1

[0103] Average of campaigns duration: 35.26

[0104] Median of campaigns duration: 21.0

[0105] Std Div of campaigns duration: 43.27

[0106] FIG. **11** illustrates an example of the importance of various features that can be used. FIG. **12** illustrates an example model feature value impact. FIG. **13** illustrates example comparative results.

[0107] FIG. **10** illustrates example features of a system that can be used to generate the reach and frequency calculations. Any or all of the following may be included in such a system: An AWS, a 256 Gb Random Access Memory (RAM), 32 cores, and parallel computer.

[0108] Methods described herein may represent processing that occurs within a system. The subject matter described herein can be implemented in digital electronic circuitry, or in computer software, firmware, or hardware, including the structural means disclosed in this specification and structural equivalents thereof, or in combinations of them. The subject matter described herein can be implemented as one or more computer program products, such as one or more computer programs tangibly embodied in an information carrier (e.g., in a machine readable storage device), or embodied in a propagated signal, for execution by, or to control the operation of, data processing apparatus (e.g., a programmable processor, a computer, or multiple computers). A computer program (also known as a program, software, software application, or code) can be written in any form of program-

ming language, including compiled or interpreted languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program does not necessarily correspond to a file. A program can be stored in a portion of a file that holds other programs or data, in a single file dedicated to the program in question, or in multiple coordinated files (e.g., files that store one or more modules, sub programs, or portions of code). A computer program can be deployed to be executed on one computer or on multiple computers at one site or distributed across multiple sites and interconnected by a communication network.

[0109] The processes and logic flows described in this specification, including the method steps of the subject matter described herein, can be performed by one or more programmable processors (e.g., processor **1410** in FIG. **14**) executing one or more computer programs to perform functions of the subject matter described herein by operating on input data and generating output. The processes and logic flows can also be performed by, and apparatus of the subject matter described herein can be implemented as, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application specific integrated circuit).

[0110] FIG. **14** illustrates an example computer **1405**, according to some embodiments of the present disclosure. Computer **1405** can include a processor **1410** suitable for the execution of a computer program, and can include, by way of example, both general and special purpose microprocessors, and any one or more processor of any kind of digital computer. A processor can receive instructions and data from a memory **1430** (e.g., a read only memory or a random access memory or both). Processor **1410** can execute instructions and the memory **1430** can store instructions and data. A computer can include, or be operatively coupled to receive data from or transfer data to, or both, a storage medium **1440** for storing data (e.g., magnetic, magneto optical disks, or optical disks). Information carriers suitable for embodying computer program instructions and data can include all forms of nonvolatile memory, including by way of example semiconductor memory devices, such as EPROM, EEPROM, flash memory device, or magnetic disks. The processor **1410** and the memory **1430** can be supplemented by, or incorporated in, special purpose logic circuitry.

[0111] The computer **1405** can also include an input/output **1420**, a display **1450**, and a communications interface **1460**.

[0112] While various embodiments have been described above, it should be understood that they have been presented by way of example and not limitation. It will be apparent to persons skilled in the relevant art(s) that various changes in form and detail can be made therein without departing from the spirit and scope. In fact, after reading the above description, it will be apparent to one skilled in the relevant art(s) how to implement alternative embodiments. For example, other steps may be provided, or steps may be eliminated, from the described flows, and other components may be added to, or removed from, the described systems. Accordingly, other implementations are within the scope of the following claims.

[0113] In addition, it should be understood that any FIGS. which highlight the functionality and advantages are presented for example purposes only. The disclosed methodology and system are each sufficiently flexible and configurable such that they may be utilized in ways other than that shown.

[0114] Although the term "at least one" may often be used in the specification, claims and drawings, the terms "a", "an", "the", "said", etc. also signify "at least one" or "the at least one" in the specification, claims and drawings.

[0115] Finally, it is the applicant's intent that only claims that include the express language "means for" or "step for" be interpreted under 35 U.S.C. 112(f). Claims that do not expressly include the phrase "means for" or "step for" are not to be interpreted under 35 U.S.C. 112(f).

1. A method, comprising:

analyzing duplicated watching habits of viewers in a designated market using machine learning algorithms and auto-generated synthetic campaigns to obtain campaign spot plans;

applying the campaign spot plants to single viewer data to calculate campaign spot plan Television Average Ratings Points (TARP) pattern information; and

translating the TARP pattern information into reach information determining how many people were uniquely exposed to each campaign spot.

* * * * *