Journal of Engineering and Applied Sciences 14 (Special Issue 9): 10560-10568, 2019
ISSN: 1816-949X
© Medwell Journals, 2019

CTR Prediction with Deep Neural Networks

Mustafa Radif and Atheer Alrammahi
University of Al-Qadisiyah, School of Computer Science and IT, Al-Diwaniyah, Iraq

Abstract: Ad effectiveness on online platforms including Facebook and Google is a challenge for businesses. Since, ad networks use algorithms, ad effectiveness as measured by CTRs is not well understood by marketing and sales executives. CTR prediction with deep neural networks can improve ad CTRs. The AI solution for ad CTRs is useful across industry sectors. In the solution, RNN Models learn representations of sequences by maintaining internal states which are updated sequentially and used as proxy for target prediction. Evidence from research shows that deep neural networks could help businesses enhance CTRs.

Key words: Ads, artificial intelligence, CTRs, recurrent neural networks, industry, networks

INTRODUCTION

Knowing whether a customer will click an ad is one of the dilemmas for ad spending. From an information systems perspective, this is an optimisation problem which involves learning how to make predictions with algorithms. The objective of the machine learning and AI approach is to engage in deep learning to solve the ad optimisation problem. This presentation provides evidence for enhanced predictions on ad networks across industry sectors.

Problem description: Ads have become cost-intensive for businesses across industry sectors. Since, businesses operate in competitive environments, it is critical for the executive management to focus on good decision-making. Regardless of the business sector, it is critical to enhance the decision-making process as this could mean a significant reduction in ad spending. Machine learning and AI could enhance the ad effectiveness decision-making. Smith (2018) argues that several factors should be considered while analysing whether Click-Through Rate (CTR) is good. Since, achieving higher CTRs is challenging to accomplish (Cheng *et al*., 2012), machine learning and AI could be deployed for improving CTR. Markus (2017) explains how Google and Facebook use CTRs to determine the quality of ads. Each ad is given a quality score which is based on the CTR. Ads use bidding systems which consider the quality score for ad placement. A high-quality score could outrank the competition and do away the need for outbidding the competition. This means the ad shows higher up in the user's feed at a lower cost-per-click. Thus, the challenge for advertising is to increase the CTRs. Strong CTRs could be achieved by good targeting and creativity. Several factors come into play to make a good CTR. Ad positioning, imagery quality and keywords are the main factors that impact CTR. Though these guidelines are useful there is no clear solution for determining what is a good CTR.

Literature review: This section has a review of relevant literature on ad effectiveness, the most significant problem for online ad spending. The discussion shows the relevance of the metric CTR and how AI could solve this problem for businesses based on evidence from academia and industry experts.

MATERIALS AND METHODS

Ad effectiveness: Columbus (2018) argues that challenges for marketing include optimising marketing campaigns, improving the precision and profitability of pricing and understanding, anticipating and acting on the problems of sales prospects better than the competition among others (Fig. 1). Current practices include reliance on analytics and key performance indicators for monitoring progress towards revenues and customer growth goals.

Machine learning and artificial intelligence and could help improve existing processes. Marketers continue to devise and launch new ad campaigns. Machine learning could contribute to every aspect of attracting, selling and customer service by deploying machine learning to predict outcomes. Certain areas of marketing have higher levels of complexity and provide greater financial benefits. Traditionally, marketers have placed emphasis on must do areas that offer high benefit and low complexity. Chatbots, facial recognition, product and service recommendations, revenue analysis and virtual assistants are some applications benefitting from machine learning and AI.

An area of significance is ad effectiveness which requires high levels of design accuracy, context and

Corresponding Author: Mustafa Radif, University of Al-Qadisiyah, School of Computer Science and IT, Al-Diwaniyah, Iraq

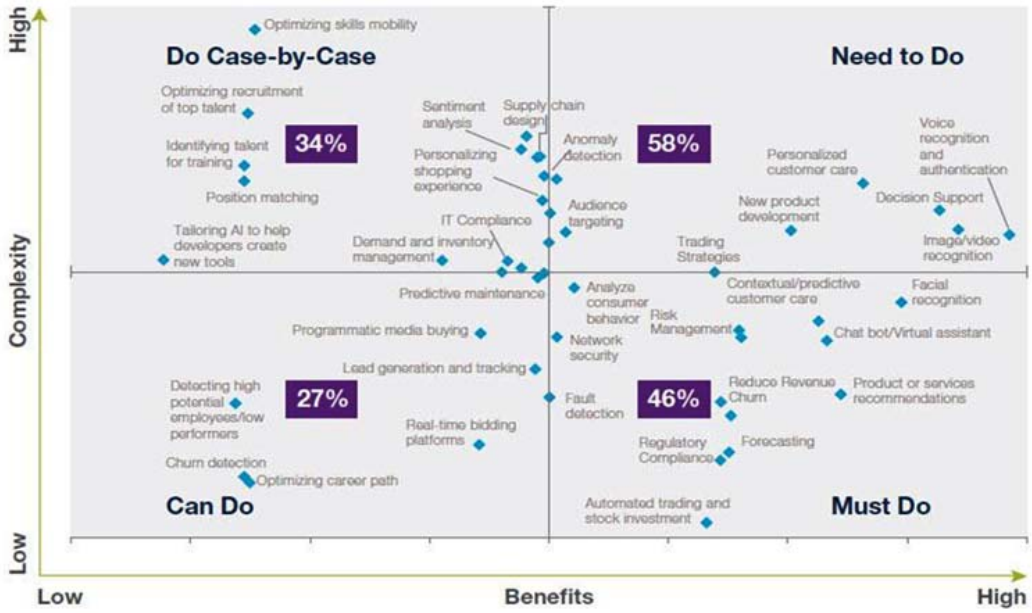

Fig. 1: Marketing benefits and complexity (Columbus, 2018)

precision. Solving this problem could contribute to increased sales. As suggested by Linoff and Berry (2011), marketing and sales executives need ways to increase ad effectiveness. Traditional approaches for ad optimisation are expensive and time-consuming. Khoshgoftaar *et al*. (2010) provide evidence for machine learning to deal with such challenges.

CTR: Cost Per Click (CPC) is the standard model for internet advertising. In the CPC Model, CTR is the index for measuring the effect of ad placement. The metric helps marketers gauge the results and the effectiveness of ad campaigns. The metric is a measure of the percentage of viewers who click on the ad. It is a ratio of the number of the number of viewers who click on the ad and the total number of impressions. For marketers who implement paid search ads, this is a critical metric. A high CTR means the targeting is effective. Jiang *et al*. (2018) explain a high CTR means that the marketing campaign is contributing to more leads and sales. On the other hand, a low CTR means the ads are not appropriate for the audience. Jiang (2016) argues that CTR prediction with deep learning could become extremely valuable for contextual ads.

CTR prediction: CTR prediction is a critical task for ad spending. This technique is useful because higher CTR means better engagement with the audience and lower ad spending. Chen *et al*. (2017) recommend filtering strategies with neural networks. Recurrent Neural Networks (RNNs) are capable of learning embedded high-dimensional representation of sequences. Zhai *et al*. (2016) argue that the technique incorporates models for solving sequential problems. The models learn representations of sequences by maintaining internal states that are updated sequentially and used as proxy for target prediction. Since, the method allows for the stacking multiple layers, it is possible for building deeper representations that result in significant improvements. Chen *et al*. (2016) and Greff *et al*. (2017) argue that a notable architecture of RNNs is the Long-Short Term Memory (LSTM) as shown in Fig. 2.

Machine learning and AI: Machine learning is the automated detection of meaningful data. Deep learning is the driving force contributing to the development of AI. Covington *et al*. (2016) argue that layered artificial neural networks or deep neural networks are contributing to this phenomenon. Hall (2018) argues that four computational elements make up a deep neural network. The first element is involved in performing data pre-processing augmentation tasks. The second element is involved in training the deep learning model. The third element involves storing and adapting the trained model for deployment. The final element involves engineering a scale-out system for deploying the model. Experts argue that training is the most challenging element that often encounters peak computing demands. While training a deep neural network, data is passed through the network in a forward pass and network weights are updated in a backward pass. With millions of parameters, bigger models can result in better performance. In a phenomenon known as epoch, several passes through the data are required for training a single model. A model may not

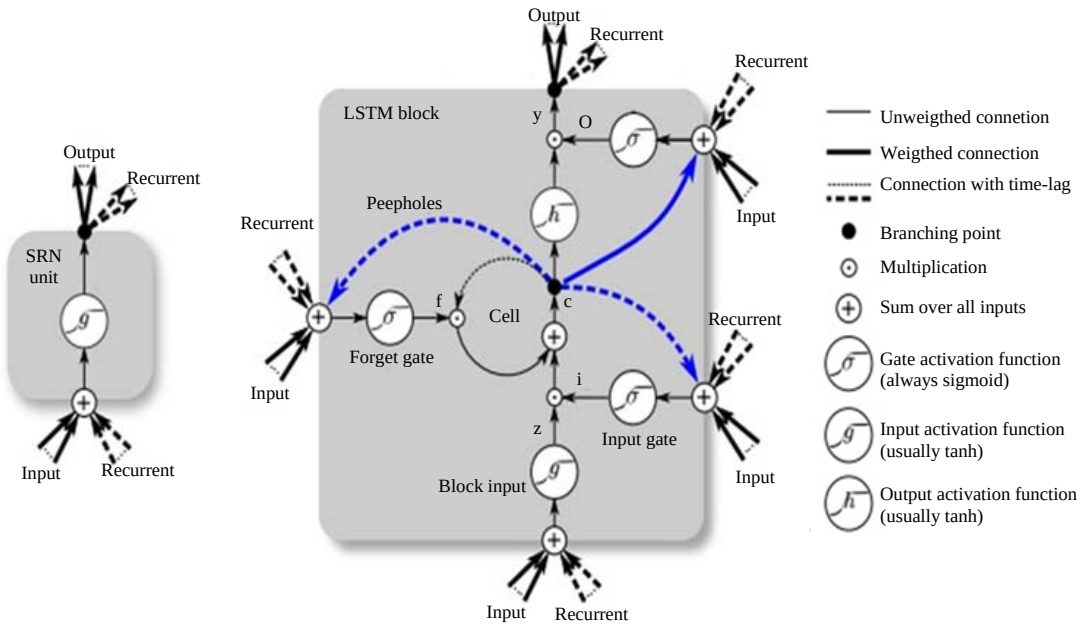

Fig. 2: RNN and LSTM Model (Chen *et al*.,)

converge to a solution or the solution may not be optimal, however in the case of a single model. In the industry, experts conduct experiments to find the ideal network architecture and search for the optimal solution for this architecture. This approach is also known as hyper-parameter. This process is deployed for implementing new ideas or training newly available data. Deep leaning could help solve the ad effectiveness problem by predicting CTRs.

The role of machine learning is to analyse data from the past into accurate predictions for predicting CTRs (Yan *et al*., 2014). Machine Learning could enhance the marketing decision-making process while also improving the customer experience. Herbrich (2017) provides evidence from industry leaders including Amazon as a testimony to the capabilities of machine learning but the majority of businesses do not enjoy the luxury of machine learning. Thus, there is a significant demand for machine learning with tools dedicated to marketing practices.

User ad experience: A useful solution could help marketers and organisations navigate the complexity of ad effectiveness with machine learning and AI. Good CTR also means businesses can get a clearer understanding of audience ad experience. In real-world practices, sales executives are unable to fully comprehend ad effectiveness. Machine learning and AI help executives analyse data on ad CTRs for targeted marketing. This is one of the unique characteristics of the solution which involves optimisation of ads based on predictive insights. Based on the patterns or insights, executives can take high-quality decisions.

In the past, marketers emphasised on establishing, sustaining and expanding relationships with consumers. Today, the customer is in the driver's seat. The challenge for businesses is to understand customers better for high-quality ads. Since, consumers spend more time researching, discovering, browsing and purchasing products, the implication for marketers is to enhance the engagement by delivering more meaningful ad experiences. Data on past behaviour is highly valuable for machine learning and the insights could help deliver suggestions relevant to their preferences. Machine learning could also open up new possibilities which include presenting further information such as the odds of a user clicking on an ad.

Machine learning could work for many more businesses beyond the companies like Amazon, Google and Facebook that make extensive use machine learning and AI for ad placements. Boothe (2018) predicts these capabilities could change ad optimisation practices for the vast majority of firms. Firms could eliminate the uncertainty associated with targeting consumers.

Learning paradigms: Several learning paradigms contribute to the machine learning world view. Supervised versus unsupervised learning, active versus passive learning, helpfulness of the teacher and online versus batch learning are the four learning paradigms (Shalev-Shwartz and Ben-David, 2014).

Learning tasks could be classified according to the nature of interaction. The main distinction is the

difference between supervised and unsupervised learning. The process involves using experience for gaining expertise. The experience or training contains important information that is missing in the unseen examples where the learned expertise is applied. In this type of training, the acquired expertise helps predict the missing part of the test data. The environment has the role of a teacher responsible for supervising the learner by including the extra information. In the case of unsupervised learning, there is no distinction between training and test data. Learning involves processing input data with the objective of summarising or compressing the data. A typical task is the clustering of data into subsets of similar objects. In between these two paradigms, there is an intermediate learning setting. The training examples contain more information that the test examples and the learner predicts more information for the test examples. Also, known as reinforcement learning, this learning framework could be deployed for predicting the outcomes of events based on previous actions.

Learning paradigms could also be classified by the role played by the learner. Active learners interact with the environment during training by performing experiments or posing queries. On the other hand, a passive learner observes the environment without trying to influence or direct the performance.

Another learning paradigm is based on the role of a helpful teacher whose role is to feed the learner with useful information for achieving the learning goal. In this paradigm, learning scenarios are modelled based on the belief that the training data is caused by some random process. This method is part of the statistical learning approach. Learning also occurs when inputs are generated by an adversarial teacher. This approach also helps model worst-case scenarios when the parameters of milder scenarios can easily be assumed. The world view is that is a subject can learn against an adversarial teacher, success is guaranteed while interacting with other circumstances or teachers.

The final learning paradigm is that a learner responds to experiences. The approach involves engaging the expertise after enjoying opportunities to process large amounts of data. This is feasible in an online environment as a result of the availability of large data sets that can be analysed in a variety of ways before generating intuitive conclusions. In the process of becoming an expert, the learner has an opportunity to make mistakes during the training process.

Statistical learning: Statistical pattern recognition, nonparametric classification and estimation and supervised learning are statistical learning techniques. For solving the ad CTR problem, it is useful to examine methods for learning. The problem is to observe an object and classify it into one of two classes which are say 0 and 1. For making this decision, measurements of various properties of the object are necessary. These measurements could be from sensors that capture physical variables or features of interest. The set of features are put together in the form of a feature vector, say there are d features, the value of the features are given by x1, x2 , …, xd. Thus, the feature vector $\bar{x} = (x1, x2, ..., xd)$ is considered as a point or a vector in d-dimensional space R^d which is known as the feature space. For a feature vector, the problem is to decide either 0 or 1. A decision rule could be considered as a mapping $c:R^d \rightarrow \{0, 1\}$ where $c(\bar{x})$ indicates the decision when feature vector is $\bar{x}$ observed. A decision rule could be considered a partition of R^d into two sets Ω0 and Ω1. These correspond to the feature vectors that get classified as 0 and 1. In reality, it is not necessary to classify the set of all possible subsets of R^d as potential decision rules. The emphasis is on decision rules that can be measured. Though there are many possible decision rules, it is useful to select one that performs well. A correct decision means the decision $c(\bar{x})$ is equal to y. The criterion for success is the correct decision. The decision rule should maximise this probability or minimise the probability of error. Thus, for a decision rule c, the probability of error is given by the following expression:

$$R(c) = E\left[c((\bar{x})-y)^2\right] = P(0)P(\Omega 1|0)+P(1)P(\Omega 0|1)$$

Ideally, the underlying probabilistic structure is known and the solution to the classification problem is a result of statistics. Based on the assumption that the underlying probabilities P(0), P(1) and $p(\bar{x}|0)$ and $p(\bar{x}|1)$ are known, the posterior probabilities are computed by using Baye's theorem:

$$P(\bar{x}) = P(0)P(\bar{x}|0)+P(1)P(\bar{x}|1)$$

For this expression, the probabilities $P(0|\bar{x})$ and $P(1|\bar{x})$ are derived as follows:

$$P(0|\bar{x}) = \frac{P(0)P(\bar{x}|0)}{P(\bar{x})}$$

$$P(1|\bar{x}) = \frac{P(1)P(\bar{x}|1)}{P(\bar{x})}$$

With the posterior probabilites, the ideal decision involves choosing the class that has the larger conditional probability. This means choosing 0 if

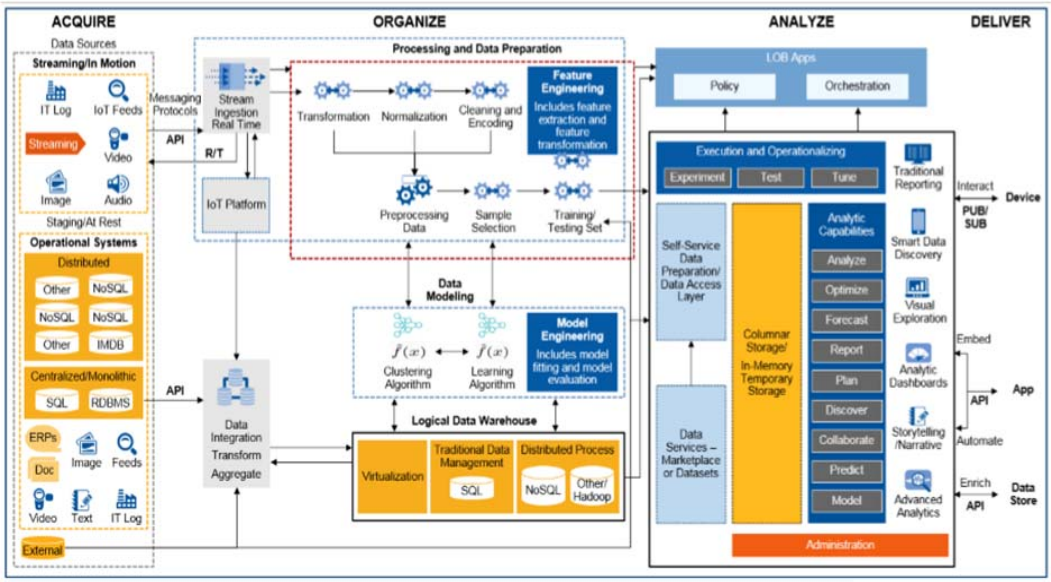

Fig. 3: End-to-end architecture (Sapp, 2017)

$P(0|\bar{x}) > P(1|\bar{x})$ or 1 if $P(1|\bar{x}) > P(0|\bar{x})$. If both are equal, the choice does not matter. Since, Baye's decision rule is optimal because there is no other decision rule with a smaller probability of error (Barber, 2012). The error rate R^* is expressed by the following expression:

$$R^* = \int_{R^d} \min\{P(0|\bar{x}), P(1|\bar{x})\} P(0|\bar{x}) P(\bar{x}) d\bar{x})$$

$$0 \le R^* \le \frac{1}{2}$$

Knowledge of the underlying distributions is required. In the case of ad CTR, the distributions are not known. This lack of knowledge is solved by using labelled examples. Say we have access to scenarios $(\bar{x}_1, y_1)(\bar{x}_n, y_n)$ that are independent and distributed according to the unknown distribution $P(\bar{x}, y)$. By using these scenarios, a decision rule can be created to classify a new feature vector $\bar{x}$.

The concept of supervised learning arises from the fact that there is access to training examples $(\bar{x}_1, y_1)(\bar{x}_n, y_n)$ that are correctly labelled as supervisor or teacher. This approach is in contrast with unsupervised learning in which many scenarios of objects are available but the class to which the objects belong is not known.

ML and AI Architecture: Sapp (2017) argues that an end-to-end architecture shown in Fig. 3 contributes to flexibility, thereby allowing for a variety of tools to process different volumes of data with varying powers of processing. The data acquisition phase involves data collection and preparation. Data processing includes pre-processing, sample selection and training of data sets. The feature analysis phase is part of this phase with the analysis of structures within the data. Data modelling involves data model designs and algorithms for machine learning. This phase also incorporates model fitting and model evaluation. In the execution phase, the trained data is forwarded for execution of machine learning routines. The deployment phase includes business-usable results of machine learning. Models or insights are deployed for data stores, enterprise applications and systems.

Implementation: The Acumos A.I. (2018) application or solution has four stages in its life-cycle as shown in Fig. 4. The first stage involves creating and on-boarding models. Stage two involves executing the solution in the target environment. In stage three, the model is enhanced with application data sets. The final stage involves sharing the models in the marketplace.

Ad CTR predictions could be accomplished in minutes, thereby helping businesses shorten the decision-making process and reduce resource utilisation. Model developers and trainers could explore the opportunities to improve the algorithm and quality of prediction. The solution incorporates high-quality algorithms which are ready for deployment or even built their in-house machine learning algorithms. Users can not only select algorithms but also improve the algorithms before running them on relevant data.

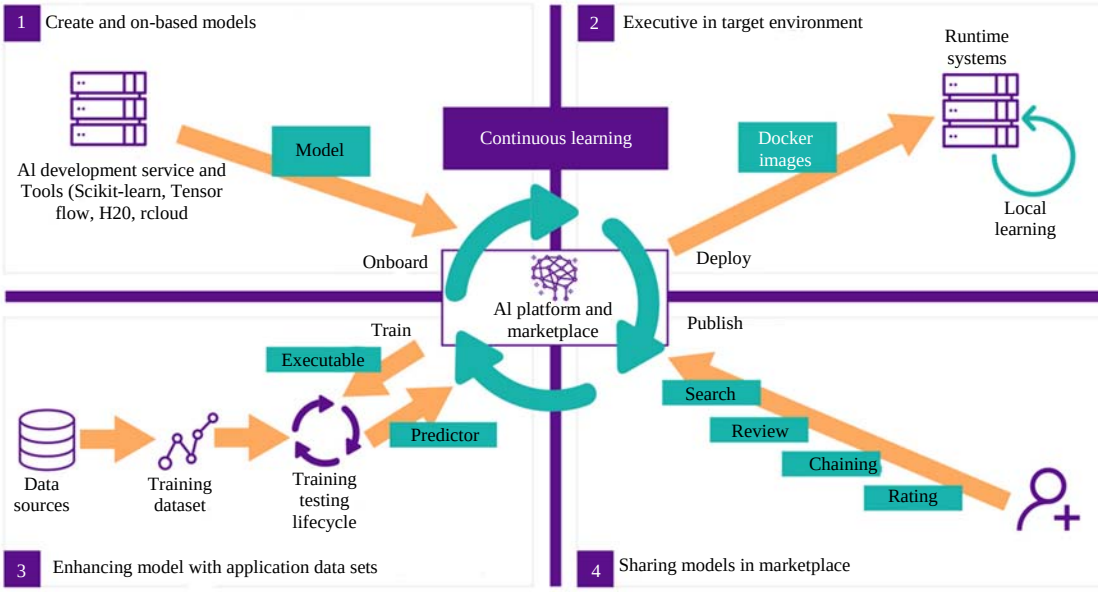

Fig. 4: Stages of ML solution development (Acumos AI., 2018)

RESULTS AND DISCUSSION

Acquisio performed a benchmark study based on 50,000 campaigns for 11,000 advertiser accounts on its platform. CTR is one of the core metrics reviewed in this study. Based on the results, Smith (2018) published the average CTRs for 12 industry groups as shown in Fig. 5 with and without machine learning.

Effects: Ad networks are contributing to complexity for marketers as it is challenging to fully comprehend the algorithms. Han *et al*. (2011) recommend machine learning to mine data effectively. The data is collected from various channels for various brands across industry sectors. Such data sets are imbalanced which could benefit from learning algorithms as observed by He and Garcia (2009). Business executives realise they are incapable of synthesising the deluge of data created by digital channels. Software tools help comb through data points for finding insights which could inform the planning, execution and optimisation of campaigns. Software tools are limited in their capabilities and the insights generated are of limited value. Despite the limitation, it is critical for marketers to discover the customer ad experiences. A critical discovery is understanding how user ad experiences across various channels could converge to a purchase. Without the ability to seamlessly access the data from various channels and combine them coherently and consistently, this aspiration becomes a challenge. The other challenge for marketers is to test and optimise campaigns with thousands of variables for achieving the correct formulation of messaging, frequency, channels and spend. Several decisions need to be made even after segments have been identified and targeted. Adjustments include channel, device and format for the campaigns. The ad ecosystem is dynamic and it is not possible to account for all the variables for maximising an ad campaign's impact. In the conventional approach, executives solve this problem by dedicating personnel and teams who focus their efforts on ads. This approach to ads is cumbersome because consumers display complex behaviour, thereby making organic discovery and attribution nearly impossible. Thus, the channel-centric approach is in contrast to the channel-agnostic behaviour of consumers. Since, the problem gets worse across channels, business executives are unable of developing a coherent cross-channel strategy. This problem indicates that marketers need an accurate and efficient technique for analysing data, attribution, budget allocation and organic discovery of customer ad experiences. This understanding could help busineses spend their resources on understanding consumer needs and crafting high-impact ad campaigns.

The AI solution is capable of handling the data across channels, thereby helping marketers save time used in managing spreadsheets and dashboards. The technology is also capable of performing this task with a high level of accuracy, vigilance and efficiency. This AI capability is affected by machine learning and AI which is the ability of the computer to learn from past actions and observations and applying those learnings to future decisions.

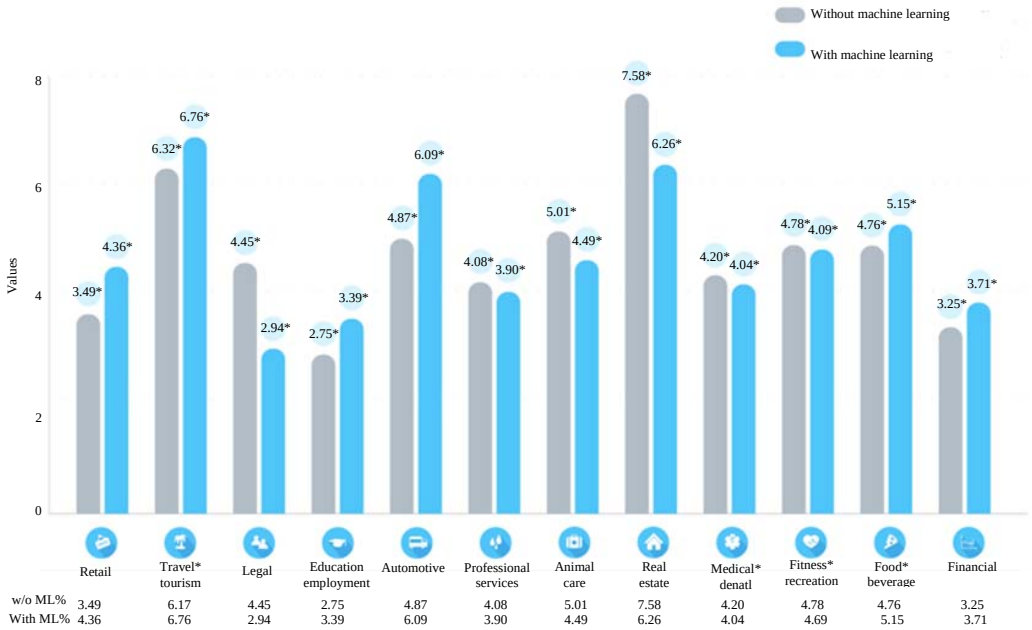

Fig. 5: CTRs by Industry (Smith, 2018)

Several concerns or issues continue to pose challenges for the widespread deployment of AI in marketing. Concerns relate to cost, expertise, technology and security (Fig. 6). There are also perceptions that machines could change the way some marketing functions could be achieved or even replace jobs. Gurin (2017) argues that one of the biggest challenges is that people do not understand what machine learning does or its limitations. There is a lack of clarity or confusion of what could be achieved by AI.

Several other obstacles are caused by technology. Some executives believe that the technology is too complex, immature or unproved for deployment. There are concerns about the sophistication of the technology and there could be problems for integrating the technology into the existing systems. Overall, there is a lack of understanding about which AI technologies are available and how to use them in their practices. Large companies have the financial resources while firms with limited resources face a steep learning curve in the absence of appropriate skills and talent within the organisation. In the majority of organisations, marketing teams lack the technical skills to leverage AI effectively. This disconnect between technologists and marketers are problematic (Fig. 7).

One concern is associated with the potential of displacing human jobs. AI has the potential for displacing some human jobs. For this reason, some marketers are hesitant to adopt AI. While the concern is genuine, it is useful to consider AI as a tool to augment the human decision-making process rather than replace jobs.

Best practices could help marketers deal with the risks. One approach is to define specific projects in their marketing function and specify goals for the solution. This approach could help deploy AI technology effectively. A top-down approach to deploying AI is counter intuitive but by reversing the approach to a bottom-up approach, marketers find an appropriate solution for AI deployment.

Since, marketers lack the technical know-how, choosing a solution wisely could help create value. While choosing solutions, it is important to understand the technical capabilities of the vendor. For some projects, a pilot project could provide an understanding of the capabilities of vendors. This could also contribute to the client's own learning. There is some ambiguity with regard to AI and data processing and data analytics. Understanding the technology as much as possible is important for the best results. By contacting experts in the field and interacting with industry professionals, marketing executives can ensure that they have bridged the knowledge gap that could help them make smart decisions. For some types of ads, there may not be adequate data sets for deep learning. In this case, the AI solution could be superficial and may not yield any valuable insights or not work over time. This could be an opportunity to innovate.

The search for a solution should also consider the human side of the problem. Though AI is intuitive, there is also a people aspect to AI. In the absence of this, the

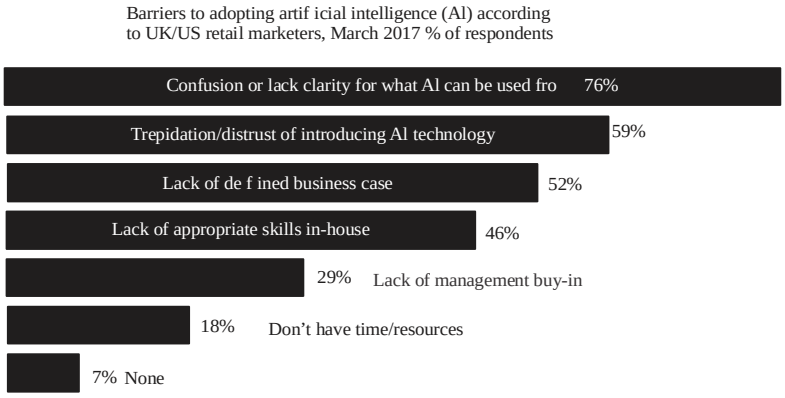

Fig. 6: Barriers to Al adoption (Gurin, 2017)

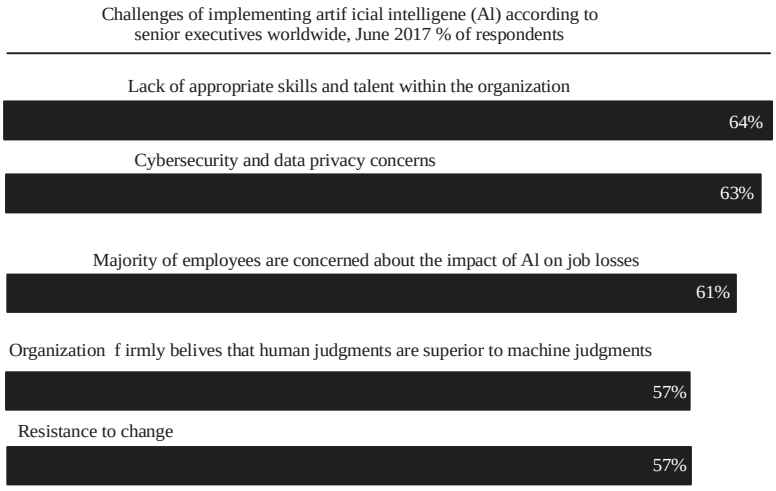

Fig. 7: Challanges for implementing Al (Gurin 2017)

solution may not be optimal. While looking to deploy an AI solution, it is also useful to plan for the future. Things are unlikely to remain the same of function the same way in the future. For this reason, it is useful to have a forward-looking approach to deploying AI. Though the same approach of finding the appropriate architecture for the problem and optimising the solution with the architecture is likely to remain the same, the underlying phenomena could be different. Thus, this machine learning platform eliminates many of the problems marketers face within their organisations by making solutions accessible.

CONCLUSION

The primary advantage of using deep neural networks for ad CTR is enhanced decision-making in comparison to conventional approaches. The solution is automated, thereby reducing the time and effort involved in the ad CTR optimisation process. Some of the challenges of this solution are the need for volumes of data and the dynamic nature of user ad experiences.

REFERENCES

Acumos, A.I., 2018. The Acumos AI platform is a complete environment for the full lifecycle of AI and ML application development. The Linux Foundation Projects, ?San Francisco, California, USA. https://www.acumos.org/platform.

Barber, D., 2012. Bayesian Reasoning and Machine Learning. Cambridge University Press, Cambridge, MA., USA., ISBN-13: 9780521518147, Pages: 697.

Boothe, M., 2017. Using machine learning for enhanced marketing results. Adobe Inc., San Jose, California, USA. https://blogs.adobe. com/digitalmarketing/digital-marketing/using-machin e-Learning-enhanced-marketing-results.

Chen, Q.H., S.M. Yu, Z.X. Guo and Y.B. Jia, 2016. Estimating ads click through rate with recurrent neural network. Proceedings of the 3rd Annual International Conference on Information Technology and Applications (ITA 2016) Vol. 7, November 21, 2016, EDP Sciences, Les Ulis, France, pp: 1-6.

Chen, T., Y. Sun, Y. Shi and L. Hong, 2017. On sampling strategies for neural network-based collaborative filtering. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2017, ACM, New York, USA., ISBN:978-1-4503-4887-4, pp: 767-776.

Cheng, H., R.V. Zwol, J. Azimi, E. Manavoglu and R. Zhang *et al*., 2012. Multimedia features for click prediction of new ads in display advertising. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12), August 12-16, 2012, ACM, New York, USA., ISBN:978-1-4503-1462-6, pp: 777-785.

Columbus, L., 2018. The 10 ways machine learning is revolutionizing marketing. Forbes, Jersey City, New Jersey, USA. https://www.forbes.com/sites/louiscolumbus/2018/02/25/10-ways-machine-learning-is-revolutionizing-marketing/#48163c725bb6

Covington, P., J. Adams and E. Sargin, 2016. Deep neural networks for youtube recommendations. Proceedings of the 10th ACM International Conference on Recommender Systems (RecSys '16), September 15-19, 2016, ACM, New York, USA., ISBN:978-1-4503-4035-9, pp: 191-198.

Greff, K., R.K. Srivastava, J. Koutnik, B.R. Steunebrink and J. Schmidhuber, 2017. LSTM: A search space odyssey. IEEE. Trans. Neural Netw. Learn. Syst., 28: 2222-2232.

Gurin, C., 2017. Artificial Intelligence for Marketers 2018: Finding Value beyond the Hype. eMarketer Inc., USA.,..

Hall, D., 2018. Machines for Machine Intelligence. Digital Catapult, London, UK.,.

Han, J., M. Kamber and J. Pei, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Morgan Kaufmann Publishers, USA., ISBN-13: 9780123814791, Pages: 744.

He, H. and E.A. Garcia, 2009. Learning from imbalanced data. IEEE Trans. Knowledge Data Eng., 21: 1263-1284.

Herbrich, R., 2017. Machine learning @ Amazon. Amazon.com Inc., Seattle, Washington, USA.

Jiang, Z., 2016. Research on ctr prediction for contextual advertising based on deep architecture model. J. Contr. Eng. Appl. Inf., 18: 11-19.

Jiang, Z., S. Gao and M. Li, 2018. An improved advertising CTR prediction approach based on the fuzzy deep neural network. PloS One, 13: 1-24.

Khoshgoftaar, T.M., Van J. Hulse and A. Napolitano, 2010. Supervised neural network modeling: An empirical investigation into learning from imbalanced data with labeling errors. IEEE. Trans. Neur. Netw., 21: 813-830.

Linoff, G.S. and M.J.A. Berry, 2011. Data Mining Techniques: For Marketing, Sales and Customer Relationship Management. 3rd Edn., Wiley Publishing Inc., Hoboken, New Jersey, USA., ISBN:9781118087459, Pages: 888.

Markus, J., 2017. What is Click-Through Rate (CTR)?. Oberlo Company, Vilnius, Lithuania. https://www.oberlo.com/ecommerce-wiki/click-through-rate-ctr

Sapp, C.E., 2017. Preparing and architecting for machine learning. MSc Thesis, Gartner Inc., Stamford, Connecticut, USA.

Shalev-Shwartz, S. and S. Ben-David, 2014. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, Cambridge, UK., ISBN:978-1-107-05713-5, Pages: 410.

Smith, B., 2018. What is a good Click through Rate (CTR) and how to improve it. Acquisio, Brossard, Canada. https:/ /www. acquisio.com/ blog/agency/what-is-a-good-click-through-rate-ctr/.

Yan, L., W.J. Li, G.R. Xue and D. Han, 2014. Coupled group lasso for web-scale ctr prediction in display advertising. Proceedings of the 31st International Conference on Machine Learning (ICML'14) Vol. 32, June 21-26, 2014, ACM, Beijing, China, pp: 802-810.

Zhai, S., K.H. Chang, R. Zhang and Z.M. Zhang, 2016. Deepintent: Learning attentions for online advertising with recurrent neural networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), August 13-17, 2016, ACM, New York, USA., ISBN:978-1-4503-4232-2, pp: 1295-1304.