Simulation

Special Issue Article

Simulation: Transactions of the Society for Modeling and Simulation International
2018, Vol. 94(3) 195–212

DOI: 10.1177/0037549717743106
journals.sagepub.com/home/sim
SAGE SCS

Data-driven agent-based exploration of customer behavior

David Bell and Chidozie Mgbemena

Abstract

Customer retention is a critical concern for mobile network operators because of the increasing competition in the mobile services sector. Such unease has driven companies to exploit data as an avenue to better understand changing customer behavior. Data-mining techniques such as clustering and classification have been widely adopted in the mobile services sector to better understand customer retention. However, the effectiveness of these techniques is debatable due to the constant change and increasing complexity of the mobile market itself. This design study proposes an application of agent-based modeling and simulation (ABMS) as a novel approach to understanding customer behavior through the combination of market and social factors that emerge from data. External forces at play and possible company interventions can then be added to data-derived models. A dataset provided by a mobile network operator is utilized to automate decision-tree analysis and subsequent building of agent-based models. Popular churn modeling techniques were adopted in order to automate the development of models, from decision trees, and subsequently explore possible customer churn scenarios. ABMS is used to understand the behavior of customers and detect reasons why customers churned or stayed with their respective mobile network operators. A CART decision-tree method is presented that identifies agents, selects important attributes, and uncovers customer behavior – easily identifying tenure, location, and choice of mobile devices as determinants for the churn-or-stay decision. Word of mouth between customers is also explored as a possible influence factor. Importantly, methods for automating data-driven agent-based simulation model generation will support faster exploration and experimentation – including with those determinants from a wider market or social context.

Keywords

agent-based modeling, decision trees, customer behavior

1. Introduction

Simulation models that describe domain-specific agents are widely used and popular tools for understanding phenomena. These models have been used across industries and provide insights into a range of complex problems. Agent-based models (ABMs) allow researchers and practitioners to study how system-level properties emerge from the adaptive behavior of individuals and conversely how systems affect those individuals.[1] Domain-specific data sources, such as log or transaction files, are typically used to construct agents and their operating environments.

ABMs consist of a number of entities with individual rules of behavior. Entities in such models interact with one another and with their surrounding environment. Such interaction may influence the behavior of agents. Harnessing this information and understanding the influence of agents' interaction with other agents and agent interaction with the environment can provide useful insights to business problems – in this case, customer churn.

There are various ways of developing ABMs. Adopting a design science paradigm, this paper presents a novel data-driven approach to agent-based modeling in which agents are derived from determinants of interest using decision-tree analysis. The Customer Agent DEcision Tree (CADET) approach is a data-driven approach that provides key drivers that collectively uncover the decision-making of individual agents in a mobile phone marketplace. The CADET approach, a design method, is not industry specific but does require customer and determinant specific

Brunel University London, Kingston Lane, London, UK

Corresponding author:
David Bell, Brunel University London, Kingston Lane, Uxbridge, London, UB8 3PH, UK.
Email: david.bell@brunel.ac.uk

datasets for the domain of interest. It is a data-driven development for ABM development that aims to more effectively extract agents, attributes, and behaviors from data. Subsequent validation is carried out through instantiation in an agent-based simulation (ABS), investigating customer retention in the mobile services industry (MSI).

The next sections present background on customer retention in the MSI before coverage of agent and social network analysis. A design science methodology is presented before CADET description – a design method – and evaluation though implementation in an ABS tool – a design instantiation.

2. Background on customer retention

Over the last decade the number of mobile phone users has increased, reaching a staggering seven billion.[2] In developed countries, telecommunication companies have mobile penetration rates above 100%, with no new customers.[3] Consequently, customer retention receives a growing amount of attention from telecommunication companies. The high penetration rate of $>100\%$ in the mobile services sector has motivated substantial research into customer retention. It has been shown in the literature that customer retention is profitable to a company because: (1) acquiring a new customer costs five times more than retaining an existing customer[4,5]; (2) existing customers generate higher profits, become less costly to serve, and may provide new referrals through providing positive word of mouth (WOM) while dissatisfied customers may spread negative WOM[6–8]; and (3) losing customers may lead to opportunity costs because of reduced sales.[9,10] In such large markets it is no surprise that a small improvement in customer retention can lead to a significant increase in profits.[11] A number of studies in the area of customer retention have revealed that customer satisfaction is a strong predictor of customer retention.[12] This is as a result of the relationship between both conceptual elements.[13] Factors that drive customer satisfaction (such as service quality) can also drive customer retention.[14] Furthermore, customer satisfaction is often seen as a motivator for customer retention.[15] While it seems that satisfied customers will remain with their mobile service provider, this is not always the case because satisfied customers can defect, while dissatisfied customers can be retained.[15]

Customer churn is a term that is widely used in the area of customer retention to describe customers who switch to a different mobile service provider or leave the market entirely. There are two basic approaches that can be used to address customer churn, namely untargeted and targeted approaches.[16] Untargeted approaches rely on outstanding products and mass advertising to increase brand loyalty and retain customers, while targeted approaches rely on identifying customers who are likely to churn and providing them with either a direct incentive or a customized plan to stay.[16] Various types of information can also be used to predict customer churn, such as information on socio-demographic data (e.g., sex, age, or post-/zip-code) and call behavior statistics (e.g., the number of international calls, billing information, or the number of calls to the customer help-desk).

The main factors that influence customer churn in the mobile services market are: (1) customer satisfaction, (2) switching costs, (3) relationship quality, and (4) price.[17–19] Price is the most important factor for customer churn, followed by customer service, service quality, and coverage quality.[20] However, social influence is another key driver to customer churn in the MSI.[21,22] ABMs provide a means to explore social influence on customer churn behavior. The next section presents approaches to modeling customer behavior with respect to customer retention.

3. Customer modeling behavior

Customers often interact with other customers about products and services they purchase. In addition to the advertisement campaigns carried out by mobile network operator (MNOs), interaction among customers also influences customer purchase and repurchase decisions. There are various theories in social science and marketing concerned with understanding and modeling customer behavior.[23] Customer behavior may change as a result of an act of a consumer changing preference on a product or service.[24] A number of approaches have been applied to understand the concept behind consumers changing their preferences.[24]

Customer retention can be achieved if a company is able to understand patterns by which customers behave and the likely triggers for such behavior.[25] Understanding customers and managing interactions and relationships with them is a vital part of customer relationship management (CRM). A company with good CRM should be able to predict possible changes in customer behavior.[25] Predicting customer behavior can be achieved through customer behavior modeling,[26],applying tools and techniques to gain a better insight on customer behavioral patterns and in turn predict future behavior.[27] Neslin and colleagues[28] characterized CRM models as either analytical or behavioral models. Analytical models typically utilize large datasets stored in data warehouses. These datasets require models that can easily scale the dataset and provide results to increase company revenue.[28] Behavioral models often make use of surveys to analyze cognitive responses to services provided.[28] Furness[29] classified customer behavior modeling into: (1) descriptive modeling, (2) predictive modeling, and (3) a combination of descriptive and predictive modeling. Descriptive modeling attempts to answer *why* questions using techniques such as clustering. When a customer clustering exercise is conducted, customers belong to a certain cluster because they collectively possess similar attributes or behaviors. Predictive modeling describes models that answer the *who* questions. For example, who will buy a product or service? In the context of customer churn, predictive models can give insight on who is likely to churn.[30] Predictive models typically predict future customer behavior based on their past behavior.[30] Finally, the combination of descriptive and predictive modeling addresses problems by integrating both descriptive and predictive models to provide a more concise answer on *who* and *why* questions at the same time.[3] Descriptive and predictive models are typically carried out using data-mining approaches. Data-mining approaches offer opportunities to both investigate customer churn and uncover elements of an ABS.

3.1. Applying data-mining techniques for modeling churn

Data-mining techniques are commonly used to build churn prediction models.[31,32] Data mining simply means extracting hidden knowledge from data, and it is a popular technique for understanding customer behavior from raw data. Data-mining methods are widely used in the literature for analyzing and investigating customer churn as they demonstrate better prediction results[33] and are more suitable for analyzing large datasets.[34]

Numerous studies have applied a range of data-mining techniques to study customer churn.[35–37] The most popular data-mining techniques for predicting churn include decision trees, logistic regression, and support vector machines (SVM).[35] Companies have attempted to apply data-mining techniques to study customer churn. Industries include banking,[38,39] insurance,[40] retail,[41,42] and economics.[28] Most studies that address customer churn using data-mining techniques do not typically account for social effects on customer retention.[43] As such, this area has also been neglected and is considered a major driver for customer retention.[21,22] Social network analysis provides access and reach when considering large markets (and wider social environments), and as such provides insight into possible ABS scenarios for exploration.

4. Social network analysis

Social network analysis (SNA) is an evolving scientific research area, due in part to the ever-changing feature sets of social networking platforms. SNA has developed to be a primary technique for describing the social structure and interaction between network elements.[44] A social network is often represented as interconnections between nodes using various links.[24] For the purpose of this study, nodes represent customers and the links represent the relationships between customers.

SNA typically focuses on static networks through the use of historical snapshots. Static networks are the mapping of relationships between discrete entities.[45] These networks do not change their structure over time. Recently, researchers have focused on dynamic networks that are capable of representing the continuous transmission of information and influence.[46] Dynamic networks are a vital aspect of agent-based modeling and simulation (ABMS). Using dynamic network analysis involves understanding agent rules that govern network structure and growth, and how networks and their embedded relationships convey information.[47] SNA is an approach to anticipating and modeling society as different sets of people or groups linked to one another.[24] SNA is a method of enquiry that focuses on the relationships between subjects. This approach seeks to understand subjects by collecting information from different sources, analyzing the information, and visualizing the results. SNA has proved fruitful in explaining some important phenomena, such as investigating the spread of disease, understanding the internet, and explaining the small-world effect on the spread of information.[48] Sociologists and market researchers believe that the life of an individual depends on how that individual engages with the web of social connections.[45] The ''social network'' term is loosely used to refer to social and professional networking sites including but not restricted to Facebook, LinkedIn, and Twitter, with each networking site representing an online community. An online community is an example of a social structure. The people (also known as the nodes) who sign up on these online communities, together with the relationship between them, represent a social network. Due to the growing interest of SNA, researchers have investigated the principles of the network approach. These principles include[24]:

1. Actors and their actions should be viewed as autonomous and independent units rather than as interdependent units.
2. The links between actors should be viewed as channels for transfer of material and non-material resources.
3. Social network models focusing on individuals view the structural environment as a network imposing certain constraints on individual actions.
4. Social network models conceptualize structure (social, economic, political, and so on) as long-lasting patterns of relations among actors.

Milgram[48] presents an empirical research study of social structure and introduces the ''six degrees of separation'' phenomenon while addressing the ''small-world problem.'' Participants were chosen at random and asked to deliver a letter to a target person using only a chain of friends and acquaintances. Milgram concludes by describing how people of a population are connected. Although the study was not subject to an evaluation process, his concept of the small world is widely adopted in social networks research to provide an explanation about how information spreads in the real world. The small-world network has become one of the most widely used social network models.[49] The next section provides a background on the social impact of customer retention.

4.1. Social impact on customer retention

More recently, the influence of social networks has been found to be a key contributing factor to customer retention.[50,51] Social network influence is typically carried using WOM.[52] WOM is the informal communication between private parties regarding the evaluation of goods and services.[53] Seventy-five percent of customers with defective products spread a negative WOM to one or more customers.[54] WOM will likely have a social influence on customer retention.

Phadke and colleagues[21] developed a model that integrates SNA with traditional churn modeling concepts. The model was applied to a dataset of over half a million subscribers, provided by a large mobile network provider. The dataset contained customer call detail records. To compute social tie strength, the authors used three attributes: (1) The number of calls placed between two users; (2) the total duration of calls between two users; and (3) neighborhood overlap of the two connected users. The study found that users who make phone calls to numbers on a different network are also likely to churn in future in order to save costs. Similarly, Verbeke and colleagues[22] conducted a study investigating the impact of social networks on customer retention. However, the latter differs from the former in that it uses both networked and non-networked (customer-related) information about millions of users. The key finding of this study was that churn not only had an impact on customers' friends, it also had an impact on friends of friends.

Although the studies mentioned above have contributed to knowledge of customer retention, they do not capture possible factors as to why customers make their decision to churn. ABMS is able to explore the dynamics of customer behavior.[55] Exploration of possible determinants is important with limited theory. Accessible customer datasets enable ABMS approaches, including SNA insight.

5. Agent-based modeling and simulation

Over the years, researchers and industry practitioners have attempted to apply different techniques to understand customer behavior in the market place. The ABMS approach is a typical example of one such technique.[56] ABMS provides an understanding of how systems work under certain conditions, creating scenarios that imitate real-life conditions. For example, carrying out an ABMS exercise with customers who enter a retail space. ABMS can be used in this context to derive insights into customer behavior within the retail store.[57] This process provides an explanation of the relationship between elements of a complex system, including customers, products, and the built environment. ABMS is composed of two main activities: modeling and simulation. Modeling is the process of representing real-life events into a model, while simulation is the process of executing the represented models such that they imitate the proposed system. ABMs are composed of agents and a structure for agent-based interaction.

Agents can represent anything from a number of patients in a hospital to consumers of a product or service. ABMs are often characterized by rules and these rules define the behavior of agents in the system.[47] These behaviors are often influenced by agent interactions with other agents in the system, making the outcome difficult to predict. In such cases, a balance may be difficult to reach, making the ability to study the underlying system and the dynamics of the behavior imperative. ABM is distinct from traditional modeling approaches where characteristics are often aggregated and manipulated.[49] Generalization can lose important insight. Traditional modeling techniques are often suitable for their own purposes but they may not be able to provide adequate levels of detail when considering independent behaviors of agents. Although commonly used data-mining techniques for modeling consumer markets are powerful with regard to their purposes, they are generally not able to provide sufficient levels of detail with regards to the modeling of interdependent behaviors of consumers in a complex marketplace.[55]

In addition, ABM is able to sufficiently represent interdependent systems even on a large scale (i.e., incorporating a high number of factors, including the required level of detail and the behavioral complexity of chosen factors).[57] ABM is a relatively new approach for modeling complex systems that are composed of interacting independent elements.[47] ABM techniques can be applied to any aspect of a phenomenon[47] – in various areas including economics,[58] healthcare,[59] management science,[47] artificial life,[60] and geography.[61] In business, ABM has been applied to help decision-makers understand underlying market structures and anticipate dynamics in the marketplace.[47] ABM has also been utilized in artificial life research, to explore life in order to uncover how it might be, rather than how it actually is.[60] ABM is also used in consumer modeling to understand and to predict consumer behavior processes.[62] Consumers are represented as independent agents with individual characteristics and an independent decision-making process.[63] Sellers are represented as agents who present their products with different characteristics into the market.[64] Data-driven approaches to behavioral and social simulation are reported in the literature.[65,66] with a natural

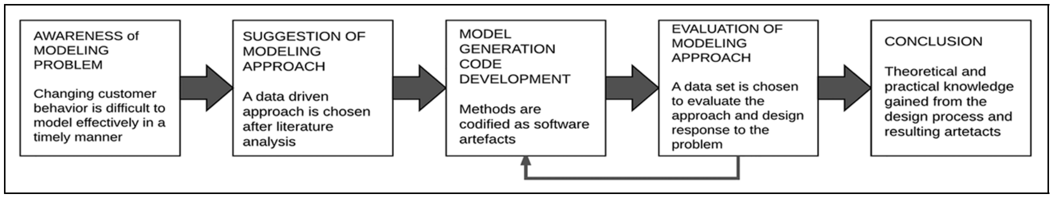

Figure 1. Design science research process.

focus on uncovering distinct behavior (e.g., rule mining). These promising approaches typically identify agents upfront, whereas here we uncover agents from the data analysis itself – in relation to determinants of interest. Domain complexity warrants a looser customer typing in order to adapt to new or changing data content and structure. In the study of social behavior and interactions, ABM starts with a set of assumptions derived from the real world (deductive), and produces simulation-based data that can be analyzed (inductive).[67] Importantly, ABM must create a clear representation of what happens in reality so that each agent performs a task associated with an individual as if it is happening in social reality.[64]

ABM has a number of benefits such as the ability to model individual decision-making while incorporating heterogeneity and interaction/feedback.[68] In addition, ABM has the ability to incorporate social/ecological processes, structures, norms, and institutional factors.[69] These advantages make it possible to couple human and natural systems within an ABM. This paper applies the problem-solving design science paradigm as an overarching methodology in order to uncover techniques to rapidly create agent models from both the data analysis of the market and social interaction within it. The problem being addressed is that markets are dynamic and models need to be speedily created in response. The CADET approach captures the dynamics in customer behavior in a simulated environment. The next section provides a background on design science research (DSR) methodology.

6. Design science research methodology

DSR is a multidisciplinary approach that primarily uses design as a research method or technique to solve a problem and learn from the process of solving that problem.[70] The ability to synthesize a range of artifacts in response to a problem make DSR directly applicable to the problem being explored. Apart from its popular adoption in information systems, it is also widely used in disciplines such as education, engineering computer science, and healthcare.[71] March and Smith[72] describe DSR as a research methodology that allows research to produce relevant and improved effectiveness by strategically combining research output (product) and research processing (activities) from both natural and design science in a two-dimensional framework. Design science output or artifacts includes constructs, models, methods, and instantiations, while the natural science activities include build, evaluate, theorize, and justify.[72] The DSR output classification defined by March and Smith[72] can help establish appropriate measures to build, evaluate, theorize, and justify a DSR. The four artifact types comprise:

Constructs: Articulate set of concepts that are used to describe problems within a domain and specify their solutions. Constructs also form the vocabulary of a discipline.
Models: Define a set of statements which express relationships among constructs and represent real-world design activities in a domain.[72] Models can also be used to suggest effective solutions.
Methods: A sequence of steps used to execute a task. These steps provide guidelines on how to solve problems with the use of constructs and models. Furthermore, methods can be described as a set of methodological tools that are created by design science and applied by natural science.[72]
Instantiations: The utilization of constructs, models, and methods to showcase an artifact in a domain. They demonstrate the effectiveness of constructs, models, and methods.[72] Newell and Simon[73] describe the importance of instantiations in computer science by explaining how they offer a better understanding of a problem domain, and as a result, provide improved solutions. Instantiations provide working artifacts that can drive significant advancement improvement in both design and natural sciences. A DSR methodology incorporates five stages of a design cycle to address design research problems. These phases are designed to aid sustainable development during the research and transfer knowledge from one iteration to the next iteration until a desired result is achieved. The next section explains the DSR processes.

6.1. The design science research process

The DSR process follows a stepwise approach structured as five phases, shown in Figure 1.

Awareness of problem: The DSR process begins by identifying the problem under study. The identified problem may arise from multiple sources, such as the literature or current problems in the industry. The research problem needs to be clearly defined and articulated. The output of this phase is a formal or informal proposal for new research. In this study, the core problem is identifying an effective approach to modeling changing customer behavior as a means to understand and improve customer retention. Included is the need to carry out experimentation in areas of little theory, such as the impact of social connections. A rich environment of internal and external data provide additional context, motivating a data-driven strategy.

Suggestion: Possible solutions about the research problem are explored and evaluated, leading to the acquisition of further insights to the domain under study. The specifications of the appropriate solutions to the research problem are defined. The output of this phase is a conditional design or representation of proposed solutions. In this design study, the CADET approach to model building is suggested. This is articulated in diagrammatic and narrative form. Machine learning approaches are selected and tested using the dataset and R code.

Development: This phase involves further developing and implementing DSR artifacts based on the suggestions from the previous phase. During this phase, the CADET approach is further developed using a selected machine learning algorithm, refining R code to better generate the required outputs.

Evaluation: Developed artifacts are analyzed and evaluated according to the criteria set (awareness of problem phase). Deviations and expectations should be noted and explained. If the outcomes derived from the development or evaluation phase do not meet the objectives of the problem, the design cycle returns to the first phase, along with the knowledge gained from the process of the first round of work. The phase may be iterated until the evaluation of the artifacts meets the solution requirements. The outputs of this phase should improve the efficiency and effectiveness of the artifacts. The CADET approach is operationalized – constructing, extending, and executing an ABS from CADET-derived data. The evaluation itself uses the Telco dataset and the TEA-SIM ABMS platform.

Conclusion: This is the last phase of the DSR cycle. The results of the research are written up and communicated to a wider audience in forms of professional and scholarly publications.[74] Kuechler and Vaishnavi[75] categorize the knowledge gained in this phase as either firm or loose ends. Firm knowledge are facts that have been learned and can be repeatably applied, or behavior that can be repeatably invoked, while loose ends are anomalous behavior that defies explanation and may well serve as the subject of further research.[75] The CADET approach to ABMS can be categorized as firm knowledge because it can be replicated on different datasets (both structured and unstructured) in various industries.

6.2. Design science research evaluation

Evaluation is an integral part of the DSR process. Usually, it is concerned with answering the question "How well does the artifact work?"[72] The evaluation process provides an avenue to validate the performance of an artifact and measure progress according to the defined metrics.[72] Artifacts are constructed to carry out specific problems, thereby demonstrating their effectiveness in solving the problems. The process of developing an artifact may result in deviations from expectations. In this case, these deviations should be properly explained.[75] Knowledge gained from the evaluation phase of one iteration can be applied into further iterations. Evaluation plays an essential role in DSR as it is iterative in manner. Hence, it is important to develop appropriate evaluation metrics to assess artifact performance and to measure the efficiency and effectiveness of the artifact developed.[72] The criteria for evaluating the quality of an artifact depends on the artifact type.[72]

Table 1 presents types of artifacts and their evaluation criteria. The core artifact derived in this study is the CADET approach, which is evaluated though application in a Telco context. The CADET approach was derived as a means to address the problem of customer retention analysis in the MSI. A number of studies in the literature have applied ABMS to address customer retention[76,77]; however, the CADET approach differs in its use of decision trees to model agents and agent behavior feeding an ABMS environment. Decision trees are a central component of the CADET approach.

7. Decision trees

A decision tree is an analytical decision tool that uses a tree-like graph with possible consequences to arrive at a decision. The tree-like structure represents the relationships between events. Decision-tree analysis is a method commonly used in data mining.[78] The goal of a decision tree is to create a model that predicts a target variable based on the input variables. Each leaf on a decision tree shows a connection to the target variable.

Decision trees are sequential models that logically combine a series of simple tests; each individual test compares a numeric attribute against a threshold value or a nominal attribute against a set of possible values.[79] Decision trees consist of nodes and branches in different stages and various conditions.[80] Decision-tree analyses are popular for use in customer churn analysis in the MSI,[35] classified for simplicity into churn or no-churn. Decision-tree models are usually represented in a top-down manner. The main aim of a decision tree is to derive a tree that solves a particular business problem and is relatively easy to understand.

Table 1. DR artifact evaluation criteria.[71]

Artifact	Brief description	Evaluation criteria
Construct	The conceptual vocabulary and symbols describing a problem within a domain.	Completeness, clarity, elegance, ease of understanding, and ease of use.
Model	A set of propositions or statements expressing relationships between constructs. Models represent the situation as problem and solution statements.	Precision with real-world phenomena, completeness, level of detail, robustness, and internal consistency.
Method	A sequence of steps used to perform a task. A method can be tied to a particular model. A method may not be articulated explicitly but represents tasks and results.	Operationality (ability for the method to be reused), efficiency, generality, and ease of use.
Instantiations	Application of constructs, models, and methods to provide working artifacts.	Efficiency and effectiveness of artifacts. Also, the influence of the artifact on its users and on the environment at large.

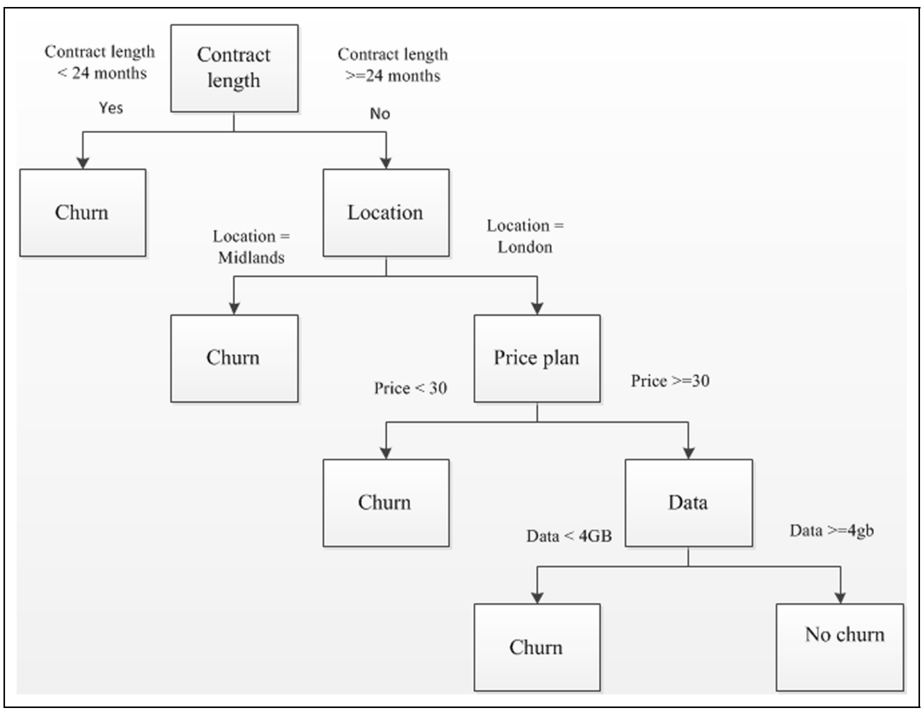

Figure 2. Decision tree classification process.

To achieve this the decision tree undergoes two stages – tree building and tree pruning. Tree building is carried out using a top-down strategy (also known as a divide-and-conquer strategy). The process of tree building involves:

- selecting the attribute for the root node;
- splitting instances into subsets;
- repeat recursively for each branch;
- stop if all instances have achieved the same class.

A root node is selected by comparing the number of bits (splitting based on information gain) for possible root nodes and choosing the node that has the most bits of information. After selecting the root node, the next step is to look at the branches that emanate from the root node. The process of selecting the node with the most bits of information is repeated. This process continues until all instances have the same number of bits – that is, when there are no more classes to split on (accuracy is 100%). The tree pruning process involves eliminating error-prone branches. A pruned tree can improve a classifier's performance and can facilitate further analysis of the model for the purpose of knowledge acquisition. The pruning process should never remove predictive parts of the classifier. For better understanding of how decision trees work, consider the example below in relation to Figure 2:

If (Contract length $\geqslant$ 24 months, Location = midlands, price-plan < 30 and data $<$ 4gb), then Churn = Yes.

A number of algorithms can be used to build decision trees. These algorithms include CART (classification and regression trees), Chi-squared automatic interaction detector (CHAID), iterative dichotomizer (ID3), and C4.5, which is the successor of ID3. The CART algorithm is used as part of the CADET approach to ABM. Decision trees have many advantages in the binary classification context, and are popular because they are relatively easy to interpret and understand.[81] They have the ability to handle covariates, which are measured at different measurement levels[36] and they can process both numeric and categorical data. However, decision trees have some limitations; they have a high degree of instability.[82] A small change in data can result in different series of splits, which often affects the quality of prediction when validating the trained model.[36] Advantages of CART[83] also include its non-parametric nature, capability with outliers, ability to select variables of interest, and, importantly, the ability to re-use variables. These benefits provide an effective approach for agent and rules elicitation – our primary concern here. Evaluating CART against competitor methods also substantiated its effectiveness. The accuracy rates for predicting customer churn on coverage quality are 52.64% for the CART model and 50.00% for the random forest model. Finally, the random forest model produced a higher accuracy rate for predicting churn based on price fairness. The accuracy rate for the random forest model was 73.53%, while the accuracy rate for the CART model was 70.58%. A combination of relative effectiveness and underlying strengths of the approach led to CART being included in our design.

8. The CADET approach

A number of modeling techniques can and have been used to undertake ABMS. Object-oriented (OO) methodologies and knowledge-engineering methodologies are popular.[84] Three common applications of OO technology for describing ABMs are *static* for describing the structure of objects; *dynamic* for describing object interaction; and *functional* for describing the data flow of the methods of the objects.[84] Flow-charting is another popular example for representing the process flow of an ABM.[1]

Our CADET approach (summarized in Figure 3) focuses primarily on structures resulting from decision-tree analysis. Key variables must first be understood and then found in datasets. Importantly, however, the approach itself involves four central steps:

1. dataset selection and normalization;
2. decision-tree generation;
3. tree interpretation;
4. model building.

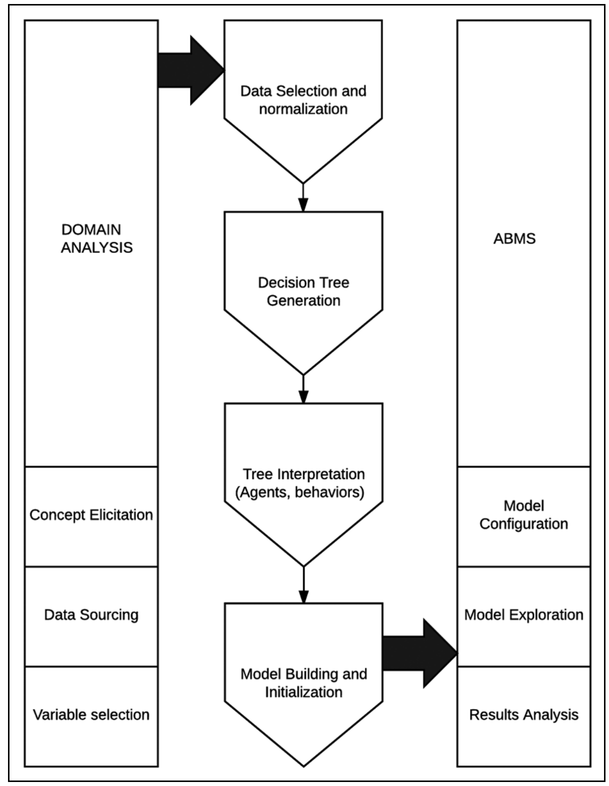

Figure 3. CADET method.

A dataset is initially chosen for a specific time period (or set of time periods) containing variables of interest. Customer, behavior, and outcome attributes are normalized for analysis. The CADET approach applies the CART algorithm on a Telco customer dataset and then to visualize the tree-like structure derived from the analysis, specifically the leaf nodes and the splitting rules. Agents are derived from unique leaves and agent attributes are derived from the decision-tree flow process. As stated earlier, data-driven approaches typically pre-select agents and use data to uncover behavior. Attributes of interest are identified in the decision tree through the splitting rules. CADET selects possible outcomes and allows the agent types to emerge from the tree itself. An initial model provides results that can be loaded into the agent-based platform. The CADET approach is also general-purpose and can be applied to various domains and industries. To further explain the structure and functionality of the CADET approach, we present a simple automotive scenario.

A Mercedes-Benz car dealership provides a number of products and services. In order to understand customer purchasing behavior, a decision-tree analysis is constructed. The analysis is carried out to improve understanding on customer purchase behavior. Trends and patterns in customer decision-making are uncovered. Analysis shows that customers who have certain attributes purchase Mercedes-Benz cars from the Mercedes-Benz car dealership. The

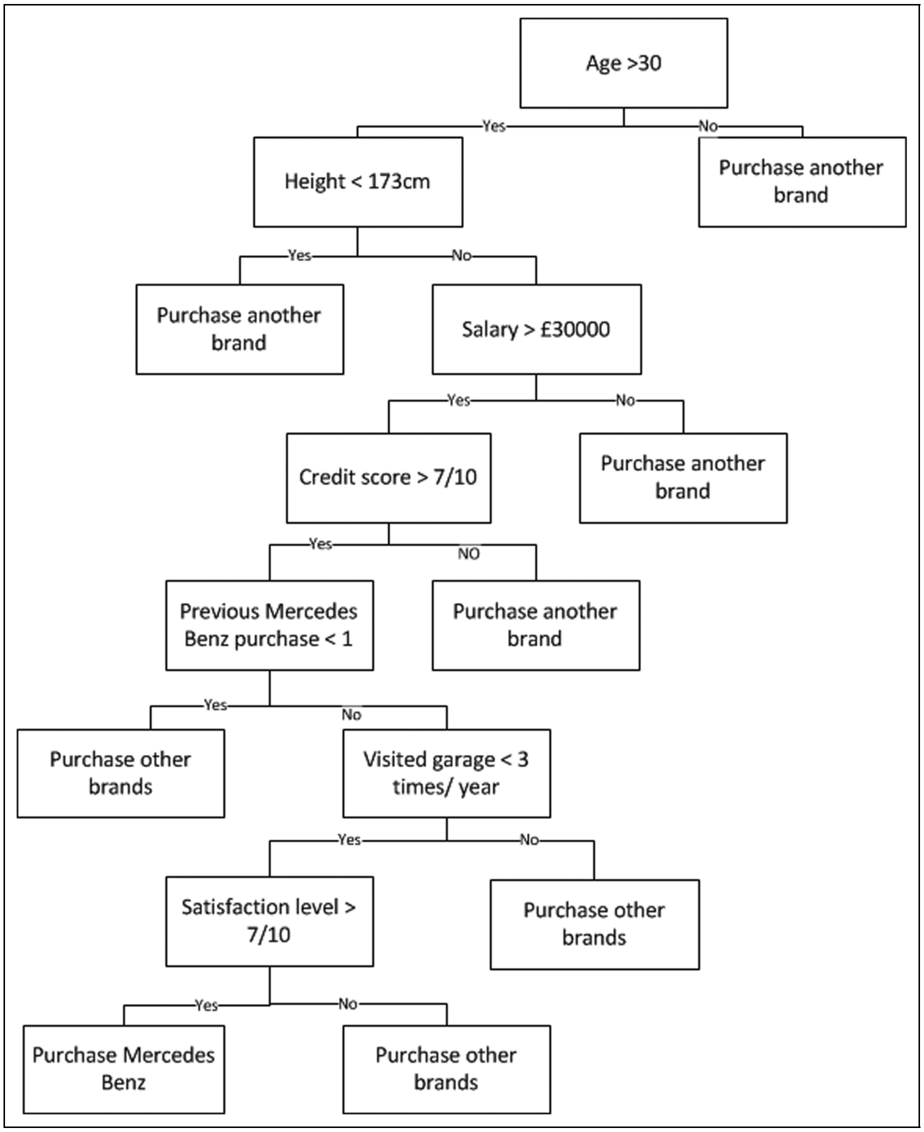

Figure 4. Automotive example description of an ABM using the CADET method.

variables used for this analysis are age, height, salary, credit score, previous Mercedes-Benz purchase, number of services/repairs, and satisfaction level. The target variable is either to purchase a Mercedes-Benz car or to purchase another car brand.

Decision-tree analysis shows that customers that are over 30 years old are often interested in purchasing a Mercedes-Benz while customers who are under 30 years old often buy other car brands. The next variable on the decision tree is height. Customers with height less than 173 cm often buy other brands, while taller customers are often interested in purchasing Mercedes-Benz cars. The next variable on the decision tree is salary. If salary is greater than 30,000, move to the next step; else, consider purchasing other brands. The next step is credit score. If credit score is greater than 7, move to the next step; else, consider purchasing other brands. This process is followed until the end of the tree. Figure 4 presents the decision-tree analysis resulting from the automotive exemplar. Later evaluation uses an ABS tool (TEA-SIM in this case) to assess the efficiency and effectiveness (from Table 1) of the CADET method with a Telco dataset by instantiating elements derived from the tree.

8.1. TEA-SIM tool

Over the years, companies have invested substantial resources in products and services that enable them to manage and understand their customer behavior more effectively.

Advertisement and WOM can be powerful tools for customer retention. However, customers are often skeptical about advertisement and may turn to their family or friends within their network to seek advice before deciding whether to purchase or repurchase a product or service. Companies have manipulated WOM by running schemes that offer customers benefits for expressing positivity about their product or service. A positive WOM is a great tool for marketing; negative WOM can be a damaging tool for businesses. Dissatisfied customers of a product or service may share their experience about their dissatisfaction with members of their social network, which can include family and friends.

CADET analysis can direct any agent modeling tool, but here we use our TEA-SIM tool because of its data-driven capabilities[85] – allowing the incorporation of cognitive processes for understanding how the members of a small-world network make decisions.[48] In addition, the TEA-SIM tool is a decision-support tool that can be adopted by various industries to model various entities such as customers, products, and services with file-based configurations. It also provides a medium for companies to see the interaction process between agents and how the interactions influence agent decisions with its web-based user interface. The results derived from this process can be used to provide information within organizations so that they can strengthen their CRM strategies, exploring the effect of WOM to improve customer retention.

TEA-SIM agent behavior is developed in PHP:Hypertext Preprocessor (PHP) and agent description is in JavaScript Object Notation (JSON) (see the initialization example in Figure 7). The JSON files can be extracted from a relational model where entities represent agent supertypes. The column names in each of the defined entities describe agent attributes. In both our automotive example and MSI evaluation we have a single customer agent supertype – table CUSTOMER. Column names define the attributes used in spitting rules (e.g., CREDIT_SCORE and SALARY for the automotive example). Two rows will be created in the table CUSTOMER to describe a Mercedes-Benz and ''other brand'' purchaser. These are agent types. JSON files are generated from this relational schema and experiments described within initialization files. Traditional initialization is available, including grid sizes and number of specific agents. Step functions defined in PHP operate on agent attributes to alter behavior as required – for example, salary changes. Once started, agents will be instantiated in memory and visualized using a web-based user interface. Setting up the simulation requires two configuration files for initialization and model execution – init.json and model.json. Files are extracted from a relational representation. The init.json file is where the initialization is carried out. Grid size, number of instances for each agent, the position of each agent, and the simulation steps are defined in init.json. Agent attributes of class MercedesBenzPurchaser are defined in model.json with behavior(step functions) in steps.php. Agent behavior specification is described in a module called steps.php. The stepping function, step() describes how each agent of a specific type progresses from one iteration to another, including state changes. In steps.php, agent states or internal rules are being processed. See Figure 9 for the rules applied to the MSI evaluation with churn and stay customers, and Figure 10 for depiction of an example web visualization. Rules are easily implemented using PHP syntax, including morphing into another agent or state change.

The dynamic nature of TEA-SIM initialization provides a unique approach to understanding customer behavior. It can be used to observe the pattern of customer interaction and perform further analytics to explore the possibilities of incorporating those patterns into marketing strategies in order to increase revenues. The data-driven nature of the agents and behavior also allow the modeler to generate several models for different data extracts in order to explore and compare a number of unique scenarios. The TEA-SIM tool also works as a generic model that captures the key drivers behind customer change of behavior and it can also work well in a consultancy environment – configuring the model using the web interface. It is a cross-industry tool that is not unique to any industry. Experimenting with the TEA-SIM tool improves understanding of factors that can drive customer churn. The TEA-SIM tool is not a precise prediction tool, however, primarily providing insight into the behavior of a population of customers.

8.2. Applying the CADET method to a real-world dataset

Models are built for the purpose of mimicking real-world events. The CADET method for conducting agent-based modeling is a novel approach for building ABMs using decision trees. In this study, the decision-tree analysis is used to describe customers that either remain with their MNO after their contract period and customers that leave their MNO after their contract period. The CADET method shows that customers' decision to stay or leave their MNO is driven by a set of subjective attributes. The attributes are the individual nodes on the decision tree. Mobile network customers may have attributes such as mobile phone type, customer location, price plan, and data usage. The decision to churn or stay is composed of a number of phases represented on the decision tree. The leaf node of the decision tree is the customer's final decision to stay with or leave an MNO. Decision trees consist of dependent and independent variables. The dependent variables are the nodes that make up the final node. These variables influence customers' final decision. The primary purpose of applying the CADET method for ABMS using the TEA-SIM

Table 2. Dataset description.

Name	Description
Contract length	Length of contract
Gender	Customer's gender
Sales channel	Company that delivered the contract
Postcode	Postcode in which the customer lives
County name	The name of the county where the customer lives
Region	The region where the customer lives
Devices	Name and model of device used by the customer
Tenure	Number of months with the present mobile service provider
Life stage segment	Customer age
Number of complaints	Number of complaints throughout the contact period
Q2_bytes	Second quarter data usage
Q3_bytes	Third quarter data usage
Q2_voice	2nd quarter voice usage
Q3_voice	3rd quarter voice usage
No_of_repairs	Number of times the customer's phone has been repaired
Prob_handset	Number of times the customer has reported problems with handset

application is to understand how much influence a customer's environment, family, and friends within their network have on customer retention. In addition, the CADET method utilized with the TEA-SIM model provides information on the possible decisions a customer might make when they interact with other customers within their network or environment. We believe that if a customer meets another customer within the environment, and they share the same MNO, they may have a conversation on their overall customer experience and that conversation may influence a customer's decision to renew their contract. The CADET method of agent-based modeling seeks to provide an understanding of how customer variables, along with customer network, can influence customer retention. The next section presents a description of the dataset used to conduct the experiment in this study.

8.3. Dataset description

To demonstrate the usability of the CADET approach, we collected a dataset from a UK telecommunications company. The dataset comprises 19,919 observations and 16 variables. The dependent variable (output variable) for this dataset is whether the customer churned or stayed with their MNO after their contract expired. The predictor variables (input variables) are customers' data (such as type of device, price plan, and region). The dataset contains 50% of customers who churned and 50% of customers who stayed with the MNO until the end of their contract. The dataset is based on a 24-month contract. Some customers stayed with the MNO after the end of the 24-month period (i.e., they renewed their contract) while other customers left the MNO after the 24-month contract. The dataset comprises different data types as a means to represent the entire customer base. Table 2 presents a tabular description of the dataset. The next section presents details of the experiment conducted.

Data preparation consisted of data cleaning, data integration, data transformation, data reduction, and data discretization.[86] Data cleaning involves filling in missing values, identifying or removing outliers, and resolving inconsistencies. Data cleaning is usually performed in two iterative phases: discrepancy detection and data transformation. The Telco data used was of reasonable quality; however, even with limited cleaning a better understanding of the data results. Data integration was not required for this dataset. Data-reduction techniques can be applied to a dataset to obtain a reduced representation of the dataset that is much smaller in volume, yet closely maintaining integrity of the original data. Mining the reduced dataset should produce the same (or almost the same) result. Data transformation is converting data to a form that is suitable for appropriate algorithm processing. Data discretization is the process of transforming continuous data attribute values into a finite set of intervals with minimal loss of information. In this study, data cleaning was conducted with the following steps:

1. filling in missing values;
2. resolving inconsistencies;
3. identifying and removing outliers.

Filling in missing values and resolving inconsistencies was achieved by predicting missing values with the rest of the dataset. The process of identifying and removing outliers involves removing unimportant and unreliable objects. Variables were removed from the list when displaying little variation and therefore would not help in generating the best model. Data preparation should be carried out such that it suits the requirements for selected

Table 3. Steps for the decision-tree analysis.

Step 1	If tenure is $>$ or $<$ 24 consider the next node.
Step 2	If tenure is $\not>$ 24 and tenure $<$ 14, churn; else consider the next node.
Step 3	If tenure $\not>$ 24, and tenure $\not>$ 24; and device = Samsung then churn, else renew contract.
Step 4	If tenure $>$ 24, and contract length $\not>$ 12, then churn; else consider the next node.
Step 5	If tenure $>$ 24, and contract length $>$ 12, and region = Midlands, then churn; else consider the next node.
Step 6	If tenure $>$ 24, and contract length $>$ 12, and region $\neq$ Midlands, and region $\neq$ London, then churn; else consider the next node.
Step 7	If tenure $>$ 24, and contract length $>$ 12, and region $\neq$ Midlands, and region = London, and device = iPhone 5, then churn; else move to the next node.
Step 8	If tenure $>$ 24, and contract length $>$ 12, and region $\neq$ Midlands, and region = London, and device = iPhone 5, and problem with handset $>$ 3, and gender = male, then churn; else renew contract.
Step 9	If tenure $>$ 24, and contract length $>$ 12, and region $\neq$ Midlands, and region = London, and device = iPhone 5, and problem with handset $\not>$ 3, and number of complaints $>$ 2, then churn; else renew contract.

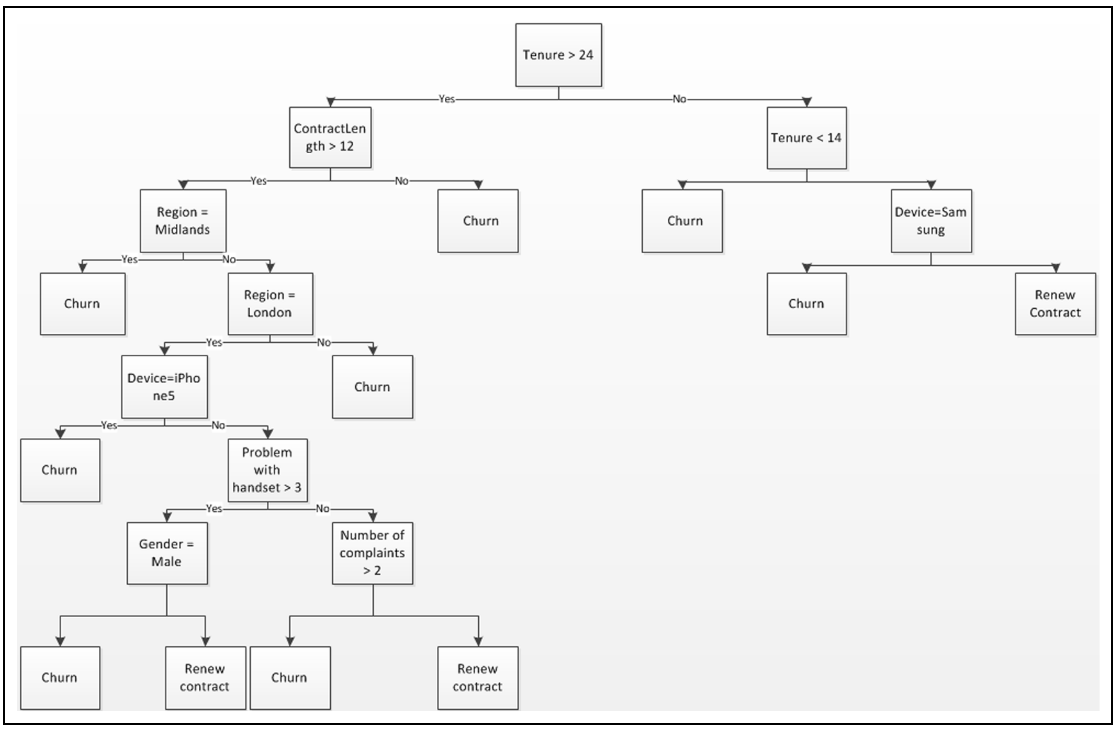

Figure 5. Decision-tree analysis.

algorithms, including more pragmatic formatting aspects. Simple comma-separated values were inputs to our R-based algorithms.

8.4. Model structure and experiment

Table 3 outlines the steps in the decision-tree analysis. To apply the CADET method with the TEA-SIM tool, the CART decision-tree algorithm is run on the dataset described above and the result of the decision tree is visualized. From the decision tree (see Figure 5), we can see the flow of the decision process for customers who have decided to stay with their MNO or move to a different MNO. Leaves of the decision tree are used to identify the required agent types (e.g., churn or renew contract) and the earlier branching directs specific agent attributes and behavior. The top of decision-tree analysis shows that a customer's tenure is either greater than 24 or not. If tenure is greater than 24 then move to the next variable on the left. However, if tenure is not greater than 24, then move to the next variable on the right-side of the tree. This process continues until the end of the tree. The end of the tree displays the end result of the process – that is either customer churned or renewed their contract at the end of the 24-month contract period. Figure 5 displays the decision-tree analysis diagram.

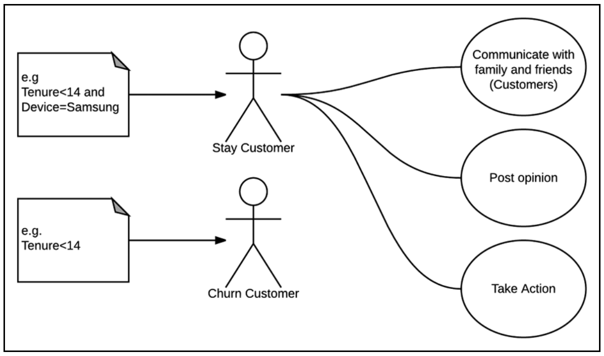

Figure 6. Customer behavior exploration.

```
{
"grid" :
        { "n" : 10, "m" : 10 }, "agents" :
        [ { "type" : "customers.1", "instances" : 6, "position" : "rnd"},
        { "type" : "customers.2", "instances" : 5, "position" : "rnd" } ],
        "simulation" : { "start" : 0, "end" : 25 } }
```

Figure 7. Init.json file.

To simulate this process, agents are fed into the TEA-SIM tool as JSON data files. Simplified agent behavior within the ABMS environment is illustrated in Figure 6. Agents typically communicate with other agents in the simulation environment and take actions after evaluating services provided by their MNO. To conduct the ABMS experiment, three files are created and run on the TEA-SIM tool. The files are init.json, model.json, and steps.php. The JSON files describe agent attributes and the ABMS environment. The PHP file describes the interaction process of the agents.

The init.json file (see Figure 7) is composed of the following attributes: grid, agent and simulation. n and m under "grid" represent the number of rows and columns required for the ABMS experiment. "Agents" is composed of agent type, instances and the position of the agents. The init.json files show that the instances are six in number and the position is random. "Simulation" on the init.json file represent the number of runs for the experiment. In this experiment, "start" is 0 and "end" is 25. This means that the agents in the grid should move 25 times.

The model.json file (see Figure 8) describes customer type and attributes. There are two types of customers with customer id 1 and customer id 2. Customer id 1 represent churn customers while customer id 2 represent stay customers. From the decision-tree analysis carried out (Figure 5), churn customers have the following attributes:

tenure >14, device = Samsung, contract >12 months, region = Midlands, region $\neq$ London, device = iPhone5, problem with handset >3, gender = male, number of complaints >2.

Furthermore, stay customers have the following attributes:

tenure >24, contract length $>$ 12, region $\neq$ Midlands, region = London, device $\neq$ iPhone5, problem with handset $<$ 3, gender = female, number of complaints $<$ 2.

The steps.php file (Figure 9) describes agent interaction. To summarize the code, if a churn customer meets a stay customer, the stay morphs to a churn customer. The crying face in Figure 10 represents churn customer and the smiley face represent stay customers. Overall, this experiment shows how customer behavior can be influenced by the environment. Figure 10 demonstrates the effectiveness and utility of the method. Further experimentation is possible. Using friend networks, agents move from one grid to another and their decisions are based on the interaction with family/friends, as represented in Figure 10.

9. Artifact discussion

The CADET method was evaluated using an experimental instantiation with a Telco dataset. The agent architecture, agent attributes and behavior, and the model structure are explained. In DSR the evaluation process is carried out to establish the design effectiveness of the CADET method to ABMS. Our initial aim was to uncover agents, attributes, and behavior from customer data for practical use in an ABM. In the experiment, agents represented mobile subscribers. Mobile subscribers were modeled to interact with other customers in the environment, thereby establishing the effect of social influence on customer retention. Agents and rules were extracted and loaded into the simulation (see Figures 7 and 8). The TEA-SIM tool provided a social environment, demonstrating how customers can be influenced by their social circle. Validating the CADET method with the TEA-SIM tool provides insights on the impact of social influence of customer retention. In addition, it demonstrates how WOM can be a vital tool for customer retention.

The CADET method can be applied to both structured and unstructured datasets. The CADET framework is clearly able to describe agents and their attributes ready for transformation into an ABMS experiment. Furthermore, the decision tree is able to identify attributes of interest and therefore provide a filter on source data. ABMS experimentation was conducted as a validation for utilizing the CADET framework as a data-driven approach to ABM. Traditional validation, where results from simulations are contrasted to test datasets, were also carried out. However, the aim here is to present an approach to more

```
{
"customers" : [
{ "id" : 1, "name": "ChurnCustomers",
"attributes":[ {"name" : “tenure”, "value" : < 14 months} ,
{ "name" : “device”, "value" : samsung } ,
{"name" : “contract”, "value" : < 12 months} ,
{ "name" : “region”, "value" : midlands } ,
{"name" : “region”, "value" : !London} ,
{ "name" : “device”, "value" : iPhone5 } ,
{"name" : “problemsWithHandset”, "value" : > 3 } ,
{ "name" : “gender”, "value" : male } ,
{"name" : “number of complaints”, "value" : > 2} ,
{ "name" : "_img" , "value" : "img/ChurnCustomer.jpeg" } ] },


{"id": 2, "name": "RenewCustomers",
"attributes":[{ "name" : “tenure”, "value" : > 24},
{ "name" : "churn", "value" : false } ,
{ "name" : “contract length“, "value" : > 12},
{ "name" : "region", "value" = !Midlands} ,
{ "name" : “region”, "value" : London } ,
{ "name" : "device", "value" : !iphone5} ,
{ "name" : “problemsWithHandset” , "value" : < 3 } ,
{ "name": “gender”, “value” : female } ,
{"name" : “number of complaints”, “value”: < 2} ,
{ "name" : "_img" , "value" : "img/RenewCustomer.jpeg" } ] }]}
```

Figure 8. Model.json.

exploratory agent-based experiments where snapshots of market activity and external forces can be tested. The TEA-SIM tool provided a generic approach for companies across industries to better understand their customers. In addition, the TEA-SIM tool also helps decision-makers in organizations to work on strategies to understand customer behavior, enhance customer retention, and subsequently improve CRM. A real-world example is presented, implementing the CADET method using the TEA-SIM tool. The CADET approach was used to effectively provide an ABM using the results derived from decision trees. Although applied to a dataset from a Telco company, the CADET approach can be extended and implemented in other industries such as healthcare, manufacturing, and financial services. Furthermore, the CADET framework provided concrete insight into customer retention in this study. In particular, the experimental instantiation demonstrates the possible impact of WOM as an important factor for

```php
<?php

class customers_ChurnCustomers extends Agent {
    function step($step) {
            if (!$this->churn && $this->anyNeighbour (1, 2, array("churn"=>false),
array("churn" => true, "_img" => "img/ChurnCustomer.jpeg"))) {
                    $this->churn = true;
                    $this->_img = "img/ChurnCustomer.jpeg";
            }
            $this->move(1);
            $this->morph(2);

    }
}

class customers_StayCustomers extends Agent {
    function step($step) {
        $this->move(1);
            }
        }
```

Figure 9. Stepper function for dataset.

customer retention. In terms of describing ABMs, the CADET approach has the following advantages: (1) it provides a clear and intuitive way to describing agents and their attributes while explaining the significance of relationships between variables; (2) it can take into account

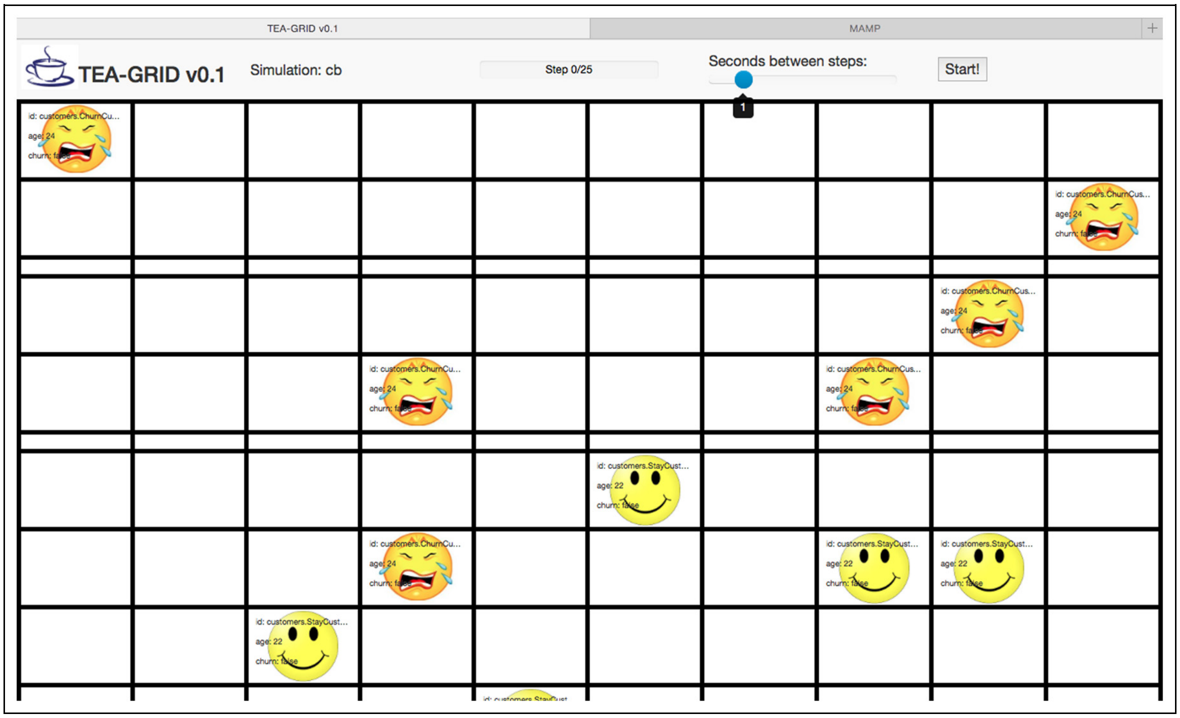

Figure 10. Agent interaction process.

the heterogeneity of customers to capture emergent phenomena; (3) it is flexible and can easily adapt to new constraints; and (4) in addition to its flexibility, the CADET approach can also be applied more generally in any industry with detailed customer datasets.

Applying the CADET framework to an ABMS tool such as TEA-SIM can provide actionable insights into the social effects of customer retention. Results demonstrate that ABMS can provide an effective approach to the exploration of inter-customer dynamics, which have so far been ignored, while still providing a robust data-driven underpinning. In addition, the simulation itself demonstrates likelihood of a customer to churn when the number of neighbors who churn increase.[87] Figure 6 depicts CART as part of an ABMS workflow that synthesizes initial agent and attribute results with inter-customer dynamics. Agents are uncovered from Telco data and then further extended with social behaviors.

10. Conclusion

Customer retention is extremely important to MNOs because of fierce competition in the mobile sector. Hence, companies in the sector are increasingly strengthening their CRM strategies in order to retain customers. A number of factors can influence the decision for customers to purchase or adopt a product or service. However, customers are also likely to trust the WOM from someone within their social network. Tools and techniques are required to build models in near real-time and also allow companies to explore external factors. Machine learning and ABMS have both been widely adopted to better understand customer behavior. However, their combination as a means to generate models using only market determinants is novel. Consequently, models can be generated in a timely manner and in response to market change. It is this level of automation that allows companies to explore possible future scenarios or active interventions. This paper presents the CADET method, a novel data-driven approach to ABMS, that addresses the effective utilization of large and rapidly changing datasets. Agents are not predefined before data analysis activities, determinants are used to uncover both agents and behaviors automatically.

March and Smith describe artifacts as constructs, models, methods, and instantiations to delineate core research activities. Artifacts naturally contribute to the process of modeling and simulation primarily as a method for model building, the model itself, and as an instantiated simulation. This study presents the CADET method as a data-driven framework for initial building (or initialization) activities – identifying agents, attributes of interest and customer behaviors. A Telco customer model is generated from data and then instantiated in order to evaluate the CADET method utilizing a data-driven ABMS tool. The application of the CADET method clearly demonstrates the effectiveness of data-driven processes as part of an ABMS industry experiment. Agents, attributes, and behaviors are uncovered before transformation into the ABM. Although applied here in the mobile telecommunication sector, future work could easily be applied more widely with only determinants and customer datasets required upfront.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Grimm V, Berger U, Bastiansen F, et al. A standard protocol for describing individual-based and agent-based models. *Ecol Model* 2006; 198(1): 115–126.
2. Ozcan A. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab Chip* 2014; 14(17): 3187–3194.
3. Verbeke W, Dejaeger K, Martens D, et al. New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. *Eur J Oper Res* 2012; 218(1): 211–229.
4. Keramati A and Ardabili SM. Churn analysis for an Iranian mobile operator. *Telecomm Policy* 2011; 35(4): 344–356.
5. Hughes AM *Strategic database marketing*. New York: McGraw-Hill, 2006.
6. Jones MA, Taylor VA and Reynolds KE. The effect of requests for positive evaluations on customer satisfaction ratings. *Psychol Mark* 2014; 31(3): 161–170.
7. Ganesh J, Arnold MJ and Reynolds KE. Understanding the customer base of service providers: an examination of the differences between switchers and stayers. *J Mark* 2000; 64(3): 65–87.
8. Blodgett JG, Wakefield KL and Barnes JH. The effects of customer service on consumer complaining behavior. *J Serv Mark* 1995; 9(4): 31–42.
9. Itzkowitz J. Customers and cash: how relationships affect suppliers' cash holdings. *J Corp Finance* 2013; 19: 159–180.
10. Rust RT and Zahorik AJ. Customer satisfaction, customer retention, and market share. *J Retail* 1993; 69(2): 193–215.
11. Epstein MJ and Westbrook RA. Linking actions to profits in strategic decision making. *MIT Sloan Manag Rev* 2001; 42(3): 39.
12. Baumann C, Elliott G and Burton S. Modeling customer satisfaction and loyalty: survey data versus data mining. *J Serv Mark* 2012; 26(3): 148–157.
13. Johnson MD, Herrmann A, Huber F, et al. *Customer retention in the automotive industry: quality, satisfaction and loyalty*. New York: Springer Science & Business Media, 2012.
14. Fullerton G. The moderating effect of normative commitment on the service quality–customer retention relationship. *Eur J Mark* 2014; 48(3/4): 657–673.
15. Ribeiro Soriano D, Jyh-Fu Jeng D and Bailey T. Assessing customer retention strategies in mobile telecommunications: hybrid MCDM approach. *Manage Decis* 2012; 50(9): 1570–1595.
16. Khan AA, Jamwal S and Sepehri M. Applying data mining to customer churn prediction in an internet service provider. *Int J Comput Appl* 2010; 9(7): 8–14.
17. Hanif M, Hafeez S and Riaz A. Factors affecting customer satisfaction. *IRJFE* 2010; 60: 44–52.
18. Silva KA and Yapa STWS. Customer retention: with special reference to telecommunication industry in Sri Lanka. In: *Proceedings of the international conference on business and information, (BAI 2009)*, Kuala Lumpur, 6–8 July 2009, pp.1–15. Kuala Lumpur, Malaysia: Academy of Taiwan Information Systems Research (ATISR). 2013.
19. Svendsen GB and Prebensen NK. The effect of brand on churn in the telecommunications sector. *Eur J Mark* 2013; 47(8): 1177–1189.
20. Chu B-H, Tsai M-S and Ho C.-S. Toward a hybrid data mining model for customer retention. *Knowl-Based Syst* 2007; 20(8): 703–718.
21. Phadke C, Uzunalioglu H, Mendiratta VB, et al. Prediction of subscriber churn using social network analysis. *Bell Labs Tech J* 2013; 17(4): 63–75.
22. Verbeke W, Martens D and Baesens B. Social network analysis for customer churn prediction. *Appl Soft Comput* 2014; 14: 431–446.
23. Richard M.-O and Chebat J.-C. Modeling online consumer behavior: preeminence of emotions and moderating influences of need for cognition and optimal stimulation level. *J Bus Res* 2016; 69(2): 541–553.
24. Vag A. Simulating changing consumer preferences: a dynamic conjoint model. *J Bus Res* 2007; 60(8): 904–911.
25. Rygielski C, Wang J-C and Yen DC. Data mining techniques for customer relationship management. *Technol Soc* 2002; 24(4): 483–502.
26. Ottar Olsen S, Alina Tudoran A, Brunsø K, et al. Extending the prevalent consumer loyalty modelling: the role of habit strength. *Eur J Mark* 2013; 47(1/2): 303–323.
27. Xu M and Walton J. Gaining customer knowledge through analytical CRM. *Ind Manage Data Syst* 2005; 105(7): 955–971.
28. Neslin SA, Gupta S, Kamakura W, et al. Defection detection: measuring and understanding the predictive accuracy of customer churn models. *J Mark Res* 2006; 43(2): 204–211.
29. Furness P. Techniques for customer modelling in CRM. *Journal of Financial Services Marketing* 2001; 5(4): 293–307.
30. Huang Y and Kechadi T. An effective hybrid learning system for telecommunication churn prediction. *Exp Syst Appl* 2013; 40(14): 5635–5647.
31. Jahromi AT, Stakhovych S and Ewing M. Customer churn models: a comparison of probability and data mining approaches. In Campbell C and Ma JJ (eds) *Looking forward, looking back: drawing on the past to shape the future of marketing*. New York: Springer, 2016, pp.144–148.
32. Zhang Z, Wang R, Zheng W, et al. Profit maximization analysis based on data mining and the exponential retention model assumption with respect to customer churn problems. In *Data mining workshop (ICDMW), 2015 IEEE*

international conference, Atlantic, NJ, 14–17 November 2015, pp. 1093–1097. New York: IEEE.
33. Ye Y, Liu S and Li J. A multiclass machine learning approach to credit rating prediction. In *Information processing (ISIP), 2008 international symposiums*, Moscow, May 23, 2008, pp. 57–61. New York: IEEE.
34. Lemmens A and Croux C. Bagging and boosting classification trees to predict churn. *J Mark Res* 2006; 43(2): 276–286.
35. Huang B, Kechadi MT and Buckley B. Customer churn prediction in telecommunications. *Exp Syst Appl* 2012; 39(1): 1414–1425.
36. Burez J and Van den Poel D. Separating financial from commercial customer churn: a modeling step towards resolving the conflict between the sales and credit department. *Exp Syst Appl* 2008; 35(1): 497–514.
37. Lee C, Kwak N and Lee C. Understanding consumer churning behaviors in mobile telecommunication service industry: cross-national comparison between Korea and China. In: *Proceeding of the international conference on information systems*, Fort Worth, TX, 2015. AIS.
38. He B, Shi Y, Wan Q, et al. Prediction of customer attrition of commercial banks based on SVM model. *Procedia Comput Sci* 2014; 31: 423–430.
39. Raju PS, Bai VR and Chaitanya GK. Data mining: techniques for enhancing customer relationship management in banking and retail industries. *IJIRCCE* 2014; 2(1): 2650–2657.
40. Sundarkumar GG and Ravi V. A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. *Eng Appl Artif Intell* 2015; 37: 368–377.
41. Miguéis VL, Van den Poel D, Camanho AS, et al. Modeling partial customer churn: on the value of first product-category purchase sequences. *Exp Syst Appl* 2012; 39(12): 11250–11256.
42. Miguéis VL, Camanho A and e Cunha JF. Customer attrition in retailing: an application of multivariate adaptive regression splines. *Exp Syst Appl* 2013; 40(16): 6225–6232.
43. Richter Y, Yom-Tov E and Slonim N. Predicting customer churn in mobile networks through analysis of social groups. in *Proceedings of the 2010 SIAM international conference on data mining 2010*, Columbus, OH, 29 April 2010, pp. 732–741. Philadelphia, PA: Society for Industrial and Applied Mathematics.
44. Scott J. *Social network analysis*. London: Sage, 2012.
45. Hassouna M, Tarhini A, Elyas T, et al. Customer churn in mobile markets: a comparison of techniques. *Int Bus Res* 2015; 8(6): 224.
46. Watts DJ. The ''new'' science of networks. *Annu Rev Sociol* 2004; 30: 243–270.
47. Macal CM and North MJ. Tutorial on agent-based modelling and simulation. *J Simulat* 2010; 4(3): 151–162.
48. Milgram S. The small world problem. *Psychol Today* 1967; 2(1): 60–67.
49. Baxter N, Collings D and Adjali I. Agent-based modelling: intelligent customer relationship management. *BT Tech J* 2003; 21(2): 126–132.
50. Zhang X, Liu Z, Yang X, et al. Predicting customer churn by integrating the effect of the customer contact network. In *2010 IEEE International conference on service operations and logistics and informatics (SOLI)*, Qing Dao, 15 July, 2010, pp. 392–397). New York: IEEE.
51. Risselada H, Verhoef PC and Bijmolt TH. Dynamic effects of social influence and direct marketing on the adoption of high-technology products. *J Mark* 2014; 78(2): 52–68.
52. Luo Q and Zhong D. Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tourism Manage* 2015; 46: 274–282.
53. Anderson EW. Customer satisfaction and word of mouth. *J Serv Res* 1998; 1(1): 5–17.
54. Keaveney SM. Customer switching behavior in service industries: an exploratory study. *J Mark* 1995; April: 71–82.
55. Hassouna MB. Agent based modelling and simulation: an examination of customer retention in the UK mobile market. PhD dissertation, Brunel University, School of Information Systems, Computing and Mathematics, UK, 2012.
56. Twomey P and Cadman R. Agent-based modelling of customer behaviour in the telecoms and media markets. *info* 2002; 4(1): 56–63.
57. North MJ, Macal CM, Aubin JS, et al. Multiscale agent-based consumer market modeling. *Complexity* 2010; 15(5): 37–47.
58. Farmer JD and Foley D. The economy needs agent-based modelling. *Nature* 2009; 460(7256): 685–686.
59. Epstein JM. Modelling to contain pandemics. *Nature* 2009; 460(7256): 687.
60. Adami C. *Introduction to artificial life*, vol. 1. New York: Springer Science & Business Media, 1998.
61. Heppenstall AJ, Crooks AT, See LM, et al. *Agent-based models of geographical systems*. New York: Springer Science & Business Media, 2011.
62. Zhang T and Zhang D. Agent-based simulation of consumer purchase decision-making and the decoy effect. *J Bus Res* 2007; 60(8): 912–922.
63. Gilbert GN. *Agent-based models*. New York: Sage, 2008.
64. Roozmand O, Ghasem-Aghaee N, Hofstede GJ, et al. Agent-based modeling of consumer decision making process based on power distance and personality. *Knowl-Based Syst* 2011; 24(7): 1075–1095.
65. Kennedy C, Theodoropoulos G, Sorge V, et al. Aimss: an architecture for data driven simulations in the social sciences. In: *Computational science: ICCS 2007*, Beijing, China, May 27–30, 2007, pp. 1098–1105. New York: Springer.
66. Hassan S, Antunes L and Pavón J. Mentat: a data-driven agent-based simulation of social values evolution. *MABS* 2009; 9: 135–146.
67. An L. Modeling human decisions in coupled human and natural systems: review of agent-based models. *Ecol Model* 2012; 229: 25–36.
68. Gimblett HR. *Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes*. New York: Oxford University Press, 2002.
69. Deadman P, Robinson D, Moran E, et al. Colonist household decisionmaking and land-use change in the Amazon rainforest: an agent-based simulation. *Environ Plan B* 2004; 31: 693–710.

70. Vaishnavi VK and Kuechler W. *Design science research methods and patterns: innovating information and communication technology*. Boca Raton, FL: CRC Press, 2015.
71. von Alan RH, March ST, Park J, et al. Design science in information systems research. *MIS Quarterly* 2004; 28(1): 75–105.
72. March ST and Smith GF. Design and natural science research on information technology. *Decis Support Syst* 1995; 15(4): 251–266.
73. Newell A and Simon HA *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972.
74. Peffers K, Tuunanen T, Rothenberger MA, et al. A design science research methodology for information systems research. *J Manag Inf Syst* 2007; 24(3): 45–77.
75. Kuechler B and Vaishnavi V. On theory development in design science research: anatomy of a research project. *Eur J Inf Syst* 2008; 17(5): 489–504.
76. Hassouna M and Arzoky M. Agent based modelling and simulation: toward a new model of customer retention in the mobile market. In: *Proceedings of the 2011 summer computer simulation conference*, The Hague, Netherlands, June 27–30, 2011, pp. 30–35. Vista, CA: Society for Modeling & Simulation International.
77. Barnes JG and Howlett DM. Predictors of equity in relationships between financial services providers and retail customers. *Int J Bank Mark* 1998; 16(1): 15–23.
78. Rokach L and Maimon O. *Data mining with decision trees: theory and applications*. Singapore: World Scientific Press, 2014.
79. Kotsiantis SB. Decision trees: a recent overview. *Artif Intell Rev* 2013; 39(4): 261–283.
80. Tsai C.-F and Lu Y.-H. Data mining techniques in customer churn prediction. *Recent Pat Comput Sci* 2010; 3(1): 28–32.
81. Duda RO, Hart PE and Stork DG. *Pattern classification*. Hoboken, NJ: John Wiley & Sons, 2012.
82. Friedman J, Hastie T and Tibshirani R. *The elements of statistical learning*, vol. 1. Berlin: Springer, 2001.
83. Nisbet J, Elder R and Miner G. *Handbook of statistical analysis and data mining applications*. London: Academic Press, 2009.
84. Iglesias CA, Garijo M and González JC. A survey of agent-oriented methodologies. In Singh MP, Rao AS and Muller JP, *Intelligent agents V: Agents theories, architectures, and languages*. New York: Springer, 1999, pp. 317–330.
85. Bell D, Kashefi A, Saleh N, et al. A data-driven agent based simulation platform for early health economics device evaluation. In *Proceedings of the agent-directed simulation symposium*, Pasadena, CA, 3 April 2016, p. 5. Vista, CA: Society for Computer Simulation International.
86. Han J, Kamber M and Pei J. *Data mining: concepts and techniques*. Amsterdam: Elsevier, 2011.
87. Dasgupta K, Singh R, Viswanathan B, et al. Social ties and their relevance to churn in mobile telecom networks. In: *Proceedings of the 11th international conference on extending database technology: advances in database technology*, Nantes, 25 March 2008, pp. 668–677. New York: ACM.

Author biographies

Dr. David Bell is a Senior Lecturer at Brunel University London. His research interests lie at the intersection of user experience, modeling and simulation - where intelligent data analysis directs digital service design. Recent funded research projects have applied this interests to urgent care health simulation, augmented museum experiences and personal data trading.

Dr. Chidozie Mgbemena has a PhD in Computer Science from Brunel University London. His doctoral research focused on applying techniques in Machine Learning and Agent Based Modeling and Simulation to addressing customer retention in the telecommunications industry. He currently works as a Data Quality Analyst at SAS Software Limited.