Optimizing Modern Marketing Campaigns Using Convex Optimization

Samarth Negi

I. Introduction

Modern marketing campaigns are heavily biased towards influencer based marketing systems due to distrust in the traditional marketing streams [1]. However the availability of multiple streams with each their own nuances and target demographics makes choosing the appropriate combination of streams a challenging task.

Recent studies have shown that there are more than 37 million influencers only on the Instagram platform [2] and there are even other platforms such as YouTube, Facebook which operate on a similar if not higher scale.

The task of allotting budget to a marketing campaign is also complicated due to a two way effect between the stream and the brand as the stream and the brand share consequences and benefits making the decision of choosing an advertisement stream as extremely crucial and missing on required due diligence can have massive effects on the brand.

It is mathematically hard to predict how effective a stream is but studies like [3] have shown that we can use stream communication metrics such as mentions and retweets/post responses can be used to approximate how effective and likable a stream can be.

<u>The main goal for this project is to allocate a budget to specific streams so as to maximize the interaction between the audience and the brand.</u>

This report is heavily based on practical usage so it uses numerous mathematical formulations to target different aspects of the problem and provide a flexible framework for the problem statements such as :
1. Allocate a budget that maximizes views for a given budget
2. Allocate a budget that focuses on high quality streams. .

Section 2 introduces the important nomenclature and terminologies used in the domain and provides mathematical interpretation and justification.
Section 3 introduces the dataset and the modification operations carried out in order to make the dataset more robust.
Section 4 lists all the formulations and optimization problems used to solve problems.
Section 5 lists the results for the problems as mentioned in section 4.

II. Setting up the problem

1. Marketing Campaign

Modern marketing campaigns are extremely streamlined and data driven with the main constraint being the **budget**.

Marketing takes place in the form of posts on a content creator/social media personality's platform, each post having a cost associated with it. Brands then use their budget to allocate a certain number of posts per stream for instance:

2 ads in YouTube channel A and 5 through instagram posts of Influencer B.

$$Budget = B$$

2. Gross Rating Point

Gross rating points are a measure of the impact by a campaign using a specific medium or schedule. It quantifies impressions as a percentage of the target population, multiplied by frequency. This percentage may be greater, or in fact much greater, than 100[4]

Mathematically:

GRPs (%) = 100 * Impressions (#) ÷ Defined population (#)

GRPs (%) = 100 * Reach (%) × Average frequency (#)

Two examples: If an average of 12% of the people view each episode of a television program, and an ad is placed on 5 episodes, then the campaign has 12 × 5 = 60 GRPs. If 50% view three episodes, that's 150 GRPs.

Traditionally, GRP's are used to determine the effectiveness in the television media domain but for this project, GRP was decided to be used as a metric to maximize which would physically translate to the most views/impressions generated.

In this work however, reach is modeled as the number of followers a stream has (number of followers on YouTube for instance.) essentially boiling the formula for GRP to the following :

$$\textbf{Gross Rating Points for Stream (Objective)} = Number\ of\ followers * ads\ allocated\ to\ stream$$

$$\textbf{GRP (cumulative)} = \sum_{n} follower\ count * x_i$$

$$\mathbf{n} = number\ of\ streams\ available$$

$$\mathbf{x_i} = i_{th}\ stream$$

III. Dataset

The original sample dataset is acquired from kaggle[5]. It has the following columns :

1. 'follower_count' : Number of followers of a stream
2. ‘'following_count' : Number of people followed by the stream
3. 'mentions_received' : Number of times the stream is mentioned by other people on its platform
4. 'reposts_received' : Number of times a post made by stream is reposted by other people on its platform.
5. 'mentions_sent' : Number of mentions made by stream.
6. 'reposts_sent' : Number of posts reposted by stream.
7. 'platform' : 0-instagram, 1-facebook, 2-youtube
8. 'CPP' : Cost Per Post : Cost charged for a stream to make a post in their space (YouTube, Instagram, Facebook)

A few modifications were made to the original dataset based on the suggestions of Aman Beriwal from Sulovi Technologies to help in streamlining the problem.

The table below shows a general sample from the modified dataset.

	follower_count	following_count	mentions_received	reposts_received	mentions_sent	reposts_sent	posts	platform	CPP
1564	203788	1069	1043.235953	558.213656	20.643643	1.371208	22.231386	1	4075.76
2149	1097135	830	73.165677	19.898366	0.100503	0.100503	0.351840	1	21942.70
4298	1834007	773	1674.266909	388.426427	29.331852	7.308293	17.659710	0	18340.07
426	85138	67047	150.340456	68.117366	5.480349	0.100503	71.879274	2	1702.76
1687	268170	2087	48.073720	6.383899	0.356943	0.356943	0.356943	1	5363.40

The dataset consists of 5500 streams as columns however in a more practical setting, brands have a list of streams which is collected over time and a dataset similar to the one shown is available for every stream.

IV. Formulation

In this section, we work with practical requirements which might be requested from brands and provide optimization problems based around the requirements. Iteratively, forming a cumulative optimization problem which focuses on all of the requirements, however it is important to note that each requirement can be targeted independently.

A. *Optimization formulation 1 :* ***Get most out of the budget.***

This was the initial approach to solve the problem of how to get the most out of the budget in terms of GRP.

A sample of 10 streams has been taken from the full dataset and the following optimization problem was used to accomplish this task :

$$x_i \ : \ i_{th} \ stream \ \ // \ x \ : \ array \ of \ allocation$$
$$c_i \ : follower \ count \ for \ i_{th} \ stream \ // \ c \ : \ array \ of \ follower \ counts$$
$$A_i \ : cost \ per \ post \ for \ i_{th} \ stream \ // \ A \ : \ array \ of \ costs$$
$$B \ : \ budget$$

$$\textbf{minimize}: \ -c^t x$$
$$\textbf{subject to}: \ x \geqslant 0, \ A^t x \leqslant B$$

B. *Optimization formulation 2 :* ***Forcing high profile streams***

The cost of each stream is generally correlated with its reach so it is a very common request to get placements on high profile streams and as few ads as possible. These brands for more targeted advertisements in their niche. This can be achieved by limiting the number of ads and the optimization adjusts accordingly.

The following optimization problem accomplishes this task:

$$\textbf{minimize}: \ -c^t x$$
$$\textbf{subject to}: \ x \geqslant 0, \ A^t x \leqslant B, \ \sum_n x < 10$$

C. *Optimization formulation 3 :* ***Force high engagement streams.***

It is often the case that companies require streams who are well liked and talked about in the community. A good metric to judge this is the mentions_recieved column.

As this column is biased due to the discrepancy in the followers, we first divide it by the follower_count column to get a more general idea of how much a stream is mentioned regardless of its number of followers. This new column is added and named quality.

In order to get a comparable statistic, we apply the softmax function to this column which translates the total sum of this column to 1 ie. it gets scaled by its weight so that an arbitrary constraint on the quality column can be placed in order to ensure that the optimization targets high mentioned streams.

The following optimization problem accounts for the softmaxed, quality column :

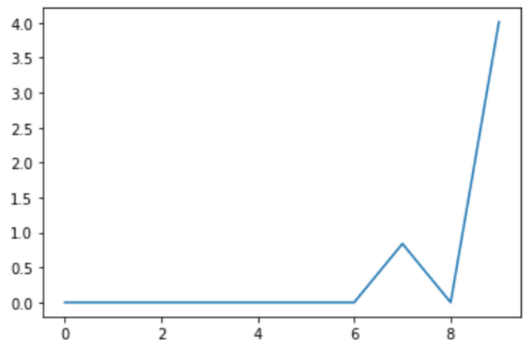
$$q_i \ : softmaxed \ quality \ for \ i_{th} \ stream \left(\sum_n q_i = 1 \right) \ // \ q \ : \ array \ of \ stream \ quality$$

$$\textbf{minimize}: \ -c^t x$$
$$\textbf{subject to}: \ x \geqslant 0, \ A^t x \leqslant B, \ q^t x \geqslant 0.5$$

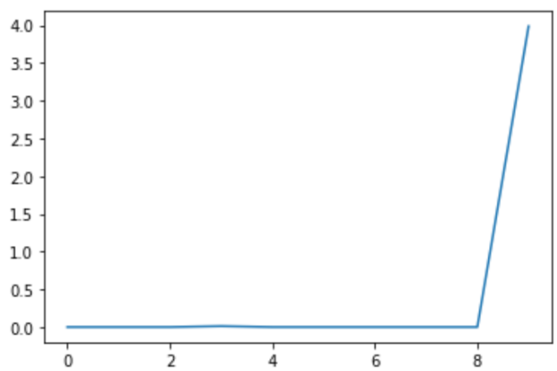
V. Results

The following 3 graphs demonstrate how using different constraints lead to different allocations for stream selection on the same set of 10 streams.

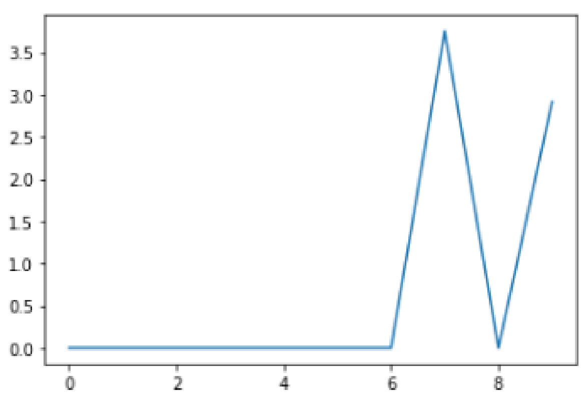
X-AXIS : stream index
Y-AXIS : budget allocated



result of formulation 1



result of formulation 2

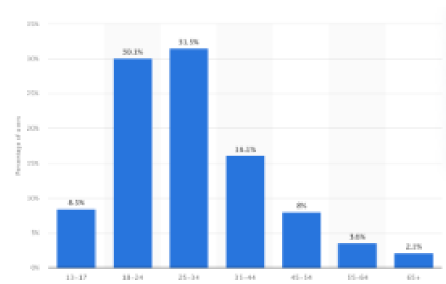


result of formulation 3

These results show how different configurations of the optimization problem are able to change the allocation of budget in order to focus on a different real word requirement.

VI. Discussion and future directions

A big part of the problem statement is based around demographics and targeted reach which is not addressed in this project due to lacking data. The actual GRP would highly depend on the demographic data as it counts for people advertised to, in the targeted demographic so the final grp we maximize isn't actually the real GRP that would be reported but a similarly correlated variable.



Graph showing distribution of demographic on Instagram

Two main issues which I look forward to addressing in the future are allocation diversity and high dimensionality.

Allocation diversity, similar to stock portfolio diversification is a common problem that exists in this project as it is highly evident that once an effective stream is found, the algorithm maximizes the budget allocated, and while this is not always the case, it does occur ie. a massive amount of money allocated to one/few streams with high effectiveness that the algorithm internally allots.

A constraint similar to covariance can be used in this case but further reading is required in order to find what feature of stream to build such a with.

VII. Code

The code can be run interactively at the url:
https://colab.research.google.com/drive/1BP76hHrBOKUsNM_k2d0AxdGtFlKGUslF?usp=sharing

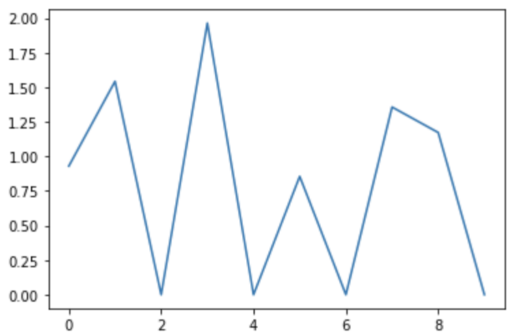
- Runtime -> Run All

Alternatively, the repository can be found here:
https://github.com/tigboatnc/Marketing-Budget-Optimization

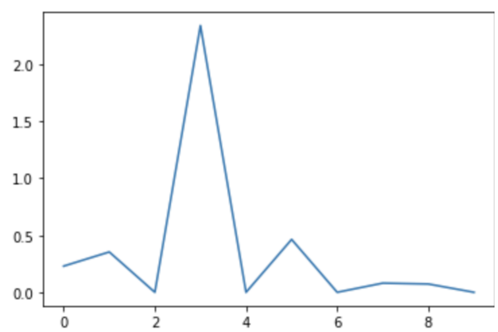
- Running instructions included in the repository.

VIII. Additional Results

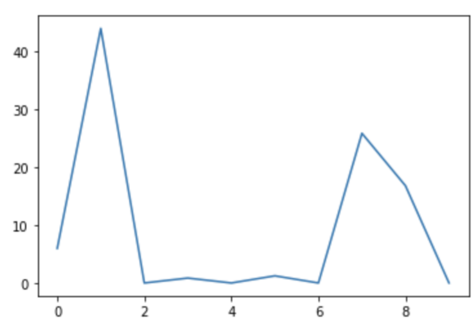
Another set of randomly picked 10 streams were ran through the algorithm to provide the following 3 outputs based on the 3 formulations :



result of formulation 1



result of formulation 2



result of formulation 3

These 3 outputs show a lot more dynamic range in terms of different allocations for different brand requirements and show how this system can effectively undertake the task of allocation.

Acknowledgment

I thank Aman Beriwal of Sulovi Technologies for his domain expertise and help in dataset exploration. I would also like to thank professor Richard La for the informative and interesting lectures.

References

[1] https://www.nielsen.com/us/en/insights/article/2012/consumer-trust-in-online-social-and-mobile-advertising-grows/#:~:text=According%20to%20Nielsen's%20latest%20Global,an%20increase%20of%2018%20percent

[2] https://thesmallbusinessblog.net/how-many-influencers-are-there-on-instagram/

[3] Choosing the right social media influencer [https://www.diva-portal.org/smash/get/diva2:1107811/FULLTEXT01.pdf]

[4] https://en.wikipedia.org/wiki/Gross_rating_point#cite_note-1

[5] https://www.kaggle.com/competitions/predict-who-is-more-influential-in-a-social-network/data?select=train.csv