December 19, 2018

Machine learning to optimize controlled experiments

Neil Hoyne

Machine learning to optimize controlled experiments

ABSTRACT

Businesses use pre-sales questionnaires to quantify the value of incoming leads. Results from such questionnaires can improve allocations of sales resources and optimize advertising campaigns. A given questionnaire may ask for dozens of pieces of information from a prospective client. However, it is well understood that the questions are not always predictive of the likelihood that a client completes a purchase.

This disclosure provides machine-learning techniques that discover the predictive ability of questions on a questionnaire. The techniques enable a business to select or design questions or experiments that correlate with the closure of a sale. The business benefits by improving lead quality and optimizing sales and advertising resources.

KEYWORDS

- pre-sales
- questionnaire
- survey design
- lead quality
- customer relationship management (CRM)
- machine learning

BACKGROUND

Businesses, e.g., engaged in business-to-business (B2B) sales, in the financial sector, etc. use pre-sales questionnaires to quantify the value of incoming leads. Results from such questionnaires can improve allocations of sales resources and optimize advertising campaigns. For example, a financial services company may ask for several pieces of information from potential customers, e.g., loan size, intended use, credit score, payback window, etc. Results from the questionnaire may be used to assign priorities, e.g., customers that apply for large loans may receive priority callbacks.

Any particular questionnaire may ask for a large number of pieces of information from a prospective client. However, it is possible that the data points obtained do not predict the likelihood that a customer will take a particular action later on, e.g., complete a transaction such as taking a loan. Further, each additional question or step in the lead process can reduce the likelihood of a target customer finishing the questionnaire. The business is therefore confronted with the problem of selecting a limited number of questions designed to collect relevant information, e.g., information that correlates with the likelihood of a sale, or that can drive optimization in operations, sales, or ad targeting.

DESCRIPTION

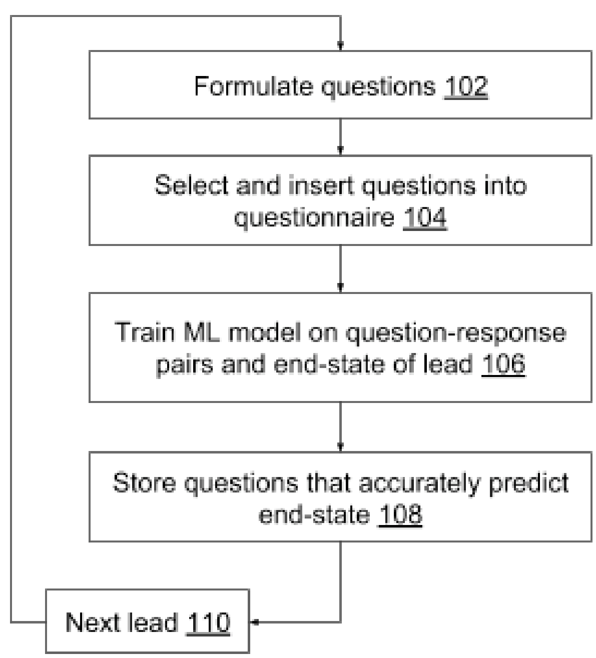

Fig. 1: Machine learning to optimize controlled experiments

Fig. 1 illustrates utilization of machine learning to optimize controlled experiments (e.g., surveys, questionnaires, etc.), per techniques of this disclosure. A list of potential questions for a customer survey or questionnaire is formulated (102). The questions are industry specific. For example, a lending company may ask questions related to credit score, marital status, account balances, etc., but may not ask for other information, e.g., gender. A dating company, however, may ask for a customer's gender, but not financial information. Certain questions, e.g., income or credit score, may be deemed sensitive by respondents and are often found to have a lower response rate. A business that serves the questionnaire (e.g., as a lead form) can set up certain questions as optional, allowing responding customers to opt out of answering such questions. Additionally, the business can set thresholds that determine a customer's pathway through the questionnaire. The business reviews and validates the questions. Customer responses to the questions can be augmented by searching for publicly available information about the customer.

A module or script inserts selected questions dynamically into the lead form (104). The questions may be inserted into particular slots. For example, when the lead form loads, the module calls a backend process that selects from available questions and dynamically inserts the selected questions into the lead generation process. On occasion, the module opts one or more customers out of dynamically inserted questions to form a control set. Alternative to dynamic insertion of questions into particular slots, the module can manage the entire process of serving the questionnaire and receiving the responses. The business that serves the lead form sets the parameters for experiment design, e.g., duration, sensitivity, thresholds, etc.

With user permission, a machine learning model is trained (106) on question-response pairs and the end-state of the lead (e.g., did the lead eventually take out the loan?). Upon training, the model discovers questions that predict customer behavior, e.g., customer's likelihood to complete a purchase.

A secondary process further guides training as follows. If a particular question scores higher than baseline (control) questions but has not reached a sufficiently high confidence level, then such a question is presented more frequently in future surveys. In this manner, the machine learning model iteratively seeks continual improvement over a predictive question set. The machine learning model can be, e.g., a generative machine learning model, a regression learning model, a neural network, etc. Example types of neural networks that can be used include long short-term memory (LSTM) neural networks, recurrent neural networks, convolutional neural networks, etc. Other machine learning models, e.g., support vector machines, random forests, boosted decision trees, etc., can also be used.

Results of the experiment, e.g., questions that predict the end-state of the lead, are stored (108), with user permission, for the purpose of serving to future leads. Future experiments thereby benefit from what the learnings of previous trials. Additionally, a second machine learning model identifies potential commonalities between highly predictive questions to provide greater precision or speculatively generate different questions not previously used. For example, in the lending company scenario described above, questions about current financial condition may be found more predictive than questions about intended use of a loan. Commonalities between questions can be used to generate new and different questions.

The trained models can be used across multiple customers, e.g., advertisers that provide the survey. Use of survey-specific data to train the models is made only if the owner of particular surveys permits it.

The lead form is served (110) to the next potential client. As mentioned before, the set of questions presented to the next lead are informed by the machine learning model that determines the correlation of questions with the end-state of the lead.

Further to the descriptions above, a user, e.g., customer, may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user’s social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user’s identity may be treated so that no personally identifiable information can be determined for the user, or a user’s geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure provides machine-learning techniques that discover the predictive ability of questions on a questionnaire. The techniques enable a business to select or design questions or experiments that correlate with the closure of a sale. The business benefits by improving lead quality and optimizing sales and advertising resources.