International Journal of Research in Marketing 37 (2020) 481–504

Contents lists available at ScienceDirect

IJRM
International Journal of Research in Marketing
journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Machine learning and AI in marketing – Connecting computing power to human insights

Liye Ma [a,*], Baohong Sun [b]

[a] Robert H. Smith School of Business of University of Maryland, College Park, United States of America
[b] Cheung Kong Graduate School of Business Americas, New York, United States of America

ARTICLE INFO

Article history:
First received on June 2, 2019 and was under review for 5 months
Available online 21 August 2020

Area Editor: Michael Haenlein

Keywords:
Artificial intelligence (AI)
Machine learning
Digital marketing
Big data
Unstructured data
Tracking data
Network
Prediction
Interpretation
Marketing theory

Artificial intelligence (AI) agents driven by machine learning algorithms are rapidly transforming the business world, generating heightened interest from researchers. In this paper, we review and call for marketing research to leverage machine learning methods. We provide an overview of common machine learning tasks and methods, and compare them with statistical and econometric methods that marketing researchers traditionally use. We argue that machine learning methods can process large-scale and unstructured data, and have flexible model structures that yield strong predictive performance. Meanwhile, such methods may lack model transparency and interpretability. We discuss salient AI-driven industry trends and practices, and review the still nascent academic marketing literature which uses machine learning methods. More importantly, we present a unified conceptual framework and a multi-faceted research agenda. From five key aspects of empirical marketing research: method, data, usage, issue, and theory, we propose a number of research priorities, including extending machine learning methods and using them as core components in marketing research, using the methods to extract insights from large-scale unstructured, tracking, and network data, using them in transparent fashions for descriptive, causal, and prescriptive analyses, using them to map out customer purchase journeys and develop decision-support capabilities, and connecting the methods to human insights and marketing theories. Opportunities abound for machine learning methods in marketing, and we hope our multi-faceted research agenda will inspire more work in this exciting area.

1. Introduction

Consider an example of a customer purchase journey. A consumer is interested in e-book readers. She searches e-readers at Google, and reads about the products at a few websites. While watching Youtube several days later, she sees an e-reader ad. Interested, she visits the firm's website, browsing through the product description and having a few questions answered via live-chat. She then reads consumer reviews at a third-party website. Over the next few days, she is frequently exposed to ads about that e-reader while web surfing. After receiving a coupon in email, she follows the link and places an order. Happy about the product, she describes her usage experience on Facebook, and posts a few pictures on Instagram. Her friends comment on the posts with interest.

* Corresponding author.
E-mail addresses: liyema@rhsmith.umd.edu, (L. Ma), bhsun@ckgsb.edu.cn. (B. Sun).

https://doi.org/10.1016/j.ijresmar.2020.04.005
0167-8116/

Unbeknownst to the consumer, much of this journey is guided by automated systems. The search results are generated by a sophisticated Google ranking system, determined partly by advertisers' bids which are automatically generated using bidding machines. The content at the websites is customized based on her profile through website morphing, and her questions are in fact answered by chat-bots. The reviews she reads are placed prominently as they are deemed helpful by an evaluation algorithm, and the ads she repeatedly sees are delivered through retargeting algorithms via read-time bidding. The coupon which offers her the personalized price is generated by the firm's pricing engine at just the right time. Finally, her posts on social media are collected by social listening engines and analyzed for sentiment and feedback. These automated systems that make split-second context-dependent decisions are known as *Artificial Intelligence* (hereafter referred to as *AI*) agents, generally implemented using state-of-the-art *machine learning* algorithms.[1]

AI is the affordance of human intelligence to machines. The concept has been in existence since antiquity, and rigorous AI research can be traced back to 1950s, when Alan Turing established the famous *Turing Test*, stating "*I propose to consider the question, 'Can machines think?'*" (Turing, 1950), and John McCarthy coined the term *Artificial Intelligence* in 1955 when he organized the 1956 Dartmouth Summer Research Project on Artificial Intelligence. In their proposal, an AI problem is defined as "*that of making a machine behave in ways that would be called intelligent if a human were so behaving*" (McCarthy, Minsky, Rochester, & Shannon, 1955). Initially, AI research focused on math and logical reasoning problems. The focus shifted to knowledge and expert systems in the 1980s. Although both waves showed initial promise, they faced considerable challenges and delivered results below the overly optimistic expectations.[2]

Since 1990s, AI research has mostly focused on machine learning methods. The widely used definition of *Machine Learning* is given in Mitchell (1997): "*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.*" Although first evolved separately, machine learning has become the main paradigm of AI research, and *is typically considered as a subfield of AI* (Goodfellow, Bengio, & Courville, 2016, p.9). The past decade has witnessed impressive breakthroughs in AI performance, on image recognition, speech processing, autonomous driving, and many other tasks typically considered as demonstrating human level intelligence. Behind much of this breakthrough is *deep learning* (LeCun, Bengio, & Hinton, 2015), methods that use multiple levels of representation typically implemented using neural networks with multiple hidden layers, combined with the advent of big data and the exponential growth in computer hardware. Other machine learning methods have also benefited from this trend and expanded their applications.

Within two decades, AI has significantly transformed fields such as biology, education, engineering, finance, and healthcare. Marketing is no exception (Huang & Rust, 2018; Rust, 2020). Interactions between firms and consumers are increasingly more individualized and ubiquitous, generating heavily digitized footprints. The abundance of data has prompted companies to invest heavily in machine learning to enhance their marketing capabilities. According to BCC Research, the global market of machine learning-enabled solutions would grow at an annual rate of 43.6%, reaching $8.8 billion by 2022.[3] Sophisticated machine learning algorithms power the recommender systems at e-commerce websites and content platforms such as Amazon and Netflix; deep learning engines analyze and tag the billions of images on social media sites such as Facebook; automated bidding algorithms examine a web surfer's profile in millisecond timescale to determine the optimal bid for ad delivery; chatbots engage human-like conversations with customers to maintain relationship and loyalty. Through these and a myriad of other applications, such as social media mining, sentiment analysis, and customer churn prevention, AI agents powered by machine learning algorithms have demonstrated their effectiveness in processing large-scale and unstructured data in real-time, generating accurate predictions to assist marketing decisions. These initiatives have greatly improved all aspects of business performance.

Meanwhile, the complex marketing contexts and big data present a formidable challenge to researchers. While statistical and econometric models with increasing levels of sophistication are being developed, researchers have also turned to machine learning methods as a valuable alternative. A diverse set of machine learning methods, such as support-vector machine (Cui & Curry, 2005; Evgeniou, Boussios, & Zacharia, 2005), topic models (Tirunillai & Tellis, 2014; Trusov, Ma, & Jamal, 2016), ensemble trees (Guo, Sriram, & Manchanda, 2018; Yoganarasimhan, 2018), deep neural networks (Liu, Dzyabura, & Mizik, 2018; Liu, Lee, & Srinivasan, 2018), network embedding (Ma, Sun, & Zhang, 2019), among others, have been used in academic marketing research for prediction and insight generation. Furthermore, they are often applied to situations where traditional quantitative methods are not a good fit, making them especially helpful. However, although the interest is rapidly growing, the use of machine learning methods in marketing is still at an early stage, and extant studies are somewhat scattered. To date, there does not appear to exist a consensus vision or a unified framework on how machine learning methods should be incorporated into marketing research.

The importance of machine learning methods warrants a systematic review and a forward-looking discussion, which we provide in this paper. We begin with an overview of machine learning tasks, including supervised, unsupervised, semi-supervised, active, transfer, and reinforcement learning, and briefly introduce common machine learning methods for handling such tasks. Compared with statistical and econometric models traditionally used in marketing, machine learning methods can effectively process large-scale and unstructured data, and have flexible structures to approximate complex functions which yield strong predictive performance. Meanwhile, machine learning methods often do not enable intuitive interpretation, especially at the causal level, and their ability to capture individual consumer level heterogeneity and dynamics is yet unproven. We then discuss several salient

[1] The definitions of artificial intelligence and machine learning can be found in Table 1. The list of acronyms used in the paper is summarized in Table 2.
[2] Both Herbert Simon and Marvin Minsky, for example, predicted in 1960s that the problem of AI will be solved within a generation.
[3] https://www.bccresearch.com/market-research/information-technology/machine-learning-global-markets.html.Accessed in November 2019.

Table 1
Key definitions.

Term	Definition
Artificial Intelligence	An artificial intelligence problem is that of making a machine behave in ways that would be called intelligent if a human were so behaving (McCarthy, Minsky, Rochester, & Shannon, 1955)
Machine Learning	A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. (Mitchell, 1997)
Deep Learning	Representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level into a representation at a higher, slightly more abstract level. (LeCun, Bengio, & Hinton, 2015)

AI-driven trends in the industry, where firms increasingly focus on the entire customer purchase journeys which contain frequent and media-rich interactions, and perform large-scale and automated context-dependent personalization and targeting. These trends both are enabled by and drive the continued improvement of the ever more powerful machine learning methods, leading to a positive feedback loop which is transforming all areas of marketing practices. Next, we review the small but rapidly growing marketing literature that has used a number of machine learning methods.

More importantly, we present a unified framework and a multi-faceted agenda for incorporating machine learning methods into academic marketing research. This framework contains five key components of empirical research: *method*, *data*, *issue*, *usage*, and *theory*. At the center of the framework is the rich family of machine learning *methods*. We call for researchers to develop a broad and in-depth understanding of machine learning methods, and to introduce more methods to and demonstrate their compelling values in marketing research. Meanwhile, we call for machine learning methods to play more central roles in addressing core marketing issues, e.g. to model consumer decisions, in contrast to playing supporting roles of feature extraction as is often done. Moving to the center helps better connect machine learning methods to the marketing literature, broadening their impact. Furthermore, we call for researchers to extend the methods, e.g. to improve interpretability, to make them more suitable for social science research. Finally, as the methods are introduced and extended, developing a clear understanding of the boundaries of their abilities is also a crucial aspect of research.

Machine learning methods need to be closely linked to the other four components in the unified framework. First, the methods are well positioned to extract rich insights from rich *data*. While studies have frequently analyzed text and image data, we encourage researchers to also focus on audio, video, and consumer tracking data, as well as network data and data of hybrid formats. Second, the *usage* of machine learning methods in marketing research should be broadened and extended. Usages, including prediction, feature extraction, descriptive interpretation, causal interpretation, prescriptive analysis, and optimization, each have different requirements on method transparency and theoretical connections, posing different challenges. While machine learning methods have been used frequently for prediction and feature extraction, we call for more effort to use them for descriptive, causal, and prescriptive analysis, and to evaluate the methods' intrinsic abilities or limitations for these usages. Third, methods should be used to address important substantive *issues*. We call for using machine learning methods to help map out the entire

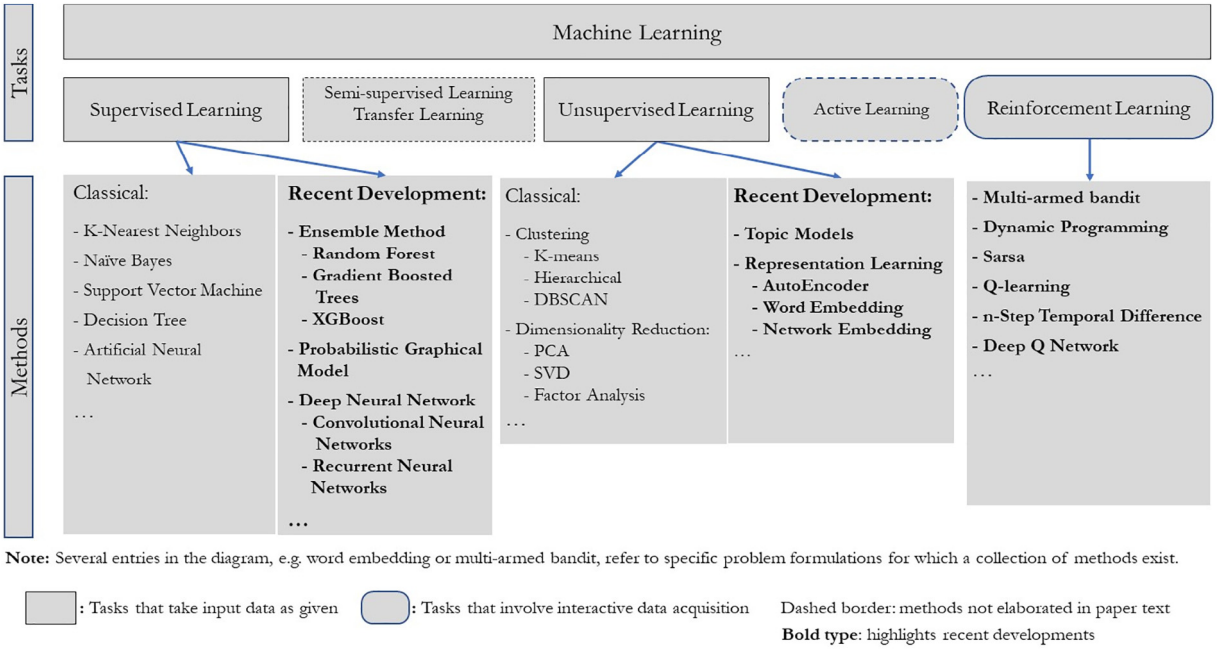

Fig. 1. Machine learning tasks and methods.

customer purchase journey, especially in early stages, to develop decision-support capabilities covering all aspects of marketing functions, and to perform holistic market structure analysis including brand positioning and competitive analysis. Fourth, we call for the unification of machine learning methods and marketing *theories*. Challenges remain on making such connections, however, given the nature of machine learning methods. We call for injecting human insights and domain knowledge into the use of machine learning methods, balancing a theory-driven perspective with a data-driven one, exploring potential theoretical connections of various types and aspects of machine learning methods, and investigating the theoretical implications of firms' adoption of AI tools. While this theoretical agenda is challenging, they also present opportunities for impactful research.

The rest of this paper is organized as follows. In Section 2, we briefly introduce common machine learning tasks and methods, and discuss their strengths and weaknesses. Following that, we discuss AI-driven industry trends and practices in Section 3. In Section 4, we briefly review the extant machine learning literature in marketing. In Section 5, we present the conceptual framework and the multi-faceted research agenda. We conclude in Section 6.

2. Machine learning tasks and methods – a brief overview

Machine learning is a vast and rapidly evolving field, encompassing a wide range of methods for addressing diverse tasks. A comprehensive review of the machine learning literature is beyond the capability and scope of this paper. Instead, we briefly discuss common machine learning tasks and methods that are likely helpful for academic marketing research. They are summarized in Fig. 1.

2.1. Machine learning tasks

Traditionally, a machine learning algorithm processes a given dataset for a specified objective, with the algorithm playing no role in data acquisition. In this paradigm, two major task categories are *supervised* and *unsupervised* learning.

2.1.1. Supervised learning

In supervised learning tasks, a *training* dataset is provided such that for each instance, both the *input*, a collection of variables commonly denoted as X, and the *output*, a target variable commonly denoted as Y, are observed.[4] Supervised learning seeks to learn from this training dataset a function, $Y = f(X)$, to predict the output when given an input. If Y is a numeric/categorical variable, the task is referred to as *regression/classification*. Prediction is the key focus of supervised learning. Typically, researchers are less concerned with uncovering the "true" linkage between the variables, than with learning a function that maximizes the accuracy of predicting the output using the input. The predictive accuracy must be evaluated using a different *testing* dataset, as absent model constraints, perfect accuracy can be achieved for the training dataset through memorization. This testing dataset can be used to select models or to tune hyperparameters. Once involved in model selection, though, the chosen model's performance on this dataset is no longer an unbiased evaluation of its out-of-sample performance. Therefore, researchers typically further split the training dataset into a *training* subset and a *validation* subset. Models will be trained using the training subset and tuned or selected using the validation subset. The final chosen model will then be evaluated using the testing dataset to assess the out-of-sample performance. Researchers also routinely perform *cross-validation*, by iteratively using different portions of the training dataset for training and validation (e.g. Hartmann, Heitmann, Schamp, & Netzer, 2019).

2.1.2. Unsupervised learning

In unsupervised learning tasks, the training dataset contains only the input variables, while the output variables are either undefined or unknown.[5] The typical goal is to find hidden patterns in or extract information from the data. Many such tasks exist. In a *clustering analysis*, the input instances are put into multiple groups to maximize within-group similarity and cross-group difference. In a *dimensionality reduction* task, high dimensional data are transformed into lower dimensional variables while retaining the information in the original data. In an *unsupervised feature learning* or *representation learning* task, features are extracted from the input data to represent them. The extracted features carry key information of the original data and can be interpreted or used as input for subsequent analysis.

2.1.3. Semi-supervised learning and transfer learning

The line between supervised and unsupervised learning can be blurred, leading to other learning tasks. In a semi-supervised learning task (Zhu, 2005), the output is known for only a subset of the data. The instances in the training dataset for which the output is not observed are nonetheless used to improve learning, e.g. through label propagation. In a transfer learning task (Pan & Yang, 2009), researchers leverage an existing model, trained using a different dataset or for a different purpose, for the task at hand. The existing model can serve as a starting point, which is then adjusted based on the current training dataset. For models that require a large amount of training data and computation time, transfer learning can be effective at leveraging existing knowledge. For example, it is routinely used in image analysis, where an existing model trained using a large set of images is updated using the specific images of the research project (e.g. Dzyabura, El Kihal, Hauser, & Ibragimov, 2019; Hartmann, Heitmann, et al., 2019).

[4] For example, the input may be a customer's past interactions with a firm, and the output the customer's purchase decision.
[5] Using the earlier notation, in an unsupervised learning task only X is known, but not Y.

While machine learning algorithms traditionally take datasets as given, situations also exist where the algorithm is involved in the acquisition of data. Commonly included in this interactive paradigm are *active learning* and *reinforcement learning* tasks.

2.1.4. Active learning

In an active learning task (Cohn, Ghahramani, & Jordan, 1996; Lewis & Gale, 1994), only limited training instances are available at first. The algorithm can acquire additional training instances to improve predictive accuracy, although such data acquisition is costly. The goal is to maximize the predictive accuracy while minimizing the data requirement. Determining the most important training instances is a key focus of active learning.

Reinforcement learning: In a reinforcement learning task (Sutton & Barto, 2018), the learning agent continuously interacts with the surrounding environment by taking actions and observing feedbacks, in order to optimize a certain objective function. These tasks are often formulated as a Markov decision process (MDP), a structure familiar to marketing researchers who investigate forward-looking behaviors using dynamic programming models. The learning algorithm needs to determine the actions to take to both learn the environment's characteristics and craft optimal policy of actions given the states. This type of machine learning tasks has received heightened attention due to recent methodological advancement and increasing usages in the industry, from autonomous vehicles to morphing websites.

2.2. Machine learning methods

Many machine learning methods transcend the boundaries of disciplines, e.g. linear regression and logistic regression are widely used in machine learning and many other fields, including marketing. While the perspective and emphasis are different in different fields, the technical foundations are the same. Meanwhile, other machine learning methods are used less frequently in marketing research, for which we provide a brief discussion. Given the vastness of the field, we cover only the common supervised, unsupervised, and reinforcement learning methods that we believe are closely related to marketing research, with emphasis given to recent developments.

2.2.1. Methods for supervised learning

2.2.1.1. Classical methods. *K-nearest neighbor* (*kNN*) is a widely used instance-based method for supervised learning (Altman, 1992; Stone, 1977). Let $\{X_1, Y_1\}$, …$\{X_N, Y_N\}$ be the training instances, where X_i is the input and Y_i is the output. To use kNN, a distance metric over the input space is first defined. For each test instance X_{test}, the k training instances with the shortest distances to it are identified, and a function, e.g. weighted arithmetic mean, of these k nearest neighbors is then used for prediction. *Naïve Bayes* (*NB*) is a classification method based on the Bayes theorem: $p(Y|X) = p(X|Y)p(Y)/p(X)$. The Bayes classifier which chooses the class that maximizes the posterior probability is theoretically sound, but empirically infeasible for high-dimensional input vectors. An NB classifier assumes each input dimension to be independent, reducing the task to the estimation of a set of univariate distributions. Despite the strong assumption, NB performs competitively and is widely used, particularly in text mining. *Support-vector machine* (*SVM*) is a powerful maximum margin classifier (Cortes & Vapnik, 1995; Vapnik, 1998). In a binary classification task, SVM seeks to establish a linear hyperplane in the input space between the two classes that maximizes the margin of error. When linear hyperplanes are inadequate, SVM can generate non-linear classification boundaries using the kernel trick, by mapping the original input space into higher dimensional spaces. As an extension, soft-margin SVM allows for but penalizes misclassifications. Naturally a binary classifier, SVM can also be used for multi-class classification through training multiple one-versus-the-rest or one-versus-another binary classifiers, or through modified methods that consider all classes at once (Hsu & Lin, 2002). Structured SVM further addresses the problems involving multiple dependent variables (Tsochantaridis, Joachims, Hofmann, & Altun, 2005). SVM can also be used for regression tasks, for which it is known as *support-vector regression* (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997). *Decision tree* is a popular method in machine learning (Quinlan, 1986),[6] where the input dataset is successively split into subsets. Each step splits an existing subset into two new ones based on the value of one input variable, chosen to optimize a criterion such as the information gain. Overfitting can be mitigated through ex-ante regularization which constrains the minimum leaf node size or the maximum tree depth, or through ex post pruning. Decision tree offers intuitive interpretations, as each split is explainable based on the corresponding input variable. *Artificial neural networks* (*ANN*) is a flexible and powerful family of machine learning methods, originally inspired by biological neural networks in animal brains (Lippmann, 1987; McCulloch & Pitts, 1943). ANN lays out a network of interconnected processing units called artificial neurons. Each neuron takes a list of variables, performs a certain calculation, and generates an output using an activation function. Model training can be conducted through back propagation.

2.2.1.2. Recent developments. While the set of classical supervised learning methods is already powerful, recent developments have further expanded the frontier. Among these important developments are *ensemble methods*, *probabilistic graphical models*, and *deep neural networks*.

Ensemble methods are meta-learning algorithms that combine multiple individual learners. The intuition is that by combining a set of weak learners complementarily, a stronger learner can emerge. Common approaches include *stacking*, *bootstrap aggregating*

[6] Also called classification and regression tree for handling classification and regression tasks.

(*bagging*), and *boosting*. In stacking, a linear combination of individual predictors is used to achieve better accuracy (Breiman, 1996). In bagging, each individual predictor is learned using a bootstrap sample of the training dataset, with the predictors then aggregated to generate a final prediction. In boosting, individual learners are trained sequentially, and combined according to their respective accuracy to produce a stronger learner. A popular boosting method is the *adaptive boosting* (*AdaBoost*), where subsequent learners are tweaked in favor of the training instances that the previous learners have misclassified.

Two popular ensemble methods using decision trees as base learners are *random forest*, or *RF* (Breiman, 2001), and *gradient-boosted trees,* or *GBM* (Friedman, 2001). Using RF, each individual tree is built from a bootstrap sample of the original data. Meanwhile, for each split, only a random subset of the input variables is considered so as to reduce correlation. RF generates the final prediction by averaging the predictions of individual trees. Using GBM, multiple trees are trained sequentially. Each tree reduces the remaining errors after applying previously built ones. Both RF and GBM are widely used and are among the methods with the best predictive performance. XGBoost, a sparsity-aware algorithm of gradient tree boosting, is behind many winning submissions to data science competitions at Kaggle (Chen & Guestrin, 2016). Recent methodological advancement has also enabled causal inference using tree ensembles (Wager & Athey, 2018).

Probabilistic graphical models (PGM), also known as structured probabilistic models, is a large family of probabilistic models that use directed or undirected graphs to encode the conditional dependence of random variables (Koller & Friedman, 2009). In a PGM, variables that are not directly connected in the graph have conditional independence properties that help simplify the analysis of the complex joint distribution. PGM includes *Bayesian networks* which use directed acyclic graphs and *Markov random fields* which use undirected graphs. This family of models has been extensively used in both machine learning and statistics. For example, the hidden-Markov model (HMM) used in numerous marketing studies (e.g. Ma, Sun, & Kekre, 2015; Netzer, Lattin, & Srinivasan, 2008) is a specific type of PGM.

Deep neural networks, i.e. ANNs with more than one hidden layer, have been a main driver behind of recent surge in AI. While already well developed early on and theoretically shown to be able to approximate any continuous function, several limitations hindered ANN's practical use. First, an ANN often has millions of parameters and requires a large amount of data for training; second, training an ANN is computationally demanding; third, the large number of parameters make an ANN prone to overfitting. These hurdles have been surpassed in the last decade. The advent of big data provided ample inputs to train complex models; the use of Graphical Processing Unit (GPU) chips enabled fast parallelized training; overfitting issues were also alleviated, e.g. through using *dropout* units. These factors combine to make deep neural networks highly successful in recent years, outperforming alternative methods on many challenging AI tasks such as image and speech processing (LeCun et al., 2015).

Two prominent examples of deep neural networks are *convolutional neural networks* (*CNN*, LeCun & Bengio, 1995) and *recurrent neural networks* (*RNN*, Rumelhart, Hinton, & Williams, 1986). CNN uses convolution in at least one layer (Goodfellow et al., 2016, p.321). A typical CNN consists of multiple convolutional and pooling layers. In a convolutional layer, each unit performs a convolution operation to a local region of the input data, and weights are shared across the units. These enable the detection of local patterns in a location-invariant manner. Meanwhile, pooling layers are used to sample local regions. The successive layers extract data representations at different abstract levels, making CNN highly successful in analyzing image and video data. RNN is a class of deep neural networks that keeps internal states. In an RNN, the output of each hidden layer is fed back to itself, allowing the network to maintain a memory. A plain RNN suffers from vanishing gradient during back propagation. This issue is alleviated using the *long short-term memory* (*LSTM*) architecture which retains long term memories. LSTM and the similar *gated recurrent-unit* (*GRU*) architecture have been successful on processing sequential data such as text, speech, and time series. Recently developed attention mechanism and the transformer architecture have made further breakthroughs in natural language processing (Bahdanau, Cho, & Bengio, 2014; Dai et al., 2019; Vaswani et al., 2017).

2.2.2. Methods for unsupervised learning

2.2.2.1. Classical methods. In clustering analysis, units are assigned to separate groups to maximize within-group similarity and cross-group distinction. Common algorithms include *K-means*, which randomly initializes groups and then iterates over recalculating group means and reassigning groups, and *Hierarchical clustering*, which follows a bottom-up approach by starting with each unit as a separate group and iteratively merging similar groups. *DBSCAN* can find groups of arbitrary shapes and is robust to data noise (Ester, Kriegel, Sander, & Xu, 1996). For dimensionality reduction, where original high-dimensional data are reduced to lower dimensional variables, common methods include *principal component analysis* (*PCA*), *singular value decomposition* (*SVD*), and *factor analysis* (*FA*). These classical methods have been widely used in machine learning and other fields, and are familiar to marketing researchers.

2.2.2.2. Recent developments. Topic models discover and extract hidden semantic structures from textual data, with the semantic information represented using *topics*. A text document is treated as a mixture of topics, and each topic is a distribution over words in the vocabulary. *Latent Dirichlet Allocation*, or *LDA* (Blei, Ng, & Jordan, 2003), uses Dirichlet priors for both the document-topic and the topic-word mappings. Extending from prior information retrieval methods such as the probabilistic latent semantic indexing (Hofmann, 1999), LDA is the first and most commonly used topic model. Extensions exist to incorporate topic correlation or covariates (Blei & Lafferty, 2007; Mimno & McCallum, 2012). Marketing research has used this method, both to process textual data as it was originally intended to (Liu & Toubia, 2018; Tirunillai & Tellis, 2014), and to analyze other data where similar semantic structures exist (Jacobs, Donkers, & Fok, 2016; Trusov et al., 2016).

Unsupervised representation learning methods have been instrumental in the progress of deep learning (Bengio, Courville, & Vincent, 2013; Hinton, Osindero, & Teh, 2006). One good example in this family is the *word embedding* method *word2vec* (Mikolov, Chen, Corrado, & Dean, 2013).[7] Traditional text mining uses a high-dimensional bag-of-words vector representation, where each word in the vocabulary occupies a separate dimension. Using word2vec, each word in the vocabulary is mapped to a much lower dimensional vector, and the semantic relationship among words is preserved in the distances among their embedding vectors. The method is a form of *autoencoder*, a type of neural networks for extracting meaningful representation of input via the hidden layers (Hinton & Salakhutdinov, 2006). Autoencoder is trained by using the raw data as both input and output, with the middle layer being learned as the representation. Word2vec has shown superior performance in many challenging text analysis tasks such as analogy. Extensions such as ELMo, which derives deep contextualized word representations, have further improved the state of the art on challenging natural language processing tasks (Peters et al., 2018). Meanwhile, embedding has also been used to process large scale network data. Similar to word embedding, these *network embedding* methods (e.g., Dong, Chawla, & Swami, 2017; Tang et al., 2015) map nodes in a large network to vectors in a lower dimensional vector space, while preserving the semantic and structural information of the network. The embedding vectors can then be processed to generate insights.

2.2.3. Methods for reinforcement learning

In the interactive paradigm of reinforcement learning, an agent repeatedly probes an environment to learn about its characteristics and to formulate an optimal policy. One of the earliest examples is the *multi-armed bandit* (*MAB*) problem and the associated solution methods. In a classic *k*-armed bandit problem, an agent repeatedly makes a choice from *k* options, each generating a reward drawn from a corresponding probability distribution, to maximize the overall reward. The *greedy* algorithm of choosing the best option given the past knowledge produces reasonable results. In comparison, the ε-greedy algorithm, which chooses the best option most of the time but also experiments with other options, can perform better due to its balance between exploration and exploitation. When actions not only generate rewards, but also change the agent's states such that future rewards are affected, an MDP framework is adopted. When the state-dependent reward functions and state transition functions are known, the optimal policy can be derived using *dynamic programming* (DP) methods. DP methods such as policy iteration, value iteration, and backward induction have been developed in numerous fields including mathematics, economics, engineering, and computer science, and have been used extensively in marketing research to analyze consumers' forward-looking behaviors (e.g. Erdem & Keane, 1996).

In modern reinforcement learning problems, reward or state transition functions are typically not known. The method needs to both learn about the environment and craft the optimal policy. One popular family of methods is *temporal-difference* (*TD*) learning, of which the *SARSA* algorithm is a standard. Using SARSA, the learning agent iteratively takes action based on the action-value function, and updates the action-value function based on the reward feedback.[8] SARSA is an example of *on-policy* method, in that the actions are generated based on the current policy. In comparison, *off-policy* TD methods use a separate behavior policy to generate actions while learning the optimal target policy, among which *Q-learning* is a prominent example (Watkins, 1989; Watkins & Dayan, 1992). Unifying the TD methods with Monte Carlo methods leads to the development of the more general *n-step Temporal Difference* methods, including the n-step SARSA and off-policy n-step SARSA methods. Another major challenge to reinforcement learning is the usually highly complex state space. To address this, deep neural networks are incorporated into reinforcement learning, leading to deep reinforcement learning methods. A prominent example of this group is *deep Q-network*, which generally follows the same process as Q-learning, and uses deep neural networks to map states and actions to values. Deep Q-network and its extensions have seen notable successes. For example, Alphabet's DeepMind has demonstrated that agents trained using such methods can surpass human level capabilities in a wide range of games.

2.3. Strengths and weaknesses of machine learning methods

Significant differences exist between machine learning methods and the statistical and econometric models commonly used in quantitative marketing research. We briefly discus their relative strengths and weaknesses, which are summarized in Table 3.

2.3.1. Strengths of machine learning methods

The first key strength of machine learning methods is that they can readily handle unstructured data such as text, image, audio, and video, and can process data with complex structures such as large scale network or tracking data. In addition, machine learning methods can accommodate data of hybrid formats, such as a combination of text, image, and structured data, in an integrated manner. Unstructured data have been the key driver of the data explosion in recent years, elevating the importance of machine learning methods.[9]

Second, compared with econometric models, machine learning methods can handle larger data volume. Using econometric models, data are typically at the scale of several hundred or thousand consumers, with a modest number of variables and restricted choice sets. Forward-looking models may use even smaller samples. In machine learning research, in contrast, millions of observations are the norm, while larger datasets abound. Efficient optimization algorithms such as stochastic gradient descent

[7] Mathematically, embedding is an injective and structure-preserving map.
[8] SARSA stands for “state-action-reward-state-action”.
[9] Industry practitioners state that more than 80% of the new data created every year are unstructured (Grimes, 2008).

Table 2
List of acronyms used in the paper.

Acronym	Full name	First described in section
AI	Artificial Intelligence	1
ANN	Artificial Neural Networks	2.2.1
CNN	Convolutional Neural Network	2.2.1
CTM	Correlated Topic Model	4
DBSCAN	Density-Based Spatial Clustering of Applications with Noise	2.2.2
DP	Dynamic Programming	2.2.3
DT	Decision Tree	2.2.1
FA	Factor Analysis	2.2.2
GBM	Gradient Boosting Machine	2.2.1
GPU	Graphical Processing Unit	2.2.1
GRU	Gated Recurrent Unit	2.2.1
HMM	Hiddent Markov Model	2.2.1
IoT	Internet-of-Things	5.3
kNN	k-Nearest Neighbor	2.2.1
LDA	Latent Dirichlet Allocation	2.2.2
LSTM	Long Short-Term Memory	2.2.1
MAB	Multi-Armed Bandit	2.2.3
MDP	Markov Decision Process	2.2.3
NB	Naïve Bayes	2.2.1
PCA	Principal Component Analysis	2.2.2
PSM	Propensity Score Matching	5.4
PGM	Probabilistic Graphical Models	2.2.1
PLSI	Probabilistic Latent Semantic Indexing	2.2.2
RF	Random Forest	2.2.1
RNN	Recurrent Neural Network	2.2.1
SARSA	State-Action-Reward-State-Action	2.2.3
SVD	Singular Value Decomposition	2.2.2
SVM	Support Vector Machine	2.2.1
TD	Temporal-Difference	2.2.3
UGC	User-Generated Content	3.1

and parallel computing enable efficient trainings on large datasets. Off-the-shelf tools with high-performance computing capacities also make implementation easy. With the ever-increasing data volume in the real world, scalable methods are clearly desired.

The third advantage of machine learning methods is their flexibility, beginning with input construction via *feature engineering*. Using econometric models, observed variables are usually entered directly for statistical inference, with manipulations restricted to normalization, monotonic transformation, or adding selected interaction terms. Machine learning, however, encourages extensive upfront efforts on creating and transforming input variables. One variable, for example, can have its original form, binned form, higher order terms, and interaction terms, all entered as input. Additional transformations are also performed based on researchers' domain knowledge. Feature engineering is frequently cited as a key success factor in winning entries of data science contests (e.g. Romov & Sokolov, 2015). Furthermore, the flexibility is also reflected in the model structure. While marketing models typically prescribe specific functional forms, e.g. linear utility functions or the solutions to DP problems, machine learning methods strive for flexibility. Using a regression tree, arbitrary regions of the feature space can be carved out. Using SVM, kernels map the original variables into higher or even infinite dimensional spaces. Using deep neural networks, many layers lay between the input and the output, performing complex transformations. Combining feature engineering and flexible model structure increases the chance of capturing the true linkage between the input and output variables.

Fourth, due partly to the above reasons, machine learning methods excel in prediction, especially in real-world settings. While econometric models typically focus on causal identification and interpretation, machine learning methods are evaluated according to their out-of-sample predictive accuracy. Popular open data contests, e.g. Kaggle competitions, are usually won by teams using machine learning methods such as deep neural networks or ensemble methods. The increasing complexity of products, markets, and decision contexts makes prediction increasingly important, making this another major advantage of machine learning methods.

Table 3
Strengths and weaknesses of machine learning methods.

Strength	• Ability to handle unstructured data and data of hybrid formats • Ability to handle large data volume • Flexible model structure • Strong predictive performance
Weakness	• Not easy to interpret • Relationship typically correlational instead of causal • Unproven on analyzing individual consumer level heterogeneity and dynamics

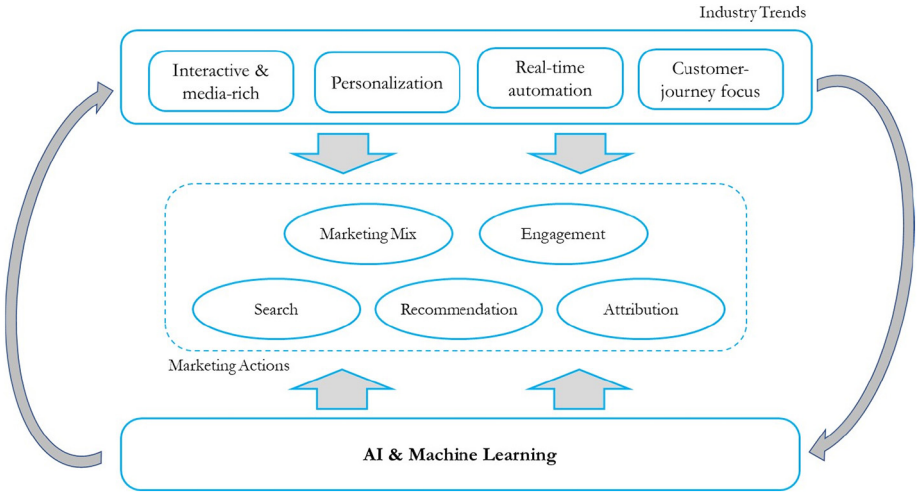

Fig. 2. AI-driven marketing landscape.

2.3.2. Limitations of machine learning methods

The first key limitation of machine learning methods is they often lack interpretability, in terms of having a transparent model structure and clear linkage between variables.[10] In research using econometric models, the type of model is chosen according to the theory, e.g. choices models represent utility maximizing decisions, and DP models represent consumers' intertemporal optimization. The functional form may also reflect theory, e.g. the Bass model represents diffusion driven by innovators and imitators. Variables are also selected according to theory, and are included in a way that statistical hypothesis testing can be performed to interpret how they are related, e.g. exposure to an ad increases utility, or to evaluate characteristics of interest, e.g. consumers demonstrate risk aversion. In contrast, machine learning methods rely on both heavily engineered features and flexible model structures, resulting in a black-box which delivers on predictive accuracy, but not on interpretive insights which marketing researchers have come to expect. Furthermore, these methods are often developed in an optimization framework, making parametric statistical hypothesis testing infeasible, presenting another hurdle for interpretation. However, we should also note several mitigating factors: 1) many machine learning methods, e.g. PGM, do have well established statistical foundations with interpretable parameters; 2) post-hoc interpretation techniques exist for machine learning methods (Lipton, 2016) which, while different from interpretations based on model structure and parameters, can still generate important insights; 3) certain methods have been successfully adapted for interpretation (e.g. Wager & Athey, 2018).

Second, relationships uncovered using machine learning methods are often correlational rather than causal. With a predictive focus, little attention has been paid to endogeneity concerns when developing machine learning methods. Issues such as selection, omitted variables, and simultaneity, which are painstakingly addressed in econometric models, are typically ignored in machine learning analysis. Consequently, functions learned from machine learning cannot be readily taken as causal, and the predictive performance may not hold in the case of regime shift or policy change. This lack of causal capacity makes it challenging to perform counterfactual analysis, which is essential for designing and evaluating marketing mix or other important decisions such as segmentation and engagement. This limits their use in a core capacity in marketing.

Third, while machine learning methods enjoy good predictive performance, their ability to capture individual consumer level heterogeneity and dynamics is yet unproven. Traditionally, dynamic data are typically handled using time series models or specific PGMs such as HMM, which arguably fall more into the field of statistics than machine learning per se. Even state-of-the-art RNN based models find success mostly in tasks such as speech recognition and translation, while their power for capturing consumer dynamics is largely unexplored. Accounting for unobserved heterogeneity and dynamics is essential in many areas of marketing research, and it remains to be seen the extent to which machine learning methods can contribute on this front.

3. AI-driven marketing industry trends

AI has been and will continue to dramatically transform marketing (Huang & Rust, 2018; Huang & Rust, 2020; Rust, 2020). Marketing practices such as digital search and advertising, social media interaction, mobile tracking and engagement, online purchase, and in-store shopping experience, are increasingly powered by scalable and intelligent algorithms, with the help of both technology powerhouses such as Google and Amazon and many smaller MarTech (Marketing Technology) companies. Machine learning has propelled several *major trends in marketing*, which in turn pose challenges to and drive the evolution of machine learning methods, creating a positive feedback loop, as depicted in Fig. 2. We now briefly discuss these marketing trends and practices.

[10] "Interpretability" is a widely discussed and debated concept in machine learning. The term can mean different things in different contexts. We refer readers to Lipton (2016) for a comprehensive discussion on this topic. The interpretability discussed here roughly corresponds to "decomposability" and "algorithmic transparency" in Lipton (2016).

3.1. Marketing trends

3.1.1. Interactive and media-rich

Internet social media and mobile devices have dramatically increased the interactions between firms and consumers, with the information encoded in rich media formats such as text, image, and video. It is imperative for firms to understand consumer perceptions and preferences and obtain brand positioning insights based on this rich media content. Similarly, designing informative and engaging content to enhance awareness, perception, and acceptance, is a crucial area of competition. While the rich media content is natural for humans, it is challenging to machines.[11] With their superior capacity, AI tools powered by machine learning methods are relied upon to generate insights and prescribe solutions in these interactive and media-rich environments.

3.1.2. Personalization and targeting

Marketing is becoming increasingly personalized (Chung, Rust, & Wedel, 2009; Rust & Huang, 2014; Chung, Wedel, & Rust, 2016). In markets where rich data are commonplace and digital channels make personalized offerings easy to deliver, machine learning methods are propelling large-scale context-dependent personalization and targeting to a new level. Segmentation is getting more fine-grained, operating on hundreds of precisely carved-out microsegments instead of a few coarsely defined large ones. Preference and behavior data, e.g. brands liked on Facebook, products searched at Google, apps used on mobile phones, etc., are increasingly incorporated into segmentation. Continued refinement leads to personalization, where each consumer constitutes a distinct segment, receiving tailored offers based on her individual profile. Taking it one step further, preferences and needs are dynamic, and opportunities temporary. Effective targeting thus requires not only matching the right offering to the right consumer, but also delivering it at the right moment in the right context. Much of personalization and context-dependent targeting is driven by powerful machine learning algorithms, and these practices also drive the methods' rapid evolution.

3.1.3. Real-time optimization and automation

The complexity of marketing environment has long surpassed the threshold of human analysts' intuitive understanding and manual capacities. Fine-grained segmentation and frequent customer interactions further make it necessary to remove human agents from the critical path. While determining targeting strategies for several segments is humanly possible, for hundreds of precisely defined microsegments, automation is a must-have. Frequent interactions also call for real-time responses. When mobile tracking detects an inbound consumer, for example, a window of only minutes exists to deliver a promotional offer. Automation and real-time optimization are becoming the modus operandi of marketing. and machine learning methods are the go-to solutions.

3.1.4. Customer-journey focus

Firms are increasingly taking a holistic view of the entire customer journey. Digital technology has made it feasible to collect fine-grained customer touchpoint information, which transforms the view of an aggregate-level purchase-funnel from the firm's perspective, to the view of continuous decision journeys with feedback loops from the perspective of individual consumers. Piecing together the picture of a consumer's entire journey is immensely valuable to firms, allowing them to monitor and guide the consumer, delivering the right information, service, and promotion at the right stage and context. In addition to nudging consumers through specific hurdles, firms can plan out a longer-term strategy to cover the entire journey. For example, to just push a consumer through the awareness hurdle, a handful of exposures to any of the firm's ad would suffice. Looking forward through the entire journey, though, more exposures of specific types may be better to create longer lasting affinity down the road. PGM, deep learning, and reinforcement learning methods can help firms master the entire decision journey: Before customers even emerge, a firm monitors user-generated content (UGC) platforms to anticipate or shape the demand. When a lead emerges, a holistic view of the consumer helps map out a program to accompany her through the purchase journey. When executing the program, the firm monitors the actual touchpoints across devices and channels, assessing and adjusting tactics to create conversion. Post conversion, the interaction continues to ensure good customer experience, positive word-of-mouth, and optimal life-time value. Machine learning methods play an instrumental role in managing consumer decision journeys, which would bring marketing effectiveness to a new level.

3.2. Marketing practices

3.2.1. Marketing mix

Marketing-mix decisions are increasingly made quantitatively instead of qualitatively. Pricing decisions are routinely made using dynamic quantitative models, so are assortment, channel, and location decisions. Meanwhile, advertising is increasingly digitized and personalized. Machine learning methods underlie many programmatic advertising tools and services, targeting users based on profile and behavioral history, with real-time bidding decisions made at millisecond timescale. Even product and advertising creative designs, traditionally in the qualitative domain, have turned to algorithms for assistance, and these transformations are accelerating.

3.2.2. Customer engagement

With firms focusing on customer decision journeys, intelligent agents assist customer engagement along the way to improve experience. When consumers browse online, targeted ads delivered through bidding machines and customized webpages created

[11] For example, research shows that human brains can process entire images the eye sees in 13 milliseconds. ("In the blink of an eye," http://news.mit.edu/2014/in-the-blink-of-an-eye-0116)

using website morphing algorithms help provide information and generate interest. When consumers visit physical stores, sales applications provide analytics to help agents deliver personalized assistance, and AI tools such as augmented reality and virtual fitting rooms make the shopping experience efficient and enjoyable. After purchase, automated follow-ups to offer usage tips keep customers engaged and loyal. Other applications abound. Machine learning algorithms extract consumer preferences from massive online data, and help create engaging text and images to attract attention. Computer-generated virtual influencers develop brand followings and promote products. Powered by sophisticated AI engines in the cloud, virtual assistants such as Amazon Alexa respond to consumers' voice inquiries to supply information or make purchase. Chatbots enabled by speech recognition and natural language processing algorithms are increasingly taking over the handling of pre- and post-purchase inquires. Along the entire customer journey, AI-driven innovations are rapidly reshaping engagement practices.

3.2.3. Search

Internet search engine is where many customer journeys begin. Originally using the PageRank algorithm, Google now processes a significant number of searches using the deep learning-based RankBrain algorithm, which helps improve the relevance and robustness of search results.[12] For marketers, search engine optimization is also conducted using machine learning-based tools such as Can I Rank or Alli AI. While keyword has been the dominant form of online search, machine learning methods are making searches based on other content types within reach. With voice recognition, natural language processing, and text-to-speech capabilities, conversation engines such as Dialogflow can handle voice search effectively. Meanwhile, tools such as Synthetic's Style Intelligence Agent make visual search possible.

3.2.4. Recommendation

Recommending the right products to the interested consumers can significantly improve marketing performance. Initial recommender systems use machine learning algorithms such as item-based or content-based collaborative filtering. Their accuracy and robustness have steadily improved with the incorporation of newer algorithms, e.g. matrix factorization. In recent years, deep neural networks and embedding methods have been leveraged to further enhance the performance. Analyzing the information of millions of consumers and products to assess relevance, recommender systems are now an essential component in marketing, effectively matching products and consumers throughout digital channels.

3.2.5. Attribution

With myriad marketing channels and touchpoint interactions comes the heightened need to attribute the final conversion to the responsible contacts. Traditionally, attributions are performed using simple rules such as first-touch or last-touch which, while easy to understand, does not accurately reflect the true contributions of the touchpoints. Firms are now moving to model-based attribution. From logistic regression to more sophisticated classification models and PGMs, a wide range of machine learning methods are being evaluated and adopted to generate accurate performance feedback and help improve channel design and allocation.

4. Review of machine learning literature in marketing

Academic marketing research started leveraging machine learning methods more than a decade ago, and the interest has increased markedly in recent years. A quick summary of this literature is presented in Table 4, which organizes the extant studies from three distinct perspectives: method, data, and usage. In this section, we briefly discuss these studies from the method perspective.

4.1. SVM

One of the first machine learning methods introduced to marketing is SVM. Cui & Curry (2005) compared SVM with multinomial logit models, and showed that SVM in general predicts better. They argued that while multinomial logit models can help draw insights, in large-scale automatic settings SVM is a better fit. Evgeniou, Boussios, & Zacharia (2005) developed robust conjoint estimation methods that are almost equivalent to SVM. Similar to Cui & Curry (2005), they also showed that the methods outperformed traditional marketing methods such as logistic regression and hierarchical Bayes. Huang & Luo (2016) developed a multi-step method which uses a fuzzy SVM (Lin & Wang, 2002) active learning algorithm to adaptively refine preference estimates for consumer preference elicitation. They showed that the method worked well for product categories with as many as 100 attribute levels, a scale that would overwhelm traditional methods.

4.2. Traditional text-mining

Lee and Bradlow (2011) performed market structure analysis and visualization using Epinions reviews. They extracted phrases and performed clustering analysis where distances were measured using cosine similarity. Their approach created informative brand maps. Netzer, Feldman, Goldenberg, and Fresko (2012) adopted a text-mining process including downloading, cleaning,

[12] https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/. Accessed in November 2019.

Table 4
Machine learning literature in marketing.

Aspect	Category	Articles
Method	SVM	Cui & Curry (2005), Evgeniou, Boussios, & Zacharia (2005), Huang & Luo (2016), Malik, Singh, Lee, and Srinivasan (2019)
	Topic models	Tirunillai & Tellis (2014), Büschken & Allenby (2016), Jacobs, Donkers, & Fok (2016), Trusov, Ma, & Jamal (2016), Ansari, Li, & Zhang (2018), Liu & Toubia (2018), Li & Ma (2019)
	Deep learning	Zhang, Lee, et al. (2017), Liu, Dzyabura, & Mizik (2018), Liu, Lee, & Srinivasan (2018), Chakraborty, Kim, & Sudhir (2019), Dzyabura, El Kihal, Hauser, & Ibragimov (2019), Li, Shi, & Wang (2019), Malik, Singh, Lee, and Srinivasan (2019), Zhang & Luo (2019)
	Tree ensembles	Rafieian & Yoganarasimhan (2018), Yoganarasimhan (2018), Zhang & Luo (2019)
	Causal forest	Feng, Zhang, and Rao (2018), Guo, Sriram, & Manchanda (2018)
	Network embedding	Ma, Sun, & Zhang (2019), Yang, Zhang, & Kannan (2019)
	Active learning	Dzyabura & Hauser (2011), Huang & Luo (2016)
	Reinforcement learning	Hauser, Urban, Liberali, and Braun (2009), Schwartz, Bradlow, & Fader (2017), Misra, Schwartz, & Abernethy (2018)
Data	Text	Lee & Bradlow (2011), Netzer, Feldman, Goldenberg, & Fresko (2012), Tirunillai & Tellis (2014), Büschken & Allenby (2016), Toubia & Netzer (2017), Ansari, Li, & Zhang (2018), Liu & Toubia (2018), Liu, Lee, & Srinivasan (2018), Chakraborty, Kim, & Sudhir (2019), Hartmann, Huppertz, et al. (2019), Li & Ma (2019), Vermeer, Araujo, Bernritter, & van Noort (2019), Zhang & Luo (2019)
	Image	Zhang, Lee, et al. (2017), Klostermann, Plumeyer, Böger, & Decker (2018), Liu et al. (2018a), Dzyabura, El Kihal, Hauser, & Ibragimov (2019), Hartmann, Heitmann, et al. (2019), Malik, Singh, Lee, and Srinivasan (2019), Zhang and Luo (2019)
	Audio	N/A
	Video	Kawaf (2018), Li et al. (2019)
	Consumer tracking	Trusov, Ma, & Jamal (2016), Rafieian & Yoganarasimhan (2018), Yoganarasimhan (2018), Kakatkar & Spann (2019)
	Network	Ma, Sun, & Zhang (2019), Yang, Zhang, & Kannan (2019)
Usage	Prediction	Cui & Curry (2005), Huang & Luo (2016), Jacobs et al. (2016), Dzyabura, El Kihal, Hauser, & Ibragimov (2019), Hartmann, Huppertz, et al. (2019), Vermeer et al. (2019)
	Feature extraction	Lee & Bradlow (2011), Netzer, Feldman, Goldenberg, & Fresko (2012), Tirunillai & Tellis (2014), Zhang, Lee, et al. (2017), Liu & Toubia (2018), Liu, Dzyabura, & Mizik (2018), Liu, Lee, & Srinivasan (2018), Dzyabura, El Kihal, Hauser, & Ibragimov (2019), Hartmann, Heitmann, et al. (2019), Malik, Singh, Lee, & Srinivasan (2019), Zhang & Luo (2019)
	Descriptive interpretation	Lee & Bradlow (2011), Netzer, Feldman, Goldenberg, & Fresko (2012), Tirunillai & Tellis (2014), Trusov, Ma, & Jamal (2016), Chakraborty, Kim, & Sudhir (2019), Li & Ma (2019), Ma, Sun, & Zhang (2019), Yang, Zhang, & Kannan (2019)
	Causal interpretation	Guo, Sriram, & Manchanda (2018), Feng, Zhang, & Rao (2018)
	Prescriptive analysis	Hauser, Urban, Liberali, & Braun (2009), Schwartz, Bradlow, & Fader (2017), Misra, Schwartz, & Abernethy (2018), Yoganarasimhan (2018)
	Optimization or estimation	Evgeniou, Boussios, & Zacharia (2005), Evgeniou, Pontil, & Toubia (2007), Hauser, Toubia, Evgeniou, Befurt, & Dzyabura (2010), Ansari, Li, & Zhang (2018), Chiong & Shum (2019)

information extraction, chunking, and identification of semantic relationships. They used conditional random field, a type of PGM, to extract entities, and used co-occurrence to form semantic networks and market structure perceptual maps. Using data on sedan cars and diabetes drugs, they showed that the approach is effective in extracting insights from UGC. Toubia and Netzer (2017) investigated the idea generation process. They developed semantic networks based on word stem co-occurrence, and showed that ideas with semantic subnetworks that have more prototypical edge weight distributions are judged as more creative. Hartmann, Huppertz, Schamp, and Heitmann (2019) compared five lexicon-based and five machine learning methods on automatic text classification, and found that RF and NB perform the best on uncovering human intuition. Vermeer, Araujo, Bernritter, and van Noort (2019) analyzed electronic word-of-mouth data on social media. They used six supervised machine learning methods to identify response-worthy messages. Other text-mining studies used topic models, which we discuss separately below given the method's importance and applicability to non-text data.

4.3. Topic models

Tirunillai and Tellis (2014) introduced LDA to the marketing literature. Analyzing product reviews using LDA, they uncovered important quality dimensions as topics, and demonstrated the informativeness and external validity of the extracted dimensions. Extending standard LDA, Büschken and Allenby (2016) developed a sentence-constrained LDA model, restricting words inside a same sentence to come from one topic, and showed that the model extracts topics that are more distinguished and coherent than those recovered using standard LDA. Liu and Toubia (2018) developed a hierarchically dual LDA model to identify topics from both consumer search queries and webpages, and showed that the topics from those two sources are related. Ansari, Li, and Zhang (2018) developed a covariate-guided heterogeneous supervised topic model to characterize products. Product tags

were incorporated using the topic model which were associated with product covariates and user ratings. Topic models have been applied not only to text, but also in other marketing settings where similar semantic structures exist. Jacobs et al. (2016) incorporated LDA to help predict purchases. Using data from an online retailer, they showed that the LDA model outperforms both collaborative filtering and discrete choice models in prediction, and is scalable to large product assortments. Trusov et al. (2016) adapted the correlated topic model (CTM), an extension to LDA, for online user profiling using consumer clickstream data. They extended the original CTM to incorporate visitation intensity, covariates, and dynamic evolution. Their analysis uncovered easy-to-interpret consumer behavioral profiles, and helped assess the information content of data from search engines and advertising networks. Li and Ma (2019) integrated CTM into a dynamic model of consumers' channel visits and conversions. They showed that incorporating search keywords using topic models improves the assessment of consumers' latent purchase stages.

4.4. Deep learning

Deep learning has been the most frequently used machine learning method in recent marketing studies, typically for analyzing text and image data. Liu, Lee, and Srinivasan (2018) analyzed consumer reviews and showed that aesthetics and price content affect conversions. They developed a full deep learning model to predict conversions and a partial deep learning model to extract features, and they used the derivative-based method developed in Simonyan, Vedaldi, and Zisserman (2013) to evaluate feature importance. Chakraborty, Kim, and Sudhir (2019) developed a hybrid CNN-LSTM model to extract attribute level sentiment from text data, and corrected for attribute self-selection of the extracted scores. Using Yelp reviews, they showed that the model does well on hard sentiment classification problems. Zhang, Lee, Singh, and Srinivasan (2017) investigated the impact of images on property demand at Airbnb. They used CNN to label images as high or low quality, and showed that verified high quality photos significantly increase property demand. Liu, Dzyabura, and Mizik (2018) presented a "visual listening in" approach. Using Flickr data, they trained classifiers using both SVM which takes pre-extracted features as input and CNN which works directly on the raw images. Analyzing an Instagram dataset, they showed that brand portrayals generated from consumer-supplied images are consistent with firm-generated images as well as the result of a national survey. Zhang and Luo (2019) analyzed Yelp photos and reviews. They found that photos are more predictive of restaurant survival than are reviews, and that information content and helpful votes are more important than photographic attributes. They used the Clarifai API[13] to identify objects from food photos, and used text-based CNN to extract quality dimensions from reviews. Dzyabura et al. (2019) investigated product returns, and showed that features extracted from product images using CNN significantly improve the prediction accuracy of return rate. Hartmann, Heitmann, et al. (2019) used CNN to classify Twitter and Instagram images, and showed that brand selfies lead to high levels of consumer brand engagement, a finding they corroborated using lab experiment. They also created gradient-weighted class activation maps (Selvaraju et al., 2017) to provide post-hoc interpretation of important image aspects. Malik, Singh, Lee, and Srinivasan (2019) investigated preference-based and belief-based attractive bias using pictures of MBA graduates, and found significant bias over a 15-year career period mainly attributed to preference-bias. They used the CNN implementation of the OpenFace project (Amos, Ludwiczuk, & Satyanarayanan, 2016) to extract image features and SVM to predict the person's attractiveness. An innovative aspect is they used the conditional adversarial autoencoder developed in Zhang, Song, and Qi (2017), a deep generative model, to age progress the pictures. Li, Shi, and Wang (2019) proposed two measures for video mining that can be automatically extracted from videos using CNN. Using data from a crowdfunding site, they showed that the proposed measures have explanatory power on funding outcomes.

4.5. Tree ensembles

Yoganarasimhan (2018) investigated search personalization using a machine learning framework with three components for generating features, search ranking, and scalability and efficiency. The search ranking component used the LambdaMART method (Wu, Burges, Svore, & Gao, 2010) which combines the lambda ranking algorithm with multiple additive regression trees. Analyzing a large dataset from Yandex, they showed that personalization improved the clicks to the top position by 3.5%, and that the return to personalization varied with user history and query type. Rafieian and Yoganarasimhan (2018) used gradient-boosted trees to predict mobile in-app ad click-throughs, and an analytical framework to conduct data-sharing counterfactuals. They showed that the method improves targeting ability by 17.95%, and that the gains mainly stem from behavioral rather than contextual information. Dzyabura et al. (2019) used the gradient-boosted tree LightGBM (Ke et al., 2017) to predict product returns using features extracted through deep learning.

4.6. Causal forest

Recent methodological advancements have made it possible to use machine learning methods for causal research. Athey and Imbens (2016) developed the causal-tree method for heterogeneous causal effects, and Wager and Athey (2018) developed the causal forest method to evaluate heterogeneous treatment effects. Guo et al. (2018) applied the causal forest method to investigate how information disclosure affects pharmaceutical companies' payments to physicians. Using a quasi-experimental difference-in-differences approach, they showed that while payments on average declined after information disclosure laws came into effect,

[13] https://docs.clarifai.com/

there was considerable heterogeneity in the effect. Feng, Zhang, and Rao (2018) analyzed online home rental market. Combing the causal forest method with propensity score matching, they showed that rental units with different values vary systematically in their choice of cheap-talk content. Furthermore, they showed that providing either objective or subjective cheap-talk communications can increase consumer search, although only the objective content affects the final renting decision.

4.7. Network embedding

Many marketing contexts can be represented as networks, e.g., consumer social networks or consumer-product networks. Analyzing large-scale networks is challenging due to their high dimensionality. Ma et al. (2019) used network embedding to analyze consumers' social curation on Pinterest. They first constructed a heterogeneous information network where users, images, and annotation words are represented as nodes, and curation actions are encoded in edges. They then used the heterogeneous embedding method *metapath2vec* (Dong et al., 2017) to map the nodes to lower-dimensional vectors, while preserving the network's structural and semantic information. They showed that the embedding vectors reveal consumer and image groups with distinct characteristics, and generate superior predictive performance. Yang, Zhang, and Kannan (2019) formulated a user-brand network from user engagement data on Facebook. They then used a deep autoencoder to perform embedding. Their analysis showed a market structure of brands that is more fluid and overlapping than what standard industry classification would suggest, and identified key competitors and complementors of brands.

4.8. Active and reinforcement learnings

Dzyabura and Hauser (2011) developed an active learning method to select questions adaptively to maximize the information about consumers' heuristic decision rules. Huang and Luo (2016) also used active learning to adaptively select questions for consumer preference elicitation. Hauser, Urban, Liberali, and Braun (2009) investigated website morphing, automatically changing the look-and-feel of a website to match the visitor's cognitive styles. They modeled morphing as a MAB problem, and showed that a close-to-optimal performance can be achieved on improving consumers' purchase intentions. Along a similar line, Schwartz, Bradlow, and Fader (2017) conducted field experiments for online display advertising by formulating the advertiser's problem as a MAB problem, and showed that the reinforcement learning policy achieved an 8% improvement in customer acquisition. Misra, Schwartz, and Abernethy (2018) developed a dynamic pricing experimentation policy which included microeconomic choice theory. They proved analytically that the proposed policy is asymptotically optimal for weakly downward sloping demand curve, and demonstrated through simulation that the policy can significantly improve profit.

While a number of machine learning methods have been adopted, many others, especially in the large families of PGM, representation learning, and reinforcement learning, are yet to be fully leveraged in marketing research. Meanwhile, Table 4 also organizes papers based on the types of data analyzed, which shows that more studies have analyzed text and image data than have used audio, video, consumer tracking, or network data. Similarly, Table 4 lists papers based on the usage of machine learning methods, which shows that more studies have used machine learning methods for prediction and feature extraction than for descriptive, causal, and prescriptive analysis. Aside from these noticeable gaps, the literature can also be expanded and deepened in several important ways. Considering these, we next proceed to discuss a comprehensive research agenda.

5. Harnessing the power of machine learning methods – a research agenda

Given their unique strengths, the use of machine learning methods in academic marketing research is expected to accelerate. Since this trend is still at an early stage, a consensus does not yet exist on the right approach of adopting these methods, and a holistic view is lacking. In this section, we articulate a multi-faceted research agenda. Our discussion begins with a conceptual framework, followed by a detailed discussion from each of the five key aspects in the framework. The research opportunities from each aspect are also summarized in Table 5.

5.1. A conceptual framework

We propose a conceptual framework for the role of machine learning methods in marketing research, which is depicted in Fig. 3. Taking an empirical researcher's perspective, the incorporation of machine learning methods into marketing research is a multi-faceted task, with focuses called for on five key aspects: *method*, *data*, *usage*, *issue*, and *theory*.

At the core of the conceptual framework is the large collection of machine learning *methods*. Quantitative marketing research has traditionally used statistical or econometric models. While these models often have similar mathematical foundations as machine learning methods, they differ in guiding philosophy and focus. Used appropriately, machine learning methods can complement econometric approaches to expand the research frontier. Researchers are encouraged to develop a broad and in-depth understanding of machine learning methods, to introduce more of them to marketing research, and to demonstrate their compelling values. For higher impact, machine learning methods should also take on a more central role in research, e.g. to model consumer decisions, in contrast to peripheral applications such as feature extraction. Meanwhile, significant effort is called for to extend the methods to make them better suited for research in marketing and other social science fields, and to also delineate the boundary of their capacities. These considerations from the method's perspective are discussed in Section 5.2.

Table 5
Call for research using machine learning methods.

Aspect	Key research opportunities using machine learning methods
Method	• Introduce more methods and demonstrate their values to marketing research (PGM, representation learning, reinforcement learning, deep generative models, ensemble methods, etc.) • Move the use of machine learning methods from periphery to center, e.g. to model consumer decisions • Extend machine learning methods in their own right, e.g. improving interpretability or enabling statistical inference • Delineate the boundary of the methods' capacities
Data	• Use the methods to analyze unstructured data, especially audio and video • Use the methods to draw insights from consumer tracking and network data • Use the methods to analyze data of hybrid format, potentially from multiple sources, preferably in integrated models
Usage	• Expand and refine the methods' usage for prediction and feature extraction • Adapt the methods for correlational, causal and prescriptive research, improving transparency and theoretical connection • Evaluate the methods' potential for correlational, causal, and prescriptive research
Issue	• Use the methods to help map out customer purchase journey, especially in early stages • Use the methods to develop automated online decision-support capabilities of various marketing functions • Use the methods and platform data for market structure analysis
Theory	• Inject human insights into the use of machine learning methods • Balance between theory-driven and data-driven perspectives • Investigate theoretical connections and implications of machine learning methods • Investigate theoretical implications of firms' adoption of AI tools

Meanwhile, machine learning methods must be effectively connected to the other key components of empirical research: data, usage, issue, and theory. First, interesting *data* are frequently what enable interesting research. Unstructured, consumer tracking, and network data are rapidly expanding in scale and scope. While often challenging for econometric models, machine learning methods exist to process them efficiently. An agenda on using machine learning methods to extract insights from rich data is presented in Section 5.3. Second, the *usage* of methods in marketing research is diverse. A method can be used for prediction, feature extraction, descriptive interpretation, causal interpretation, prescriptive analysis, or for model estimation at a technical level, among others. The different usages call for different levels of model transparency and connection to marketing theories. While machine learning methods can be used for most of these purposes, they pose different challenges for researchers to overcome, which we discuss in Section 5.4. Third, research seeks to address important substantive *issues*. The increasingly complex marketing environment and the rich interaction between firms and consumers lead to many research issues that are simultaneously interesting and challenging, which machine learning methods are uniquely positioned to help investigate. In Section 5.5, we discuss the agenda for addressing the substantive issues. Finally, impactful marketing research should be well grounded in *theory*, in order to produce generalizable findings. A close connection between machine learning methods and theories of consumer or firm behaviors, however, is a major challenge confronting researchers. In Section 5.6, we discuss unifying the methods and marketing

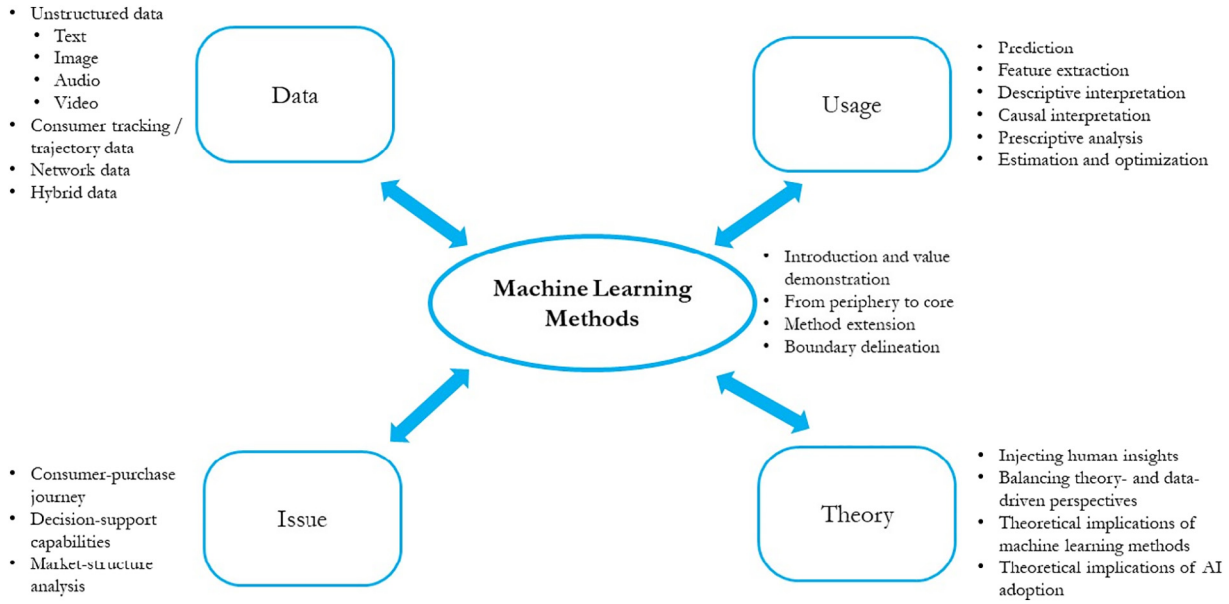

Fig. 3. Leveraging machine learning in marketing research – a conceptual framework.

theories. Through deepening and expanding the methods themselves, and through closely connecting to the other key components of empirical research, machine learning methods can make significant contributions to the marketing literature.

5.2. Enhancing the machine learning method toolset

At the core of the conceptual framework in Fig. 3 are the machine learning methods. Researchers should have deep understandings of the pertinent methods, and adapt, extend, and apply them as appropriate. From the method's perspective, we discuss the following research agenda:

5.2.1. Introducing additional methods

The collection of machine learning methods is vast. While many methods have been introduced to marketing, a lot more can and should be. The rich family of PGM methods can model complex interconnections of large numbers of random variables and produce statistically interpretable results; unsupervised representation learning methods are effective at information extraction from complex data; reinforcement learning methods enable the modeling of dynamic optimization behaviors in complex environments. Deep generative models and state-of-the-art ensemble methods are other valuable options. To expand and strengthen their usage, we call for researchers to introduce these and many other machine learning methods to marketing research.

5.2.2. Compelling value demonstration

When introducing new methods, researchers should strive to demonstrate their value in marketing research. That a machine learning method can address important issues or challenges that traditional quantitative models cannot address, for example, makes a compelling case for its adoption. Alternatively, if researchers can show that a machine learning method performs significantly better than existing quantitative marketing models on certain tasks, or complements existing models to improve their performance, that will also demonstrate the method's value. A method should not be introduced just for the sake of introducing it. A method also should not be evaluated based on its technical sophistication. A cutting-edge method that is unrelated to marketing, in our belief, is less valuable than a method that is comparatively straightforward but can be used to solve important marketing problems.

5.2.3. From periphery to center

Importantly, we call for researchers to move machine learning methods from playing a supporting role in the periphery, to addressing the central research tasks. Extant studies frequently use machine learning methods to extract variables from unstructured data, which are then analyzed using econometrics models. In contrast, fewer have used machine learning methods to directly model consumer decisions. While their application to data pre-processing is clearly valuable, the methods should also be appraised of their potentials to play a more central role. Incorporating machine learning methods into a joint statistical model is a reasonable first step (Li & Ma, 2019). Furthermore, of particular interest is the potential of using machine learning methods to model heterogeneous and dynamic consumer decisions. Accounting for observed and unobserved heterogeneity as well as dynamics is essential to analyzing consumer decisions. If machine learning methods can capture these key components, their use in marketing research will greatly expand. Machine learning methods with well-established statistical foundations have much promise for handling such tasks, and any demonstration or the adaptation or extension of the methods for such use will significantly contribute to the marketing literature. In addition to modeling consumer decisions, machine learning methods may play a central role in other ways, e.g. in designing automated decision-support systems, which we also call for.

5.2.4. Extending the methods

We also call for researchers to not just adapt existing machine learning methods, but to extend them in their own right. Properties desired for marketing research are often also needed in other social science fields. For example, interpretability is a universally desirable characteristic. While strides have been made to improve the interpretability of machine learning methods, the interpretation tends to be on a post-hoc basis (e.g. Selvaraju et al., 2017). Ex ante interpretability on model structure and connections among variables is preferred, especially for correlational, causal, and prescriptive analyses, and improvements in this respect would constitute important methodological contributions. As another example, any advancement that allows the methods to handle dynamic consumer decisions will be valuable to not just marketing but social science in general, so will any advancement that enables statistical inference. Methods such as PGM, reinforcement learning, and deep generative models, among others, are promising candidates for these advancements. We call for researchers to take on the challenge of pursuing methodological contributions at this fundamental level.

5.2.5. Delineating the boundary

While much excitement exists about machine learning methods' potential, it is equally important to know what these methods cannot do. As more methods are introduced to handle more tasks, and even more attempts are made to do so, researchers would begin to reach the boundary of these methods' capacity, and knowledge of such boundaries is valuable. For example, while it would be great to use machine learning methods to analyze consumer dynamics, any convincing demonstration that the methods are actually not able to capture such dynamics would also contribute to the literature. As the literature grows and methods mature, we also call for systematic comparisons between machine learning methods and quantitative marketing models, to obtain in-depth understandings of their comparative strengths on handling various tasks.

5.3. Generating rich insights from rich data

Machine learning methods' ability to process unstructured data is well recognized. The rich digital footprint of consumers presents great research opportunities.

5.3.1. Unstructured media data

In marketing, a reasonably large literature now exists on analyzing textual data (Ansari, Li, and Zhang, 2018; Büschken & Allenby, 2016; Chakraborty, Kim, and Sudhir, 2019; Hartmann, Huppertz, et al., 2019; Lee & Bradlow, 2011; Li & Ma, 2019; Liu, Lee, & Srinivasan, 2018; Liu & Toubia, 2018; Netzer, Feldman, Goldenberg, and Fresko, 2012; Tirunillai & Tellis, 2014; Toubia & Netzer, 2017; Zhang & Luo, 2019). Meanwhile, that on analyzing image data is quickly growing (Zhang, Lee, et al., 2017, Klostermann, Plumeyer, Böger, & Decker, 2018, Liu, Dzyabura, & Mizik, 2018, Dzyabura, El Kihal, Hauser, & Ibragimov, 2019, Hartmann, Heitmann, et al., 2019, Malik, Singh, Lee, and Srinivasan, 2019, Zhang & Luo, 2019). Fewer studies comparatively have used audio or video data (Kawaf, 2018; Li, Shi, and Wang, 2019). We call for continued effort to extract insights from unstructured data, especially audios or videos. Topic models and word embedding methods can be applied to text data; RNNs and 1-dimensional CNNs can process audio data; 2-dimensional CNNs are the standard approach for analyzing image data; 3-dimensional CNNs are applicable to videos. Opportunities still abound to use machine learning methods to extract insights about consumers and products from rich media data.

5.3.2. Consumer tracking data

Technology has enabled expansive tracking of consumer activities. The now commonplace online clickstream data keep logs of purchase, search, browsing, digital media consumption, and other web surfing actions. Mobile devices further expand the reach of activity tracking to the offline word, recording consumers' physical locations and store visits. The proliferation of Internet-of-things (IoT) devices and the increasing adoption of consumer wearables are bringing in even richer consumer tracking data, e.g. detailed in-store activities or 24/7 biometrics information (Ng & Wakenshaw, 2017). Marketing studies have just started to leverage such fine-grained tracking data (e.g. Rafieian & Yoganarasimhan, 2018; Trusov et al., 2016; Yoganarasimhan, 2018). We call for significantly more work on this front, for which PGMs and deep neural networks, among others, are pertinent methods.

5.3.3. Network data

We live in an increasingly connected world. Aside from social networks of people, other networks, of products, brands, store locations, etc., are also becoming prevalent. Even for traditional data, a network view may be beneficial, e.g. semantic networks generated through word co-occurrence may shed light on market structure (Netzer et al., 2012). Instead of treating consumers or products as standalone units, network data encode the connections among a large set of entities and present a holistic view of the marketing environment. However, analyzing large networks is also challenging due to its high dimensionality. Existing studies tend to use variables that summarize local characteristics (e.g. degree centrality), while leaving out the global structural information. Machine learning methods including network embedding, PGM, and graph deep learning, show great promise in tackling this challenge (e.g. Ma et al., 2019), and we call for adopting these methods to analyze large scale networks.

5.3.4. Integrated hybrid data

Machine learning methods can handle multi-typed data with ease. Using such methods, all data types are usually first converted into vectors, making it easy to mesh them together.[14] This capacity is important since consumers are increasingly exposed to information of multiple types. For example, an apartment posting on Airbnb has both textual description and images; an online user can read reviews and browse product specifications; a sports game viewer may tweet out comments on Twitter while watching the TV program. An in-depth understanding of consumer preferences and behaviors in these situations calls for combining data types, potentially from multiple sources. While this has been done in several studies (Dzyabura et al., 2019; Hartmann, Heitmann, et al., 2019; Klostermann et al., 2018; Zhang & Luo, 2019), many more opportunities exist. Meanwhile, extant studies often process different data using separate components, while integrated handling is preferable. Furthermore, the data combination available to practitioners and researchers also raises interesting questions. For example, Trusov et al. (2016) assessed the information content of the data available to different firms, while Kakatkar and Spann (2019) developed an approach to handle anonymized and fragmented tracking data. Researchers are encouraged to investigate similarly interesting questions related to integrated hybrid data.

5.4. Broadening and extending usages

A method can serve a range of purposes. It can handle a prediction task, extract features of interest from data, provide descriptive insights, perform causal inference or prescriptive analysis, or be used in a technical capacity for optimization. For each usage, two considerations are important to researchers: what level of model transparency is required, and how close the method should be connected to marketing theories. The more transparent and closer the theoretical connection, the higher impact the research has. So, too, is the challenge. In Fig. 4, we present a schematic view along these two dimensions. When used for *prediction*, a

[14] For example, text data are often represented using "one-hot encoding," where each word occurrence in a document is represented using a vector, in which each dimension corresponds to a word in the vocabulary.

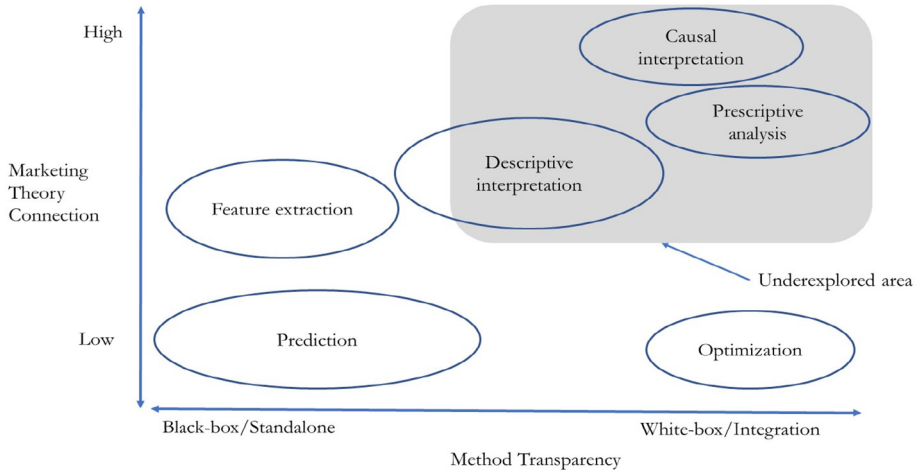

Fig. 4. The usage of machine learning methods.

black box with minimal theoretical connection may suffice, as researchers only need to ensure predictive accuracy. For *feature extraction*, a black box is fine, although moderate theoretical connection is needed to understand the features. For *descriptive interpretation*, the method needs a moderate level of model transparency and theoretical connection, especially when drawing correlational insights. For *causal interpretation* and *prescriptive analysis*, the method needs to be used as a white box and closely connected to theories, as researchers need to establish generalizable causal linkages between variables and have confidence in prescribed actions. For *estimation and optimization*, the method needs to have transparency as a white box, although connection to marketing theories is less important, since the method is used in a technical capacity.

While these usages are all compatible with machine learning methods, they present different levels of challenges and opportunities. Using the methods for prediction and feature extraction is comparatively straightforward, and extant literature has mostly adopted these usages. In contrast, it is more challenging to use machine learning methods for descriptive, causal, and prescriptive studies, and we call for more research in these underexplored areas. We discuss each usage in detail below.

5.4.1. Prediction

In accordance with their traditional focus, machine learning methods have been used for prediction in marketing research (Cui & Curry, 2005; Dzyabura et al., 2019; Hartmann, Huppertz, et al., 2019; Huang & Luo, 2016). The prediction task may be the focus of a study, or a component of a larger framework (e.g. Huang & Luo, 2016). While researchers typically prefer interpretations, the importance of prediction is elevated by the increase in scale and complexity of real-world marketing tasks. For example, a system that can predict a consumer's next store or website visit can make a strong case for methodological contribution, so can a recommender system that accurately predicts the product each consumer is interested in. Such opportunities abound given the increasing digitization and mass customization, which researchers are encouraged to pursue.

From a different angle, we call for researchers to explore using prediction in causal research. For example, propensity score matching (PSM) is often used in causal studies using observational data, where the lack of random assignment necessitates the control for selection. Using PSM, researchers first model the probability, or the propensity score, of receiving the treatment for each unit in the dataset, and then match each treated unit with an untreated unit with similar propensity score to assess the treatment effect. The first step to generate propensity scores is essentially a prediction component. While binary logit is often used in this step, a more sophisticated machine learning method may generate better predictions and thus better matchings. Similarly, machine learning methods may also improve the performance of instrumental variable regressions (Mullainathan & Spiess, 2017).

5.4.2. Feature extraction

Extracting features from unstructured data has been a popular usage of machine learning methods in marketing (Lee & Bradlow, 2011, Netzer et al., 2012, Tirunillai & Tellis, 2014, Liu & Toubia, 2018, Liu, Dzyabura, & Mizik, 2018, Liu, Lee, & Srinivasan, 2018, Zhang, Lee, et al., 2017, Dzyabura et al., 2019, Hartmann, Heitmann, et al., 2019, Malik et al., 2019, Zhang & Luo, 2019). Ample opportunities remain to extract additional and richer features, especially from audio, video, network, or consumer tracking data. Using supervised learning, researchers can manually construct a training sample of the input data and their associated features, train a supervised learning model, and then apply the model to the larger dataset to extract features. Unsupervised representation learning is also an alternative, e.g. using word embedding methods or autoencoders to generate low-dimensional representations of input data.

Extracted features can be used to draw insights directly or as input for subsequent analysis. Methods with well-established statistical foundations, e.g. PGM, can be integrated into joint statistical models (e.g. Trusov et al., 2016). Alternatively, features can be extracted in a pre-processing step, and used as input variables for subsequent statistical analysis. When used this way, however,

one note of caution is in order. Suppose a researcher trains a classifier to label documents in a dataset, and then uses the labels as an independent variable in linear regression. While commonly done, this approach introduces *measurement error*: the predicted document label, even if reasonably accurate when generated using the powerful machine learning method, is still a noisy measure of the true document label. Measurement errors of independent variables may cause attenuation bias in coefficient estimates, and appropriate corrections should be performed (e.g. Frost & Thompson, 2000).

5.4.3. Descriptive interpretation

A prime example of using machine learning methods for descriptive interpretation is topic models, which has been used to analyze both text and other marketing data (Jacobs et al., 2016; Tirunillai & Tellis, 2014; Trusov et al., 2016). Unsupervised representation learning methods such as network embedding and autoencoders can also generate meaningful descriptions from complex data, and are beginning to be introduced to marketing research (e.g. Ma et al., 2019). We call for continued effort to generate insightful descriptions of the increasingly complex marketing environments using the right methods.

The sought-after descriptions are often not about a single characteristic, but about relationships between variables. Using machine learning methods to draw such *correlational* insights, however, is challenging. One may consider training a supervised learning model, using the factors as the input variables and outcomes as the output. However, unlike linear econometric models, machine learning methods typically yield non-linear and non-monotonic functions, which may not even have known functional forms.[15] Further adding to the challenge, the input themselves can be heavily engineered features instead of the original variables.

Question remains open on how best to leverage machine learning methods for correlational description. Variable importance (Breiman, 2001), i.e., assessing how important an input variable is on predicting the output, has been used in marketing studies (e.g. Dzyabura et al., 2019). However, variable importance shows only whether a factor is related to the outcome, but not how the latter changes in response to the former. A second option is to first train a machine learning model and then provide post-hoc description of the prediction function. Recent machine learning research shows this is feasible (e.g. Selvaraju et al., 2017). Such post-hoc interpretations, though, also do not fully describe variable correlations. A third option is to use the methods in a way that preserves the interpretability, instead of fully utilizing their predictive capacity (e.g. Trusov et al., 2016). Methods with statistical foundations, e.g. PGM, have good potentials to be used this way. However, it is unclear whether this is feasible for other methods such as deep neural networks or ensemble methods. Given its importance in marketing, we call for research to develop effective ways for generating correlational descriptions using machine learning methods.

5.4.4. Causal interpretation

Recently developed causal trees and causal forests have made it possible to use machine learning methods for causal interpretation (Athey & Imbens, 2016; Wager & Athey, 2018). Rather than measure treatment effects at the population level, causal forest leverages the flexible structure of trees to measure heterogeneous treatment effects according to individual units' observed characteristics, and has been used in recent marketing studies (Feng et al., 2018; Guo et al., 2018). Such causal methods open up exciting opportunities which researchers will continue to explore. Meanwhile, an important question is whether other machine learning methods can be adapted for causal interpretation, especially in non-experimental settings involving complex environments and dynamic consumer decisions. Considerable adaptations of the methods are likely needed, e.g. establishing statistical consistency and other asymptotic properties. Delving into the underlying structure to pinpoint the linkage between variables is another necessity. While a rich econometric apparatus exists for causal analysis in non-experimental settings, its counterpart in machine learning is still yet to be developed. We call for researchers to pursue this challenging yet potentially highly impactful line of research.

5.4.5. Prescriptive analysis

Marketers seek to not only understand consumers, but also prescribe appropriate actions to achieve desired results. Machine learning methods are a natural fit for this prescriptive usage, and several studies have shown this direction to be highly promising (e.g. Hauser et al., 2009; Misra et al., 2018; Schwartz et al., 2017; Yoganarasimhan, 2018). Many methods amenable to automation and interpretation are promising candidates, and we highlight the family of reinforcement learning methods which fit naturally to prescriptive marketing tasks. For example, as in reinforcement learning, a personalized ad targeting engine needs to learn about each consumer through interactions, while crafting the optimal type, frequency, and timing of ad delivery. Similar situations exist for other marketing decisions such as recommendation, promotion, and engagement. Given its nascent stage, we call for researchers to pursue the abundant opportunities to develop prescriptive systems or frameworks. Decision-support capabilities are especially important in today's highly complex and increasingly automated marketing environment.

5.4.6. Estimation and optimization

One additional way to use machine learning methods is for estimation, taking advantage of their ability to perform large-scale optimization (Ansari et al., 2018; Chiong & Shum, 2019; Evgeniou et al., 2005; Evgeniou, Pontil, & Toubia, 2007; Hauser, Toubia, Evgeniou, Befurt, & Dzyabura, 2010). Researchers are encouraged to explore the collection of machine learning methods for such usage. For example, variational inference methods (Jordan, Ghahramani, Jaakkola, & Saul, 1999; Braun & McAuliffe, 2010;

[15] For example, in kernel SVM where inputs are transformed to higher dimensional representations, the function linking a specific input to the output cannot be easily described.

Blei, Kucukelbir, & McAuliffe, 2017) have been used in marketing for model estimation, and given the similarity in basic model setup, reinforcement learning methods may be leveraged for approximation estimation of dynamic forward-looking models.

5.5. Addressing important substantive issues

The rapid AI-driven transformation presents many interesting and important substantive issues, especially in the domain of digital marketing (Kannan & Li, 2017). The major trends and marketing practices discussed in Section 3 provide ample topics for impactful research, and machine learning methods can play an important role in addressing them. In this subsection, we organize these substantive issues into several research themes.

5.5.1. Mapping out customer-purchase journey

Understanding consumer's temporal behavioral dynamics, multitasking behavior, and context-dependent decisions is becoming ever more crucial, and many substantive issues along the customer-purchase journeys call for researchers' attention. First, continued knowledge discovery from rich media data is called for. On average, U.S. consumers are spending over eleven hours a day connected to media, almost six hours on video content (Nielsen, 2018). What can we learn about the consumers who supply the digital content? How do consumers adopt and consume the content? What do they learn from it and how do they respond to it? Machine learning methods are well positioned to help answer these important questions. Second, the rich media and tracking data can improve our understanding of consumer's choice and decision process. Traditionally, marketing contacts are analyzed only at the incidence level, e.g., using the number of contacts as a covariate. Relaxing such simplifications can lead to richer findings (Braun & Moe, 2013). Topics extracted from product reviews, for example, provide more insights into consumers than those obtained from ratings alone (Tirunillai & Tellis, 2014), and attribute-level sentiment further expands the understanding (Chakraborty et al., 2019). With the help of machine learning methods, researchers can conduct deeper and more refined investigations on how and why consumers make their decisions, and the new findings could potentially correct previous misconceptions. Third, integrated dynamic modeling of the entire customer journey is an exciting research opportunity. A holistic view of the customer journey can generate understandings of consumer decisions in a broader context, enable the analysis of the short-term and long-term impacts of marketing interactions, and help prescribe program designs that address the whole customer lifecycle. We call for research to map out the entire customer purchase journey, especially the early stages which are comparatively underexplored. Both machine learning methods that can handle rich-media and fine-grained tracking data, e.g. deep learning methods, and those that can incorporate dynamics, e.g. PGMs or reinforcement learning methods, can be instrumental for such research.

5.5.2. Developing automated decision-support capabilities

The concept of adaptively learning from customer behaviors, and automatically optimizing marketing response in a forward-looking framework, has been applied to service management and cross-selling (Li, Sun, & Montgomery, 2011; Sun & Li, 2011; Sun, Li, & Zhou, 2006). With the increasingly complex marketing environment, similar decision-support capabilities are needed in all aspects of marketing functions, with automated systems making real-time context-dependent decisions in an online fashion. The heighted speed and scale requirements of these automated systems raise important substantive issues that call for researchers' attention. First, in areas where machine learning algorithms are commonly used in industry, we call for the incorporation of marketing insights. For example, the rich machine learning literature on recommender systems focuses on the somewhat narrow angle of predicting ratings and adoptions. For managers, a more decision-oriented perspective is needed, e.g., to assess the incremental effect of recommendations (Bodapati, 2008). Using marketing insights to strengthen recommender systems is thus an interesting research issue. Second, in areas rich with theory-driven models, we call for research to scale up the model to meet real-world demands. For example, sophisticated choice models exist to account for consumer dynamics and unobserved heterogeneity. How should such models be scaled up to handle thousands of microsegments instead of two or three, while maintaining the grounding in marketing theories? We encourage research into the adaptation or approximation using machine learning methods, or the integration of machine learning methods into theory-driven marketing models, to connect marketing insights to real-world scales. For example, one may consider using reinforcement learning methods to approximate DP models. Finally, in areas where neither machine learning methods nor theory-driven models dominate, we call for developing decision-support capabilities which combine scale and insights. Automated agents are likely to reach every aspect of marketing, from marketing mix decisions, to customer engagement and relationship management (Ma et al., 2015; Schweidel & Moe, 2014), to cross-channel coordination and attribution (de Haan, Kannan, & Verhoef, 2018; Li & Kannan, 2014), to search and recommendations, and to even traditionally qualitative areas such as product and advertising design. We call for extensive research on developing automated decision-support capabilities at scale.

5.5.3. Market structure analysis

Understanding brand perceptions and product positioning in the digital age becomes more challenging given the ever more complex environment. A holistic view of how consumers perceive of a brand requires analyzing the myriad of channels such as paid search, social media, and e-commerce sites. Machine learning methods have shown promise to tackle such challenges (Netzer et al., 2012), and we call for continued research in this area. Many online platforms, e.g. search engine platforms such as Google, social curation platforms such as Pinterest, and e-commerce platforms such as Amazon marketplace, hold the market-level information that enable such research. On a social media platform, brands with overlapping followers may be close competitors; on an e-commerce platform, brands with frequent co-purchases may be close complements; on a social

curation site, a holistic view of content and consumers may reveal distinct interest groups (Ma et al., 2019). Machine learning methods have the capacity to address such market structure and brand positioning questions. We call for research to both incorporate the competitive perspective commonly taken in the empirical IO literature (e.g. Berry, 1994; Berry, Levinsohn, & Pakes, 1995), and go beyond it to account for complex complementary relationships and dynamic evolutions.

5.6. Unifying machine learning methods with marketing theories

The final important component in the conceptual framework is marketing theory. To obtain generalizable insights, marketing research needs a strong theoretical foundation. Compared with using econometric models, though, researchers face more challenges in connecting machine learning methods to marketing theories. To this we provide the following discussion.

5.6.1. Human insights

When using machine learning methods, researchers should use human insights for guidance. Machine learning methods typically address data at an abstract mathematical level. A kNN classifier, for example, is simply an exercise of calculating and sorting distances, irrespective of what each dimension or the distance represents. The method can be applied to geographical data in the same way as to consumer preference data. While this generalizability is desirable, it also means that when using machine learning methods for marketing research, it is incumbent upon the researchers to inject human insights into the analysis. This can begin with feature engineering guided by domain knowledge, which studies have shown to be important for generating substantive insights (e.g. Yoganarasimhan, 2018). More importantly, human insights should also guide the other research aspects. A deep theoretical understanding of the research issue, the context, and the related factors can help researchers choose the right machine learning method, to adapt and extend it in a way that is consistent with theories and domain knowledge, and to perform the appropriate analysis to answer the research questions. For example, after a consumer purchases a product, a collaborative filtering algorithm may give other products in the same category higher score. For an infrequently purchased category, however, such recommendation immediately after a purchase will likely be ineffective, and such domain knowledge should be used to adjust the model. With methods becoming increasingly sophisticated and data growing in scale and variety, research can easily be consumed by technical complexities. How to effectively incorporate human insights into studies using machine learning methods is thus a key item on the research agenda.

5.6.2. Balancing theory-driven and data-driven perspectives

Marketing research tends to be theory-driven. Researchers seek to obtain or verify generalizable results on consumer or firm behaviors or market characteristics. Machine learning, in contrast, focuses on extracting knowledge from data, and tends to impose few constraints a priori so as to give data the opportunity to speak. A successful unification between the two calls for a hybrid approach which balances both perspectives. Theories could form the foundation and guide the research design, which allows researchers to draw insights that are generalizable. Meanwhile, a data-driven perspective would ensure that researchers recognize the real-world complexity, and allow for sufficient model flexibility to accommodate complex situations. Integrating these two perspectives seamlessly is both challenging and important. This is another item on the research agenda.

5.6.3. Investigating theoretical implications of machine learning methods

Machine learning methods may have theoretical implications from marketing perspective. For example, in reinforcement learning models an agent interacts with the environment to learn its characteristics and to simultaneously optimize a policy. This is conceptually similar to a consumer who learns from the market while maximizes her utility, or to a firm which seeks to learn about individual consumers and to optimize marketing interactions. Modeling consumers or firms as reinforcement learners thus can be an interesting avenue of research. As another example, the different layers of deep neural networks represent different levels of abstractions, which may also be connected to how consumers process rich media content. Investigating marketing theoretical implications of machine learning methods is an underexplored topic, and another key item on the research agenda.

5.6.4. Investigating theoretical implications of AI-adoption in industry

With firms using AI to power their business decision makings, the practice of adopting AI tools may have theoretical implications to marketing research. For example, the use of machine learning methods by practitioners raises endogeneity concerns for studies that use non-experimental data. If advertising data are obtained from a firm that performs retargeting, researchers have to take that into account when using the data to analyze consumers' ad response and to develop targeting strategies. Such endogeneity issue may also manifest at the method level. For example, if a firm uses a specific method to generate recommendation, then that method would have better predictive performance than other methods when assessed using the firm's data, even though such comparison does not generalize to other settings. From a broader perspective, the adoption of AI tools can alter the economy in more fundamental ways, e.g. by creating a feeling economy, with significant social welfare implications (Huang, Rust & Maksimovic 2019). With the complex environment created by deploying AI technologies, other theoretical issues may also emerge, and researchers are encouraged to consider their implications to marketing research at a deeper level.

Compared with the other perspectives, the research agenda from the perspective of unifying machine learning methods and marketing theories may be the most challenging. Meanwhile, breakthroughs in this area would also be impactful.

6. Conclusion

The coming decades will witness the proliferation of automated AI agents powered by machine learning methods into every aspect of business and marketing, driven by technology, big data, and competition. For academic research, it is imperative to leverage the rich digital information to deepen the understanding of firms and consumers, to address emerging substantive issues in the field, and to develop scalable and automated decision support capabilities that will become essential to business managers. Machine learning methods have great potential to help address important research issues from all these perspectives. Effectively incorporating machine learning methods for marketing research is challenging, given the different focuses of the two fields. However, the opportunities are broad-based, and the potential contributions from successfully leveraging these methods well justify the effort required to address the challenges. We hope this paper can encourage the application of machine learning methods in many areas of marketing research, to generate knowledge about the constantly evolving business world.

References

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185.

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, *6*.

Ansari, A., Li, Y., & Zhang, J. Z. (2018). Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach. *Marketing Science*, *37*(6), 987–1008.

Athey, S., & Imbens, G. (2016). Recursive partitioning for estimating heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Bahdanau, D., K. Cho and Y. Bengio, (2014), "Neural machine translation by jointly learning to align and translate," *arXiv preprint*, arXiv:1409.0473

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *The Rand Journal of Economics*, *242–262*.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*, 841–890.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(June), 17–35.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bodapati, A. V. (2008). Recommendation systems with purchase data. *Journal of Marketing Research*, *45*(1), 77–93.

Braun, M., & McAuliffe, J. (2010). Variational inference for large-scale models of discrete choice. *Journal of the American Statistical Association*, *105*(489), 324–335.

Braun, M., & Moe, W. W. (2013). Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing Science*, *32*(5), 679–826.

Breiman, L. (1996). Stacked regression. *Machine Learning*, *24*, 49–64.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, *35*(6), 953–975.

Chakraborty, I., M. Kim, and K. Sudhir, (2019), "Attribute Sentiment Scoring with Online Text Reviews: Accounting for Language Structure and Attribute Self-Selection," working paper, SSRN: https://ssrn.com/abstract=3395012

Chen, T. and C. Guestrin, (2016), "Xgboost: A scalable tree boosting system," In *Proceedings of the 22n acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, ACM.

Chiong, K. X., & Shum, M. (2019). Random projection estimation of discrete-choice models with large choice sets. *Management Science*, *65*(1), 256–271.

Chung, T. S., Rust, R. T.., & Wedel, M. (2009). My Mobile Music: An Adaptive Personalization System for Digital Audio Players. *Marketing Science*, *28*(1), 52–68.

Chung, T. S., Wedel, M., & Rust, R. T.. (2016). Adaptive Personalization Using Social Networks. *Journal of the Academy of Marketing Science*, *44*(1), 66–87.

Cohn, D. A., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. *Journal of Artificial Intelligence Research*, *4*, 129–145.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machines. *Marketing Science*, *24*(4), 595–615.

Dai, Z., Z. Yang, Y. Yang, J. Carbonell, Q.V. Le, and R. Salakhutdinov, (2019), "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint* arXiv:1901.02860.

de Haan, E., Kannan, P. K., & Verhoef, P. C. (2018). Device switching in online purchasing: Examining the strategic contingencies. *Journal of Marketing*, *82*(5), 1–19.

Dong, Y., Chawla, N. V., & Swami, A. (2017). *metapath2vec: Scalable representation learning for heterogeneous networks.* (In KDD).

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. N. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, *9*, 155–161.

Dzyabura, D., S. El Kihal, J.R. Hauser, and M. Ibragimov, (2019), "Leveraging the power of images in managing product return rates," working paper, SSRN: https://ssrn.com/abstract=3209307

Dzyabura, D., & Hauser, J. R. (2011). Active machine learning for consideration heuristics. *Marketing Science*, *30*(5), 801–819.

Erdem, T., & Keane, M. P. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, *15*(1), 1–20.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD*, *96*(34), 226–231.

Evgeniou, T., Boussios, C., & Zacharia, G. (2005). Generalized robust conjoint estimation. *Marketing Science*, *24*(3), 415–429.

Evgeniou, T., Pontil, M., & Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, *26*(6), 805–818.

Feng, F., L. Zhang, and V.R. Rao, (2018), "Multidimensional Cheap Talk: An Analysis of the Online Home Rental Market," working paper.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Frost, C., & Thompson, S. G. (2000). Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *163*(2), 173–189.

Goodfellow, I., Y. Bengio, and A. Courville, (2016), "Deep learning," *MIT press.*

Grimes S., (2008), "Unstructured data and the 80 percent rule," Clarabridge BridgePoints, URL: http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/,accessed in November 2019.

Guo, T., S. Sriram, and P. Manchanda, (2018), "The Effect of Information Disclosure on Industry Payments to Physicians," working paper.

Hartmann, J., M. Heitmann, C. Schamp, and O. Netzer, (2019), "The Power of Brand Selfies in Consumer-Generated Brand Images," working paper, SSRN: https://ssrn.com/abstract=3354415

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, *36*, 20–38.

Hauser, J. R., Toubia, O., Evgeniou, T., Befurt, R., & Dzyabura, D. (2010). Disjunctions of conjunctions, cognitive simplicity, and consideration sets. *Journal of Marketing Research*, *47*(3), 485–496.

Hauser, J. R., Urban, G. L., Liberali, G., & Braun, M. (2009). Website morphing. *Marketing Science*, *28*(2), 202–223.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554.
Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.
Hofmann, T. (1999). *Probabilistic latent semantic indexing.* (Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval).
Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*(2).
Huang, D., & Luo, L. (2016). Consumer preference elicitation of complex products using fuzzy support vector machine active learning. *Marketing Science*, *35*(3), 445–464.
Huang, M. -H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, *21*(2), 155–172.
Huang, M. -H. and R.T. Rust (2020), "Engaged to a Robot: The Role of AI in Service," Journal of Service Research, forthcoming.
Huang, M. -H., Rust, R. T., & Maksimovic, V. (2019). The Feeling Economy: Managing in the Next Generation of AI. *California Management Review*, *61*(4), 43–65.
Jacobs, B. J. D., Donkers, B., & Fok, D. (2016). Model-based purchase predictions for large assortments. *Marketing Science*, *35*(3), 389–404.
Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*, 183–233.
Kakatkar, C., & Spann, M. (2019). Marketing analytics using anonymized and fragmented tracking data. *International Journal of Research in Marketing*, *36*, 117–136.
Kannan, P. K., & Li, H. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, *34*, 22–45.
Kawaf, F. (2018). Capturing digital experience: The method of screencast videography. *International Journal of Research in Marketing*. https://doi.org/10.1016/j.ijresmar.2018.11.002.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *In Advances in Neural Information Processing Systems, pp.*, 3146–3154.
Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2018). Extracting brand information from social networks: Integrating image, text, and social tagging data. *International Journal of Research in Marketing*, *35*, 538–556.
Koller, D. and N. Friedman, (2009), "Probabilistic graphical models: principles and techniques," *MIT press.*
LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10).
LeCun, Y., Y. Bengio and G. Hinton, (2015), "Deep Learning," *Nature*, Vol. 521, pp 436–444.
Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*(5), 881–894.
Lewis, D.D. and W.A. Gale, (194), "A sequential algorithm for training text classifiers," In *SIGIR'94* Springer, London, pp. 3–12.
Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, *51*(1), 40–56.
Li, H. and L. Ma, (2019), "Uncovering the Path to Purchase Using Topic Models," working paper.
Li, S., Sun, B., & Montgomery, A. L. (2011). Cross-selling the right product to the right customer at the right time. *Journal of Marketing Research*, *48*(4), 683–700.
Li, X., Shi, M., & Wang, X. (2019). Video mining: Measuring visual information using automatic methods. *International Journal of Research in Marketing*. https://doi.org/10.1016/j.ijresmar.2019.02.004.
Lin, C. F., & Wang, S. D. (2002). Fuzzy support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 464–471.
Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, *4*, 4–22.
Lipton, Z.C., (2016), "The mythos of model interpretability," *arXiv preprint* arXiv:1606.03490
Liu, J., & Toubia, O. (2018). A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science*, *37*(6), 855–882.
Liu L., D. Dzyabura, and N. Mizik, (2018) "Visual Listening In: Extracting Brand Image Portrayed on Social Media," working paper.
Liu, X., D. Lee and K. Srinivasan, (2018) "Large Scale Cross Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," working paper.
Ma, L., Sun, B., & Kekre, S. (2015). The squeaky wheel gets the grease—An empirical analysis of customer voice and firm intervention on twitter. *Marketing Science*, *34*(5), 627–645.
Ma, L., Sun, B., & Zhang, K. (2019). *"Image network and interest group – A heterogeneous network embedding approach to analyze social curation on Pinterest," working paper.*
Malik, N., P.V. Singh, D.K. Lee, and K. Srinivasan, (2019), "A Dynamic Analysis of Beauty Premium," working paper, SSRN: https://ssrn.com/abstract=3208162
McCarthy, J., M.L. Minsky, N. Rochester, and C.E. Shannon, (1955), "A proposal for the Dartmouth summer research project on artificial intelligence," Retrieved from http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html
McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.
Mikolov, T., K. Chen, G. Corrado, and J. Dean. "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
Mimno, David, and Andrew McCallum (2012), "Topic Models Conditioned on Arbitrary Features with Dirichlet-Multinomial Regression," *arXiv preprint arXiv*, 1206.3278.
Misra, K., Schwartz, E. M., & Abernethy, J. (2018). *Dynamic online pricing with incomplete information using multi-armed bandit experiments.* Ross School of Business: Paper.
Mitchell, T., (997), "Machine learning," *McGraw Hill*, p. 2, ISBN 978-0071154673.
Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.
Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*(3), 521–543.
Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). A hidden Markov model of customer relationship dynamics. *Marketing Science*, *27*(2), 185–204.
The Nielsen Total Audience Report. https://www.nielsen.com/us/en/insights/report/2018/q1-2018-total-audience-report/(2018)..
Ng, I. C. L., & Wakenshaw, S. Y. L. (2017). The internet-of-things: Review and research directions. *International Journal of Research in Marketing*, *34*, 3–21.
Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.
Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, (2018), "Deep contextualized word representations," *arXiv preprint* arXiv:1802.05365.
Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106.
Rafieian, O., & Yoganarasimhan, H. (2018). *"Targeting and privacy in mobile advertising," working paper.*
Romov, P. and E. Sokolov, (2015) "Recsys Challenge 2015: Ensemble Learning with Categorical Features," *In Proceedings of the 2015 International ACM Recommender Systems Challenge*, page 1. (ACM.)
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.
Rust, R. T. (2020). The Future of Marketing. *IJRM*, *37*(1), 15–26.
Rust, R. T., & Huang, M. -H. (2014). The Service Revolution and the Transformation of Marketing Science. *Marketing Science*, *33*(2), 206–221.
Schwartz, E. M., Bradlow, E. T., & Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, *36*(4), 500–522.
Schweidel, D. A., & Moe, W. W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research*, *51*(4), 387–402.
Selvaraju, R.R., M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, (2017), "Grad-cam: Visual explanations from deep networks via gradient-based localization," In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618–626).
Simonyan, K., A. Vedaldi, and A. Zisserman, (2013), "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," arXiv preprint arXiv:1312.6034.
Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, *5*, 595–645.
Sun, B., & Li, S. (2011). Learning and acting on customer information: A simulation-based demonstration on service allocations with offshore centers. *Journal of Marketing Research*, *48*(1), 72–86.

Sun, B., Li, S., & Zhou, C. (2006). "Adaptive" learning and "proactive" customer relationship management. *Journal of Interactive Marketing, 20*(3–4), 82–96.
Sutton, R.S. and A.G. Barto, (2018), "Reinforcement learning: An introduction," *MIT press.*
Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). Line: Large-scale information network embedding. *In WWW, pp.*, 1067–1077.
Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, *51*(4), 463–479.
Toubia, O., & Netzer, O. (2017). Idea generation, creativity, and prototypicality. *Marketing Science*, *36*(1), 1–20.
Trusov, M., Ma, L., & Jamal, Z. (2016). Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Science*, *35*(3), 405–426.
Tsochantaridis, I., Joachims, T., Hofmann, T., & Altun, Y. (2005). Large margin methods for structured and interdependent output variables. *The Journal of Machine Learning Research*, *6*, 1453–1484.
Turing, A., (1950), "Computing Machinery and Intelligence," *Mind*, Vol. LIX, (No. 236).
Vapnik, V. (1998). *Statistical learning theory.* Wiley.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *In Advances in Neural Information Processing Systems, pp.*, 5998–6008.
Vermeer, S. A., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing, 36*(3), 492–508.
Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113* (523), 1228–1242.
Watkins, C.J.C.H., (1989), "Learning from delayed rewards," PhD Thesis, University of Cambridge, England.
Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning, 8*(3–4), 279–292.
Wu, Q., Burges, C. J., Svore, K. M., & Gao, J. (2010). Adapting boosting for information retrieval measures. *Information Retrieval, 13*(3), 254–270.
Yang, Y., K. Zhang, and P.K. Kannan, (2019), "Identifying Market Structure: A Deep Network Representation Learning of Social Engagement," working paper.
Yoganarasimhan, H., (2018), "Search Personalization using Machine Learning," *Management Science*, forthcoming.
Zhang, M. and L. Luo, (2019), "Can User-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp," working paper, SSRN: https://ssrn.com/abstract=3108288
Zhang, S., D. Lee, P. Singh, and K. Srinivasan, (2017), "How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics," working paper.
Zhang, Z., Y. Song, and H. Qi, (2017), "Age progression/regression by conditional adversarial autoencoder," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 5810-5818.
Zhu, X. J. (2005). *Semi-supervised learning literature survey.* University of Wisconsin-Madison Department of Computer Sciences.