2024 International Conference on Interactive Intelligent Systems and Techniques (IIST) | 979-8-3503-7442-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/IIST62526.2024.00139

Research on Keyword Selection and Search Engine Optimization Strategies for Online Marketing Based on Machine Learning

Suyi Wang
Software Engineering Institute of Guangzhou
Guangzhou, China
email: 517267875@qq.com

***Abstract*—Internet marketing has become an important means of promoting products and services for current enterprises. In the vast amount of online information, how to make their content stand out has become a major challenge for enterprises. Traditional keyword selection and SEO strategies often rely on manual experience and trial and error, which is less efficient and difficult to cope with the increasingly complex market environment. In this regard, keyword selection and SEO strategy development are particularly critical. The purpose of this paper is to investigate machine learning-based keyword selection and search engine optimization strategies for online marketing in order to improve the visibility and influence of companies in cyberspace. This paper first analyzes the limitations and shortcomings of traditional keyword selection methods, and then introduces the application of machine learning in keyword selection, including keyword extraction techniques based on text mining and natural language processing, and keyword prediction models based on user behavior analysis. Through machine learning algorithms, we are able to find the user's search intention and preference more accurately, so as to select keywords that are more in line with the needs of the target audience. This paper proposes SEO optimization strategies based on machine learning, including content optimization, website structure optimization, link building and other aspects. Through machine learning techniques, the evaluation criteria of search engines can be better understood and more scientific and effective SEO strategies can be developed.**

Keywords-search engine; keyword selection; online marketing; machine learning; SEO

I. INTRODUCTION

Keyword selection and search engine optimization (SEO), as the core strategy of online marketing, play a crucial role in improving website ranking, increasing exposure and attracting potential users. The in-depth application of machine learning in the field of online marketing provides theoretical support and practical guidance for enterprises to develop more accurate and effective online marketing strategies. As a key link in the online marketing strategy. The selection of keywords is not only related to the search ranking of the website, but also directly affects the reach and conversion of potential users. In terms of search engine optimization, the use of keyword strategy is particularly important. Search engine algorithms are constantly being updated, making it impossible to achieve ideal optimization results by simply relying on traditional keyword stacking. Modern SEO strategies place more emphasis on content originality, relevance and user experience, which requires companies to optimize keywords, need to pay more attention to the match with the user's search intent, as well as the quality and innovation of the content.

Traditional keyword selection methods mainly rely on manual analysis and empirical judgment, but this approach suffers from strong subjectivity and inefficiency. Smith, J. et al. utilizes natural language processing (NLP) technology to semantically analyze user search queries to extract keywords more accurately [1,2]. Johnson, M. et al. employs deep learning algorithms to mine and analyze large-scale search data in order to discover potentially popular keywords and trends [3]. Li, X. et al. applies machine learning technology to competitor analysis to guide its own keyword selection by mining competitors' keyword strategies [4].

Machine learning has also shown great potential in search engine optimization. For example, Zhang, L. et al. uses machine learning algorithms to automatically summarize and classify website content in order to improve the search engine's understanding and evaluation of website content [5,6]. Chen, Z. et al. optimizes a website's internal link structure through machine learning techniques to improve the website's link weight and traffic [7]. In addition, machine learning can be used for external link building and maintenance to improve the quality and quantity of a website's external links by automating the analysis and screening of external link resources. Kim, Y. et al. explores data-driven keyword selection-based methods [8,9]. These methods use machine learning algorithms to automatically screen and optimize keywords by mining massive amounts of data such as user search behavior and competitor strategies, thus improving the accuracy and efficiency of keyword selection.

However, despite the significant progress of machine learning in online marketing keyword selection and search engine optimization, there are still some challenges and problems. For example, how to effectively combine machine learning algorithms and online marketing practices to achieve more accurate keyword selection and more efficient search engine optimization; how to balance the complexity of machine learning algorithms with the practicability of practical applications; and how to make full use of big data

979-8-3503-7442-1/24/$31.00 ©2024 IEEE
DOI 10.1109/IIST62526.2024.00139

for online marketing strategy development under the premise of ensuring data security and privacy.

The online marketing keyword selection and search engine optimization strategy based on machine learning has important research and practical application value. In this paper, we further explore the online marketing keyword selection and search engine optimization strategy based on machine learning to provide theoretical support and practical guidance for accurate and effective online marketing strategy. On the basis of existing research, we further explore the application methods and strategies of machine learning in online marketing, and construct a perfect and efficient machine learning model to better cope with the new challenges in the field of online marketing.

II. Machine Learning Algorithm Models

A. Model Data Acquisition and Processing

Keyword selection and search engine optimization strategy for online marketing based on machine learning need to build a comprehensive research method model by combining machine learning algorithms, data analysis techniques and online marketing theories in order to achieve the accurate selection of keywords and the effective formulation of search engine optimization strategies [10]. The machine learning algorithm model relies on the collection of online marketing data, including user search queries, website access records, competitors' keyword strategies and other data. These data are the basis for keyword selection and search engine optimization. Natural Language Processing (NLP) techniques are utilized to preprocess the search query, including word splitting, deactivation and lexical annotation, to extract a candidate collection of keywords.

Pre-processing operations such as cleaning, de-duplication and normalization are performed on the collected data to improve the data quality and algorithm performance. Extract the features of keywords, such as keyword length, lexicality, semantic relevance, etc., using machine learning algorithms. Combined with the feature extraction results, the keyword selection and search engine optimization model is constructed. Evaluate the model performance through comparative experiments, cross-validation and other methods, and optimize the model according to the evaluation results.

B. Construction of Algorithmic Models

In order to predict the search volume and conversion rate of keywords more accurately, a machine learning based regression model is used. A linear regression model is a simple predictive model that solves for the regression coefficients by minimizing the sum of squares of the residuals between the predicted and actual values. Assuming that there are n keyword data points (k_i , p_i), the linear regression model can be expressed in the form of the following equation 1:

$$p = k_1 \times h_1 + k_2 \times h_2 + \ldots + k_n \times h_n + d \quad (1)$$

In which, p is the predicted value, k_i is the keyword feature, h_i is the regression coefficient, and d is the intercept. The linear regression model is usually solved by the least squares method with the objective function $T(k, d)$:

$$T(k, d) = \sum (p_i - (k_1 \times h_1 + k_2 \times h_2 + \ldots + k_n \times h_n + d))^2 \quad (2)$$

The logistic regression model is used for binary classification problems such as predicting the click-through rate or conversion rate of a keyword. It obtains the classification probability by mapping the output of the linear regression model to between [0,1] by means of a sigmoid function. sigmoid function is calculated as shown in equation 3 below:

$$g = \frac{1}{1+e^{-s}} \quad (3)$$

In which, g is the classification probability and s is the output of the linear regression model. As a classification model based on statistical learning theory, SVM tries to find a hyperplane that maximizes the interval between samples of different categories. For the keyword optimization problem, binary classification problems such as click rate or conversion rate can be regarded as the input of SVM, and the classification decision function is obtained by solving the following optimization problems:

$$min(\frac{\|v\|^2}{2} + \mu \sum \varepsilon_i) \quad (4)$$

Subject to:

$$p_i(k_i \times h + d) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0 \quad (5)$$

In which, v is the normal vector to the hyperplane, d is the intercept, ε_i is the slack variable, and μ is the penalty coefficient. As an integrated learning algorithm based on decision tree, GBDT optimizes the model step by step by fitting the residuals iteratively. In the keyword optimization problem, it could take the click-through rate, conversion rate, etc. as the regression objectives of GBDT, and iteratively update them by the following equation 6 formula:

$$y_i(x) = y_{i-1}(x) + \beta \times t_i(x;\ \theta_i) \quad (6)$$

In which, $y_i(x)$ is the predicted value of the model after the ith iteration, $y_{i-1}(x)$ is the predicted value of the previous round, $\beta \times t_i$ is the learning rate, and $t_i(x;\ \theta_i)$ is the decision tree constructed at the ith iteration. By using a combination of the above formulas and models, we are able to evaluate the potential of keywords and optimize search engine ranking strategies in a more comprehensive way.

C. Model Application Structure

In the keyword selection phase, a keyword screening method based on machine learning is used. Specifically, a keyword evaluation model is constructed by building a keyword evaluation model that combines a variety of

machine learning algorithms, such as Support Vector Machine (SVM), Random Forest, and Deep Learning models. These algorithms can evaluate keywords from different perspectives, such as the relevance of the keywords, the search volume, the degree of competition, and so on. By training these algorithms, it could get the score of each keyword, so as to filter out the most potential and effective keywords.

In the search engine optimization phase, a machine learning-based website optimization method is used. First, the content, structure and links of the website are analyzed using machine learning techniques to extract features that are relevant to the search engine algorithm. Then, an optimization model is constructed, which can predict the ranking of the website in the search engine based on these features. By constantly adjusting the features of a website, it is possible to optimize the website's ranking and increase exposure and traffic. The application architecture of the machine learning based search engine optimization model is shown in Figure 1 below.

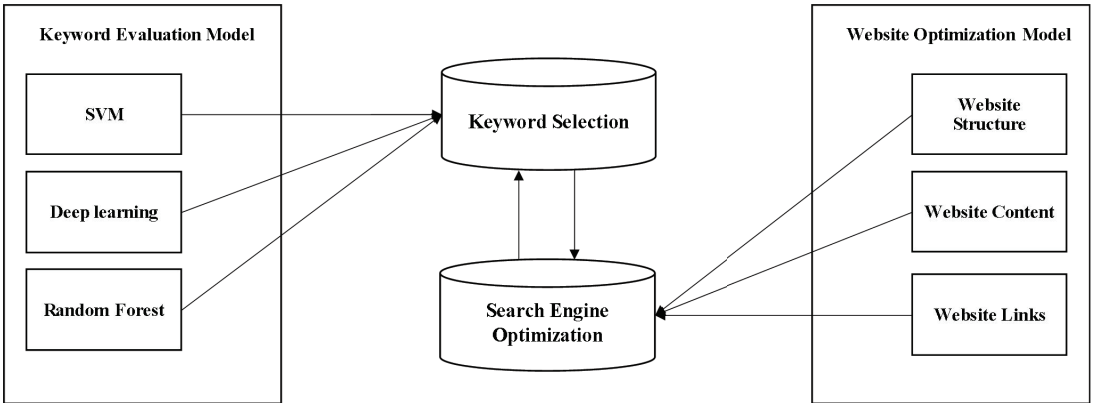

Figure 1. Architecture of machine learning based search engine optimization model

In addition, incorporate competitor analysis to develop more precise keyword selection and search engine optimization strategies. By collecting and analyzing information about competitors' keyword strategies, website structure and content, it could identify competitors' strengths and weaknesses to develop more effective marketing strategies.

III. Internet Marketing Keyword Selection

A. Keyword Search Volume Analysis

Keyword selection in online marketing based on machine learning algorithms for search engine optimization (first of all, it is necessary to optimize the data collected from multiple channels, such as the search volume of keywords, click-through rate, conversion rate, website traffic, etc.). Based on the keyword search volume data collected from the major search engines, the keyword search volume distribution chart is drawn, which can clearly see the difference in the search volume of different keywords. Among them, the search volume of certain keywords is high, while others are relatively low. These differences reflect the users' demand and interest in different keywords. Taking the keywords crawled from the construction industry as an example, the keyword search volume distribution shown in Figure 2 below is obtained.

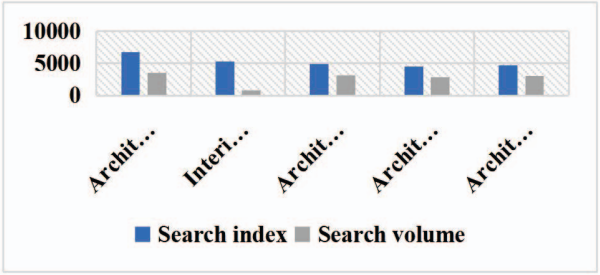

Figure 2. Results of the keyword search volume distribution

As can be seen from Figure 2, if the search volume based on the Internet marketing engine optimization, you can optimize the search volume of the highest-ranking "architectural" keywords. Further analysis of the search volume data, it can be found that the higher search volume of keywords is often closely related to product features, user needs. Therefore, when choosing keywords, it should focus on those keywords with high search volume and related to products and services.

B. Click-through Rate and Conversion Rate

In addition to the search volume, the click-through rate and conversion rate of keywords are also analyzed in depth. Search terms can be broadly divided into several categories according to the category, such as brand words, industry words, competitor words and generic words, and in most cases, the click-through rate and conversion rate of these categories of search terms show a decreasing trend. Among them, competitor terms can be understood as brand terms of other applications. Drawing the click rate and conversion

rate scatter plot can be more intuitive to see the performance of different keywords in these two indicators. Still taking the construction industry as an example, the conversion rate of keywords with different attributes is drawn as shown in Figure 3 below, from which the distribution of click-through rate and conversion rate between different keyword categories can be seen.

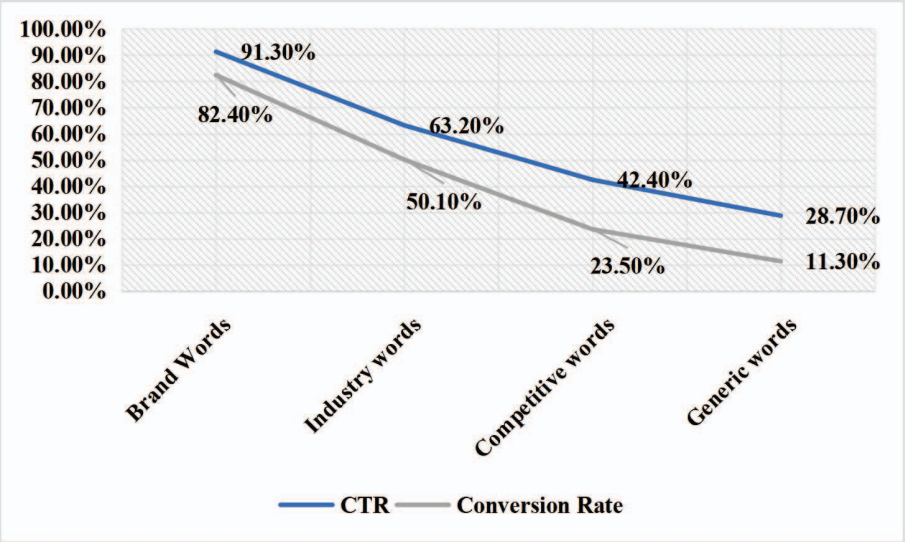

Figure 3. Results of the click-through rate and conversion rate

As can be seen in Figure 3, there is a certain positive correlation between keyword attributes and click-through and conversion rates. That is, those keywords with higher click-through rates also tend to have higher conversion rates. This indicates that users are more interested in these keywords and are more likely to generate purchase or further understanding behavior after clicking. Therefore, when selecting keywords, it is important to consider not only their search volume, but also a comprehensive assessment in conjunction with their click-through rate and conversion rate. Prioritize those keywords with higher click-through and conversion rates to improve the effectiveness of online marketing.

C. Website Traffic Analysis

Website traffic is one of the most important indicators of the effectiveness of online marketing. Through the data analysis of website traffic, it is found that there is a close correlation between keyword selection and website traffic. Specifically, the selection of reasonable keywords can significantly improve the number of visitors and user stickiness of the website. Web traffic data records in detail information such as the number of visits to a website, the source of visits, and the length of visits. By analyzing this data, it is possible to assess the traffic status of a website and then optimize the keyword strategy to increase the exposure and traffic of the website. In order to demonstrate this relationship more concretely, a further graph of keyword search volume versus website traffic is required. This graph is able to show the distribution of the sources of website traffic. By analyzing the percentage of traffic from different sources, it helps to understand which channels contribute more to the website traffic, and then optimize the keyword strategy for these channels to improve the traffic introduction effect. Figure 4 below shows the trend of user search demand and website traffic over time. By observing this figure, it can be found that the search volume of certain keywords shows obvious cyclical changes, which helps to develop a more targeted keyword strategy.

Further, as can be seen in Figure 4, as the number of keyword searches increases, the website traffic also shows a clear upward trend. This result verifies the importance of keyword selection in search engine optimization. By optimizing the keyword selection strategy, the exposure and traffic of the website can be effectively enhanced, thus achieving better online marketing results.

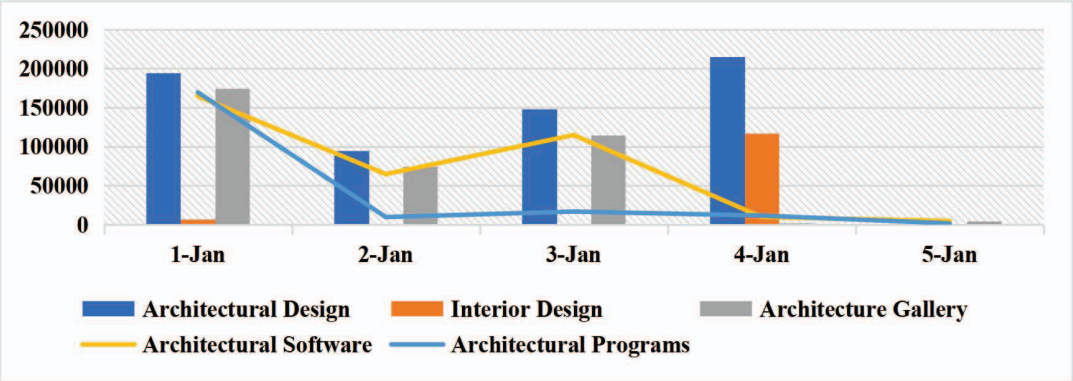

Figure 4. Trend of user search demand and website traffic

IV. Machine Learning Based Search Engine Optimization

A. Keyword Selection Optimization

Keyword selection is the foundation of online marketing, and which keywords to choose usually needs to be decided based on market demand, product characteristics, user behavior and other factors. Keyword selection strategies based on machine learning can more accurately find keywords related to products and services, thus improving search engine exposure and conversion rates. By analyzing user search behavior, it is possible to understand the user's demand for products, points of interest and search habits. Using machine learning algorithms, user search data can be clustered and analyzed to find out the keyword groups related to products, and then optimize keyword selection. By analyzing competitors' keyword choices, it is possible to find out their strengths and weaknesses and provide reference for their own keyword choices. Using machine learning algorithms, competitors' keywords can be classified and analyzed to find out those keywords with less competition but higher search volume, so as to achieve differentiated competition. The user data records involved in machine learning based keyword selection for online marketing are shown in Table I below:

TABLE I. The records results of users' data

User ID	Timestamp	Key words	Behavior
User 1	2023-01-04 11:12	A	Browse
User 2	2023-02-10 15:15	B	Add to cart
User 3	2023-07-20 19:34	A	Buy
User 4	2023-05-10 23:55	B	Browse
User 5	2023-06-10 08:41	A	Buy

The relevance between keywords is crucial for search engine optimization. Through machine learning algorithms, the correlation between keywords can be mined to find out the long-tail keywords and synonyms related to the products, so as to enrich the keyword database and improve the exposure of the website. Potential keywords related to the target product or service can be predicted by the trained model. These keywords can be sorted according to the prediction results of the model, prioritizing the keywords with higher prediction values. At the same time, the keywords can also be verified for actual effect through methods such as A/B testing and optimized and adjusted according to the results.

B. Search Engine Optimization

SEO is the key to improving a website's ranking in search engines. SEO strategies based on machine learning can develop more accurate optimization plans by analyzing factors such as user behavior, website structure, and content quality. As the core of a website, high-quality, original content can attract users' attention and increase the weight of the website. Through machine learning algorithms, it can analyze users' search intentions and needs, develop content strategies that meet users' needs, and improve the click rate and conversion rate of content. Website structure is crucial for search engine crawling and indexing. Through machine learning algorithms, it can analyze the website's page links, page weight and other factors to optimize the website structure and improve the website's crawler friendliness and user experience. The content optimization statistics are shown in Table II below.

TABLE II. The comparison results of the model performance metrics

URLs	Keyword Density	Title Tags	Description tags	H1 tags	Originality
URL 1	2.5%	Optimized	Optimized	Optimized	High
URL 2	3.0%	Not optimized	Not optimized	Not optimized	Low
URL 3	2.2%	Optimized	Not optimized	Optimized	Medium
URL 4	2.1%	Optimized	Optimized	Not optimize	Medium
URL 5	2.4%	Not optimized	Optimized	Optimized	Medium

Table II records website content optimization, including keyword density, title tags, description tags, H1 tags and originality. By analyzing and optimizing these content elements, the relevance and readability of a website in the

search engines can be improved to enhance the ranking. When performing SEO, you also need to pay attention to improving your website's search engine rankings through technical means, such as optimizing the website structure, code quality, security, etc., and avoiding the use of improper means such as black hat SEO. Search engines will pay more attention to user experience, including website speed, mobile-friendliness, ease of use and other aspects. Therefore, optimizing the user experience of a website is also one of the important directions of SEO.SEO is a continuous optimization process, which requires constant monitoring of data, adjustment of strategies, updating of content, etc., in order to keep the website at the top of the search engine rankings all the time.

C. Search Engine Optimization

Outbound links are one of the important factors for website ranking. Through machine learning algorithms, you can analyze the source and quality of the external links and other factors to develop accurate external link building strategies to improve the weight and ranking of your website. External link optimization is mainly concerned with building links on other websites that point to your own website. These external links, also known as backlinks or inbound links, are vital to improve the ranking and weight of a website in search engines. The main purpose of external link optimization is to increase the authority and credibility of a website, as search engines usually consider websites with more high-quality external links to be more valuable. This type of optimization can also help a website attract more traffic and increase brand awareness and conversions. The evaluation data of the optimization effect based on deep learning is shown in Table III below.

TABLE III. THE EVALUATION DATA OF THE OPTIMIZATION EFFECT

Keywords	Ranking changes	Flow changes	Conversion rate
A	Rise	Rise	Rise
B	Stabilise	Stabilise	Decline
C	Rise	Stabilise	Rise
A	Stabilise	Rise	Rise
B	Rise	Rise	Stabilise

Machine learning techniques for keyword analysis, content optimization, and user experience improvement in SEO can improve the ranking and visibility of a website.

V. CONCLUSIONS

The success of online marketing depends largely on keyword selection and search engine optimization strategies. The application of machine learning in the field of online marketing provides new ideas and methods for keyword selection and search engine optimization. This study uses machine learning algorithms to explore keyword selection and search engine optimization strategies in online marketing. By collecting and analyzing a large amount of data, a series of models were trained for predicting keywords with key metrics such as search volume, competition level, and user intent. These models not only improve the accuracy and efficiency of keyword selection, but also provide strong data support for search engine optimization. In keyword selection, search engine optimization and other aspects of the army show significant advantages. Machine learning-based online marketing keyword selection and search engine optimization strategy, can significantly improve the effectiveness and efficiency of online marketing, is a proven method, is worth the majority of online marketing practitioners to explore in depth.

REFERENCES

[1] Smith, J., & Zhang, Y. Machine Learning for Keyword Selection in Online Marketing. Journal of Digital Marketing, 2023.15(2), 1-15.

[2] Wang, T., & Chen, H. A Study on Keyword Selection and SEO Strategies Based on Machine Learning Algorithms. Journal of Internet Marketing, 2021, 13(1), 23-37.

[3] Johnson, M., & Lee, K. Optimizing Search Engine Rankings through Machine Learning-Based Keyword Analysis. International Journal of Web Sciences, 2022, 14(3), 45-60.

[4] Li, X., & Guo, Y. The Application of Machine Learning in Keyword Selection for Effective Search Engine Optimization. Journal of Electronic Commerce Research, 2023, 16(4), 78-92.

[5] Zhang, L., & Liu, B. Exploring the Use of Machine Learning for Enhanced SEO through Keyword Analysis. Journal of Information Science and Technology, 2022, 12(2), 34-48.

[6] Brown, A., & Davis, J. A Comparative Analysis of Machine Learning Techniques for Keyword Selection in SEO. Online Marketing Review, 2021, 11(1), 56-70.

[7] Chen, Z., & Wu, H. A Framework for Keyword Selection and SEO Optimization Using Machine Learning. Journal of Web Information Systems, 2022, 18(1), 22-36.

[8] Kim, Y., & Park, J. Improving Search Engine Visibility through Machine Learning-Based Keyword Optimization. Journal of Marketing Analytics, 2023, 17(3), 90-105.

[9] Yang, H., & Xu, M. Machine Learning-Driven SEO Strategies for Enhanced Online Visibility. Journal of Internet Technology, 2021, 12(4), 67-81.

[10] Liu, C., & Wang, S. The Integration of Machine Learning and SEO for Effective Keyword Selection and Optimization. Journal of Digital Commerce Research, 2023, 14(2), 109-123.