Time for marketing to embrace reinforcement learning

Received (in revised form): 24th May, 2022

Laura Murphy

Chief Executive Officer, Amplify Analytix, The Netherlands

Laura Murphy is the co-founder and Chief Executive Officer of Amplify Analytix. A former head of business transformation in international markets, Laura specialises in helping organisations embrace the benefits of data science to improve commercial performance.

Amplify Analytix BV, De Cuserstraat 93, 1081CN Amsterdam, The Netherlands
E-mail: laura.murphy@amplifyanalytix.com

Fernando Perales

Head of the Research Lab and Strategic Partnerships, JOT Internet Media, Spain

Fernando Perales is Head of the Research Lab and Strategic Partnerships at JOT Internet Media, where he is responsible for the coordination of innovation projects, from developing in-house technology to participating in pan-European collaborations to test cutting-edge Big Data applications.

JOT Internet Media, General Ramírez de Madrid 8, 28020 Madrid, Spain
E-mail: fernando.perales@jot-im.com

Anand Gopal

Co-Founder and Chief Operating Officer, Voiro, India

Anand Gopal is the Co-Founder and Chief Operating Officer of Voiro. An ex-operations research practitioner and Georgia Tech graduate, Anand is responsible for driving company growth, strategy and operations at Voiro.

Voiro, 43 Residency Road, Bengaluru 560025, India
E-mail: anand@voiro.com

Yordanka Gyurdieva

Head of Product Strategy, Amplify Analytix, Bulgaria

Yordanka Gyurdieva is the Head of Product Strategy at Amplify Analytix and the product owner of SOLD! — Amplify's award-winning contextual multi-armed bandit solution. She helps organisations translate business problems into hypotheses that can be treated with data science solutions, such as reinforcement learning.

Amplify Analytix, Campus X, building 3, Aleksandar Malinov Boulevard 31, floor 4, 1729 g.k. Mladost 1A, Sofia
E-mail: yordanka.gyurdieva@amplifyanalytix.com

Victor Gueorguiev

Senior data scientist, Bulgaria

Victor Gueorguiev is a senior data scientist and avid user of reinforcement learning. He uses his deep expertise in advanced analytics to develop, test and deploy data science solutions to solve customer needs and pains.

Bul. Simeonovsko Shose 110, zh.k. Gradina 14, Floor 4, Apartment 14, Sofia, Bulgaria
E-mail: victor.gueorg@gmail.com

Pratyush Shandilya
Data Scientist, Amplify Analytix, India

Pratyush Shandilya is a Data Scientist at Amplify Analytix. He has developed bespoke end-to-end data science solutions for clients in diverse industries. His expertise is in Deep learning, Reinforcement learning and Natural Language Processing. Pratyush holds a Master of Science from the Carnegie Mellon University.

No. 3 Venus Hebron, 4th G Main Road, Kalyan Nagar, Bengaluru 560043, India
Email: pratyush.shandilya@amplifyanalytix.com

Abstract Since COVID-19 upended the world, marketers can no longer rely on historical data to inform their decisions. Channel splits have changed and online conversations have exploded. Marketing budgets have decreased as a percentage of revenue, meaning marketing funds must be used more effectively and efficiently than ever. Fortunately, the relatively new application of reinforcement learning — a data science approach — in marketing offers additional opportunities to gain competitive advantage using artificial intelligence. Unlike other types of machine learning, reinforcement learning uses algorithms that do not typically rely only on historical data sets, to learn to make predictions. Rather, these algorithms learn as humans often do, through trial and error, adjusting their 'behaviour' based on the outcomes of their actions. While the algorithms and computations behind reinforcement learning can be complex and sophisticated, its ability to deal with real-time decision making makes it an attractive option for marketers. This paper shows that with the right 'business translator' — that is, a person or team operating as the 'glue' between data science and business performance — sophisticated data science becomes accessible to commercial teams looking to drive performance improvements.

KEYWORDS: reinforcement learning, data science, marketing analytics, change management, artificial intelligence, digital marketing

INTRODUCTION

The last two years have turned the world upside down and marketing has not been spared. This is most obviously seen in consumer markets, where the global COVID-19 pandemic acted as an accelerator of consumer engagement with digital media. Customer conversations have moved towards the online world, and e-commerce has grown between two and five times faster than before the pandemic.[1] Channel splits have changed dramatically, and brand loyalty has been dented. Unsurprisingly, given this trend, digital channels strengthened their position in 2021 as the most invested-in advertising medium worldwide. Nearly 59 per cent of all global advertising investments were devoted to digital promotion last year, while television received less than 25 per cent of total advertising expenditure.[2]

Add to this that marketing budgets are at their lowest as a percentage of revenue since Gartner began its annual CMO survey,[3] and the result is the perfect storm. This challenge means that using data science for 'better' decision making becomes a necessity, rather than just an opportunity. As this paper will discuss, existing data science techniques from the field of reinforcement learning, such as contextual multi-armed bandits (CMABs) and K-armed bandits, applied in new ways to marketing challenges, have proven effective at tackling common problems in digital marketing, such as investing the

right amount in paid search to show the right advertisements to the right audience at the right time, or allocating advertising inventories based on performance in digital advertising.

CONTEXTUAL MULTI-ARMED BANDITS APPLIED TO GOOGLE PAID SEARCH

One organisation that understands the challenges of paid search very well is JOT Internet Media, a Spanish company that helps its partners to reach new audiences relevant to their business, increasing and monetising their web traffic.

As an innovative organisation, JOT was looking for data-driven ways to improve the service it provides to its clients, and to save data-crunching time for its digital marketing experts, freeing up more of their time for the creative part of the job.

JOT has a long history of experimenting with data science to help achieve its goals, and in 2021 it teamed up with Amplify Analytix, a business data science company. The main goal of this partnership was to improve JOT's detection of non-trivial temporal search patterns, with a view to identifying the most relevant keywords (interests) of the day, month or season, as well as by location, to help JOT's digital marketing accounts make more effective use of their marketing budget.

The partnership was enabled by the European Data Incubator, an Innovation Action project co-funded by the European Union to facilitate sustainable business incubation around Big Data. Using JOT data to conduct the trial, Amplify Analytix explored whether applying the bandit technique could help to improve both clicks and click-through rate (CTR), thereby better serving JOT's clients — and succeeded in developing an award-winning implementation of an algorithm using the CMAB technique from the field of reinforcement learning, to optimise Google Ads spend.

To get the maximum benefit from the application of data science to business problems, it is crucial to spend time upfront to understand why the organisation wants to improve its current performance and how its teams will use the improved solution as part of their daily operational processes. This ensures that solutions are focused not just on maximum performance but also on maximum usability. Amplify Analytix business analysts invested time in understanding the various steps taken by JOT's digital marketing experts to determine the best course of action for spending Google Ads budget. Based on business context, the market trend of drastically changing consumer behaviour, data size and data patterns, CMAB was deemed the best fit for purpose.

INVESTING IN REINFORCEMENT LEARNING PAYS OFF

With the solution implemented in a live environment, the CMAB application to Google Search bid optimisation resulted in 15 per cent higher CTR, 7 per cent estimated increase in return on investment (ROI) and 38 per cent of digital expert time saved. Further, the time taken to onboard new digital specialists joining the team was reduced by an estimated 12 months.

In the words of Fernando Perales, Head of the Research Lab and Strategic Partnerships at JOT:

> 'After analysing the results, it is clear that the model helps to improve marketing campaign performance. In particular, its ability to digitise and replicate the optimisation processes carried out by accounts, resulting in a significant reduction in time consumed by analyses and delivering a data-driven decision support system. This was made possible because the team sought to understand end-user roles and pain points

and built a model that added value to daily work. We are very interested in continuing collaboration to develop these predictions one or even two steps further. The gains are definitely worth it.'

REINFORCEMENT LEARNING IN THE ALLOCATION OF ADVERTISEMENT INVENTORY

Much like JOT, Voiro, a data technology company bringing automation and intelligence to leading media organisations across the world, has also embraced the benefits that reinforcement learning can bring to digital advertising. Voiro's advertising operations suite orchestrates orders, placements, billing and integrations for leading advertisement publishers, over-the-top (OTT) businesses and e-commerce companies. Amplify Analytix and Voiro teamed up to enable Voiro to optimise the allocation of available advertising inventories for its clients by using the reinforcement learning technique — 'K-armed bandit'.

To deliver ads to viewers, ad requests are routed to supply-side partners (or also, Indirect partners). Indirect partners are the bridge between the Publisher's platform and the Ad exchanges. A Publisher is most often connected with multiple indirect partners. Based on priorities set on the Publisher's platform, ad requests get allocated to different indirect partners. Prior to the data science implementation, the percentage of advertising requests sent to a particular advertisement delivery path was assigned manually among indirect partners to advertisement publishers, a process that could not capture variations in advertisement performance. Some indirect partners provide better performing advertisement placements and help generate more revenue for advertisers than others. Without intelligent allocation of advertising inventory among a range of indirect partners, an optimal price offer for advertising could not be achieved for the advertisement publishers, thus losing them revenue without them being able to know how and why.

Thanks to Voiro's innovative approach, Amplify Analytix was able to create and deploy cutting-edge reinforcement learning models to enable the intelligent allocation of advertising inventory among its indirect partners based on specific context, such as their historic performance, inventory type, time of year and price.

K-ARMED BANDITS AND ADVERTISEMENT ALLOCATION

A K-armed bandit from the field of reinforcement learning was used to assign priority to indirect partners. Once priority levels have been identified, advertisements are delivered proportionally according to that priority. The performance of indirect partners is assessed daily by the model and priority is adjusted accordingly.

Voiro chose to work with reinforcement learning because it is an online training approach that removes model retraining overheads and becomes more flexible and scalable, thereby minimising the company's dependency on Amplify Analytix or any other data science partner in the long run, without the need to build up a new competency in-house. Amplify deployed these models on cloud technology to enable API calls from Voiro's 'Panthera' product to obtain recommendations on inventory split. The model is developed and deployed on Amazon SageMaker, which makes it instantaneous and highly scalable. This method helped reshape the allocation of advertising inventory from manual to automatic.

The data solution resulted in a 5 per cent monthly revenue increase for one of its most important clients — a well-known OTT platform that streams television shows and movies.

In the words of Anand Gopal of Voiro:

> 'This AI-driven recommendation system for publisher advertising inventory allocation has a dual benefit: in addition to driving revenue impact, it helps publishers reduce the time, effort and operational costs involved in the management of programmatic advertisements.
>
> As the model factors in holidays and special events, it can robustly handle any sudden aberrations in available advertising inventory. Reinforcement learning-based inventory allocation allows for plug-and-play optimisation across programmatic partners and reduces human dependency, to a great extent.'

TODDLERS, BANDITS AND MARKETING . . . REALLY?

The principles of reinforcement learning, and bandits in particular, are guided in no small part by learnings from the field of psychology. The way that these models work is somewhat similar to how a child learns from its actions. When a toddler is making its way around the house and discovering fun things to do, it seeks to repeat those activities to gain the 'reward' of enjoyment. Of course, some actions have unpleasant outcomes, such as accidentally burning their hand on the heater or bumping their head on a table. Over time, the child learns these are activities that it does not enjoy, and seeks to avoid the 'punishment' that such actions entail. In these circumstances, one can liken the concept of rewards to feedback for a given action; so, in circumstances that lead to a negative outcome, one obtains a negative return, while desired outcomes result in a positive reward. The model learns, as the toddler does, from the feedback that it continuously gets from its environment.

In more technical terms, the bandit seeks to maximise returns by selecting one of K-options in each round. Each option has unknown reward distribution, and the algorithm tries to learn the reward distribution of each option to then use this information and optimise its actions for the purpose of maximising returns. In other words, there is a context in which the model operates, actions (options) from which it can choose, and rewards after every choice that will define how well the action was predicted. The reward feedback 'teaches' the algorithm to learn and improve its outcome with every decision taken.

Applying this logic to digital advertising, as seen in these two examples, is a fitting application of the bandit as the context and actions are not only clearly defined but there are also frequent rewards from which the model can learn. The context for the algorithm is a complex combination of parameters, such as target audience, search category, platform, device, browser, advertising space, match type, price, date and time-related features, user demographics if available, advertisement type, advertisement content, campaign properties, etc.

The reward component requires careful consideration and must have a clear link to the business goal the organisation seeks to achieve. It could be a single objective, such as the number of clicks the advertisement receives, especially if the revenue of the demand platform depends on the number of clicks. One could also consider, for example, the total revenue obtained from all advertisement placements. The high frequency of the rewards comes from the hundreds and thousands of auctions available every day, and this is why, through frequent learning from every reward, the bandit can further improve the outcome in a short period of time.

From a process, speed and volume perspective, the algorithm mimics human learning by utilising over 1,000 daily iterations to produce the best bid options per advertising group and campaign. This process is equivalent to the collective learning of 100 digital marketing or advertising

managers working as a team, but it happens on a daily basis in a matter of seconds. The algorithm considers all past events, trends and interdependencies in the data without requiring marketer time for analysis.

REINFORCEMENT LEARNING COMPARED WITH OTHER FIELDS OF MACHINE LEARNING

In traditional machine-learning paradigms, algorithms are trained to learn patterns on a batch of historical training data. These batches come with labels or numerically valued outcomes. Once the desired outcome has been modelled reasonably well, the models are deployed into business processes. With the passage of time, new business patterns can evolve and a previously trained model may no longer be relevant. Data science practitioners refer to this as model drift. Reinforcement learning as a machine-learning paradigm offers an alternative. Reinforcement learning involves training an artificial intelligence (AI) agent that learns to take optimal decisions based on its interactions with its environment and the reward signals it obtains. At the beginning, this training does use historical data. But the design of the agent allows it to be explorative even after it has been deployed. Every now and then, the agent takes an action that cannot be considered 'optimal' based on its prior experience. This allows the agent to explore alternative action choices, even as the business patterns change. This continues, keeping the AI agent relevant with time.

UNDERSTANDING BANDITS AND CONTEXTUAL BANDITS

To understand multi-armed bandits, it is helpful to think of a practical scenario. Imagine that a casino has 20 slot machines and the cost of interacting with (spinning) any of them is US$1. Suppose then that one of these 20 machines is malfunctioning, and the probability of hitting the jackpot with it is disproportionately large compared with the other machines.

The gambler in this hypothetical scenario knows that one of the slot machines is malfunctioning, but has a budget of only US$50. He must therefore choose which machines to play based not only on payoff but also on how long he is able to interact with them. For the gambler then, the problem is thus which machines to try, and in what order, to find the machine with the highest payoff, within a limited budget.

This represents a typical statement for the 20-armed bandit problem (as there are 20 different machines). A solution to this problem involves an algorithm for deciding the sequence of slot machines to interact with.

Suppose now that other people also know that one of the machines yields bigger payoffs. In this case, imagine that the gambler consults these people every time he is about to spin a machine. As he is considering other information when choosing a machine — the opinions, beliefs and guesses of the others — it becomes a contextual 20-armed bandit problem. The problem formulation is the same as above, except that now more information — or context — is available to the decision-maker at any point in time. Again here, various algorithms exist for utilising this additional information when selecting the next slot machine to spin.

DATA SCIENCE DEPENDS ON EXPERIMENTATION

As this paper has demonstrated, reinforcement learning can bring great benefits to improving performance, helping marketing teams to identify the right actions to take across a value chain as events unfold. Thinking beyond optimising advertising choices, when integrated within personalisation and recommendation systems, reinforcement learning can help organisations personalise messaging, promotions and offers in

near real time, improving relevance for customers, thereby improving the customer experience, as well as the likelihood of sale. Getting started, however, can seem daunting, not least because expectations from the boardroom tend to be sky high, and so many companies are now promising to unlock the power of data using AI.

Here, it is worth bearing in mind that the ROI is likely to be higher with contextual bandits than with traditional machine-learning models. There are multiple reasons for this. First, reinforcement learning models are typically trained offline using historical data. This is expected to make them at least as good as a human expert at making decisions. Thereafter, the model is deployed online, in the field, where it continues to improve on its decision-making as it interacts with its environment. Its performance, as a result, stays relevant even as environmental patterns change. More relevant decisions then translate into better performance and improved ROI, as seen in the previously discussed JOT case. This stands in contrast to traditional machine-learning models, which require periodic retraining to maintain performance. Secondly, contextual bandits are 'lightweight' reinforcement learning applications that can be trained with a powerful but otherwise standard personal computer. Training standard machine-learning models on the other hand, can be CPU/GPU heavy. In conclusion, it can be said that that both cost of deployment and performance, are likely to be higher with reinforcement learning models.

Laura Murphy, CEO of Amplify Analytix, says:

> 'Over the years, we keep improving the recipe for success: a business problem or opportunity that is clearly articulated and understood by all stakeholders, a "translation" of the business problem into terms that can be handled by data science, a top-notch data science model built by the best data scientists, and a deep understanding of how the output or insights will be used by the business users in question are all crucial ingredients. Our best work happens when we create a multi-disciplinary, cross-organisation team, as we did with JOT and Voiro, made up of those who need and those who build the insights, to enable us to change established ways of working and deliver improved performance. We encourage every team to get on the journey, experiment, learn fast and embrace occasional failure as part of the learning. We are convinced that the possible benefits, for now and in the future, far outweigh any teething problems on the way to unlocking the value of your data.'

CHOOSING THE RIGHT DATA SCIENCE PARTNER

So, how does one choose the right data science partner to get started? Sasha Dzhuras-Dotta, Growth Strategist and former Marketing Director at Wella and iRobot UK, provides the following viewpoint:

> 'If you are at the beginning of your AI journey, you need a reliable partner who will seek to understand your business challenge before bringing a data-science based solution to the table. Some considerations:
> - Curiosity, skills and persistence to ask the right questions in order to articulate clearly the business problem or opportunity you are trying to tackle and why.
> - Ability to keep it simple — a data science partner should be able to make themselves understood by people who have no experience in data science, using common sense and business terms, and a willingness to offer the simplest solution possible to "get the job done".
> - Patience and persistence working with your data — no matter how challenging or "unclean" it may be, and a can-do attitude to discover what *is* possible, rather than what is not.

- Broad coverage of different data science fields, and top-notch data scientists who can offer the most appropriate solution.
- Willingness to explain the process (not a black box) and invest in development of your team, preparing them for the next stage of your AI journey.
- Courage to share unexpected or difficult discoveries.'

BRINGING IT ALL TOGETHER

Potential customers are spending more of their time shopping online than ever before, and customer journeys are increasingly complex, generating mountains of data for analysis; yet marketing budget allocation is experiencing a low as companies around the world try to recover from the impact of drastic COVID-19 regulations.

This gives marketers and data scientists a unique challenge — extracting meaningful information from vast amounts of data as efficiently and cost-effectively as possible in order to inform decision-making and plan a way forward for the company. The solution lies in the application of intelligent data-driven methods such as the K-armed bandit technique to inform decision-makers, improve customer service, and save data-crunching time for digital marketing experts.

Getting the best out of complex data requires advanced science, and an enthusiastic, clever, transparent team to develop the best algorithms to handle different problems. Getting the expertise and the experts right gives every company the best chance to handle whatever curveball is coming next.

References

1. McKinsey (2021) 'How e-commerce share of retail soared across the globe: A look at eight countries', available at: https://www.mckinsey.com/featured-insights/coronavirus-leading-through-the-crisis/charting-the-path-to-the-next-normal/how-e-commerce-share-of-retail-soared-across-the-globe-a-look-at-eight-countries (accessed 23rd June, 2022).
2. PricewaterhouseCoopers (2021) 'The global consumer: Changed for good', available at: https://www.pwc.com/gx/en/consumer-markets/consumer-insights-survey/2021/gcis-june-2021.pdf (accessed 23rd June, 2022).
3. Gartner (2021) 'The State of Marketing Budgets 2021', available at: https://emtemp.gcom.cloud/ngw/globalassets/en/marketing/documents/gartners_annual_cmo_spend_survey_2021_ebook.pdf (accessed 23rd June, 2022).