International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 8259-8269

https://doi.org/10.70517/ijhsa463706

Research on the Application of Data Mining-based Methods in the Construction of Brand Design Requirement Models for Financial Central Enterprises

Liangyun Zhu[1,*], Gaofeng Mi[1] and Dan Chen[1]
[1] School of Design and Art, Shaanxi University of Science and Technology, Xi'an, Shaanxi, 710119, China
Corresponding authors: (e-mail: lyzhu93@163.com).

Abstract Traditional brand design methods are difficult to accurately grasp audience needs, and data mining technology provides a new way to solve this problem. This paper constructs a brand design demand model for financial central enterprises based on data mining methods. The study adopts web crawler to obtain brand review text, applies word2vec to realize word vectorization; uses TF-IDF algorithm to extract brand overall imagery; analyzes brand local imagery based on syntactic relationship; realizes perceptual imagery parameterization through word vector technology; establishes brand demand element model, analyzes the correlation between brand features, emotional features, contextual features and behavioral features. The study found that "national credit" has the highest word frequency (14520) and TF-IDF value of 0.02687 among the theme words processed by data mining; the social factor has the highest weight (0.2278) in the brand design evaluation system constructed; and the W brand of a financial central enterprise designed by applying the model obtained a score of 4.39, with the highest score (4.44) for the spiritual factor. The study establishes a brand design model driven by "visual" + "semantic" dual modes, which improves the relevance and accuracy of the brand design of financial centralized enterprises, and provides a scientific method and practical path for the brand design of financial centralized enterprises.

Index Terms financial centralized enterprises, brand design, web crawler, data mining, word vector representation, imagery mining

I. Introduction

As an important material and political foundation of socialism with Chinese characteristics, the publicity, ideological and cultural work of central enterprises is of great significance [1]. The publicity, ideological and cultural work of state-owned central enterprises is an important part of the Party's publicity, ideological and cultural work, and bears the important task of gathering strength for stronger, better and larger state-owned capital and state-owned enterprises. In recent years, with the goal of shaping a first-class brand that matches the scale and volume of the enterprise, its status and role, and its development vision, it has closely followed the theme and main line of the national work, focused on the group's main responsibilities and main business, actively innovated brand publicity means, and continuously improved the effectiveness of brand publicity, so as to gather strong momentum for the construction of world-class enterprises [2], [3].

The arrival of the big data era is an inevitable trend in the future development of the world, but also the inevitable choice of the information age, enterprises want to stand in the tide of the big data era must be for enterprise change and innovation, keep pace with the development of the times, combined with the characteristics of the development of the times [4]-[7]. Under the era of big data, the most important thing for the development of an enterprise is the design and dissemination of corporate brand image, only a good corporate brand image design can be in the wave of many brands to immediately seize the consumer's psychology, and then continue to attract new consumers [8]. So that enterprises can be in a sustained and healthy development process to create more wealth for the development of society and the people's standard of living to provide a constant source of power [9]-[11]. Today's society has entered the era of informationization with rapid economic development, and data mining is an indispensable technology in the era of big data [12]-[14]. Through the mining of a large amount of information in the network, many of the data developed can be used to analyze and deal with the information that may affect the corporate brand image, helping the corporate brand image to better sustainable development.

As the backbone of China's financial system, financial central enterprises bear the important mission of promoting national economic development, serving the real economy and maintaining financial security. Under the background of economic globalization and digital transformation, the brand building of financial central enterprises faces new challenges and opportunities. As an important part of the core competitiveness of an enterprise, brand is not only related to the market recognition and reputation of the enterprise, but also directly affects the market value and long-term development of the enterprise. However, the traditional brand design method often relies on the designer's subjective experience and intuition, and lacks scientific analysis and accurate grasp of user needs, resulting in a significant gap between the design results and the actual needs of users. In the Internet era of information explosion, users express their evaluations and expectations of brands through various channels, and how to effectively collect and transform this scattered but rich information into a powerful support for brand design has become an important issue in the brand building of financial central enterprises. Data mining technology, as an effective means to discover knowledge from massive data, provides new ideas and methods to solve this problem. Through the collection and analysis of text data such as user comments, it can dig deeper into the user's perception of the brand and demand, providing a scientific basis for brand design. Especially in the field of financial central enterprises, which is highly specialized and requires high credibility, it is of great significance to enhance the relevance and effectiveness of brand design by how to construct a brand design demand model that meets its characteristics and realizes the effective transformation from data to design. However, most of the existing research focuses on the commercial brand marketing level, and there is a relative lack of research on the brand design requirements of such special subjects as financial central enterprises, especially the lack of systematic research and practical application based on data mining methods.

This study constructs a brand design demand model for financial central enterprises based on data mining methods. Firstly, obtain the text data of financial central enterprises' brand comments through web crawler; secondly, use word2vec tool to realize the word vectorization representation, and use TF-IDF algorithm to extract the overall brand imagery; then, carry out the brand local imagery mining based on the syntactic relationship, and realize the parameterization of perceptual imagery through the word vectors; then, construct the brand demand element model, and analyze the brand features, emotional features, Then, the brand demand element model is constructed to analyze the relationship between brand features, emotional features, contextual features and behavioral features; finally, the visual + semantic dual-mode driven brand design model is established, and the validity of the model is verified by taking the example of W, a financial centralized enterprise. Through this series of research steps, we aim to provide a set of data-driven scientific methods and practical paths for brand design of financial central enterprises.

II. Multi-level brand design demand big data mining methods

II. A. Text Big Data Acquisition

Financial centralized enterprise brand comment text big data acquisition channels mainly include public datasets, web crawlers, and data trading platforms. Among them, web crawlers are an effective way to obtain information in the Internet. Web crawlers are a kind of program or script that can automatically crawl information in the Internet according to set rules. In order to obtain the brand's review text data, the brand review data in the brand evaluation website is crawled by web crawler and structured as the brand review text big data to be analyzed.

II. B. Word vectorization based on word2vec

word2vec is a tool for representing words as Distributed Representation word vectors [15], by using text to train a shallow neural network to obtain a real vector representation of words. The word vectors generated by word2vec can be used for text mining tasks such as lexical clustering and synonym analysis. In this paper, we investigate the use of word2vec tool to realize word vector representation, where a large number of product review texts are de-emphasized and segmented, and subsequently input into word2vec for training, so as to obtain word vector representation of vocabulary words containing the semantics of the vocabulary words.

II. C. Overall brand imagery extraction based on keyword extraction

Keyword extraction algorithm is a method of analyzing the overall keywords of the text mainly by calculating and extracting the statistical features of the vocabulary in the document. Therefore, the key adjectives extracted by the keyword extraction algorithm can be used as the overall perceptual imagery vocabulary of the brand to participate in the analysis of the overall perceptual imagery of the brand.

TF-IDF algorithm is a method to assess the importance of a word in a document through the statistical features of the word [16]. The formula of TF-IDF algorithm is as follows:

$$tf-idf(w_{i,d})=\frac{n_{i,d}}{\sum_{d}n_{i,d}}\log\frac{M}{m_i+1} \tag{1}$$

where $w_{i,d}$ is the vocabulary w_i in document d. $tf-idf(w_{i,d})$ is the importance of vocabulary w_i in document d. $n_{i,d}$ is the word frequency of vocabulary w_i in document d. m_i is the number of documents containing vocabulary w_i. M is the total number of documents.

The TF-IDF algorithm reduces the weight value of high-frequency words appearing in multiple documents as keywords to prevent the extraction of common words of lower importance, thus ensuring the algorithm's extraction effect.

II. D.Extraction of Brand Local Imagery Based on Syntactic Relationships

II. D. 1) Noun Grouping of Brand Local Characteristics

In order to discover the brand local feature words in the text, the semantic relationship between the nouns is analyzed by using word vectors and syntactic relations, and the correlation is approximated by using the maximum spanning tree algorithm, and the nouns existed in the same tree are considered as the candidate words for the same feature, and finally, the irrelevant words are manually deleted from them, so as to obtain different expressions of the same brand feature.

When analyzing the semantic relationship, the semantic relationship between word nouns is divided into two aspects: similarity and dissimilarity for analysis. Since word vectors can reflect the degree of similarity in the semantic relationship of words, the similarity between nouns is determined by calculating the cosine similarity between word vectors. The word similarity is calculated by the cosine similarity of word vectors with the following formula:

$$s(w_a,w_b)=\frac{\sum_{d=1}^{D}(w_{ad}\times w_{bd})}{\sqrt{\sum_{d=1}^{D}(w_{ad})^2}\times\sqrt{\sum_{d=1}^{D}(w_{bd})^2}} \tag{2}$$

where $s(w_a,w_b)$ is the word similarity between noun w_a and noun w_b, which is used to evaluate the similarity relationship between the nouns. w_{ad},w_{bd} is the value of the word vector parameter in dimension d for noun w_a and noun w_b, respectively.

Subsequently, the dissimilarity relationship between nouns is analyzed. Since the nouns in the same sentence generally describe different things and are mutually exclusive, it is considered that there is a dissimilarity relationship between the nouns in the same sentence, which can be expressed as equation (3):

$$c(w_a,w_b)=\begin{cases}1, & if\ w_a\ and\ w_b\ in\ same\ sentence\\ 0, & else\end{cases} \tag{3}$$

Subsequently, the similarity and dissimilarity relationships obtained were combined to determine the correlation parameter between the words through the following equation:

$$r(w_a,w_b)=\begin{cases}0, & if\ c(w_a,w_b)=1\\ 0, & if\ s(w_a,w_b)<0\\ s(w_a,w_b), & if\ c(w_a,w_b)=0\end{cases} \tag{4}$$

where $r(w_a,w_b)$ is the semantic relationship parameter between noun w_a and noun w_b, with higher values indicating higher correlation between w_a and w_b.

II. D. 2) Brand Local Imagery Extraction

The brand feature imagery vocabulary is obtained through syntactic relations. Combining the obtained noun groups of the local features of the brand to be analyzed in section 2.4.1, the perceptual imagery of the brand features to be analyzed is obtained by analyzing the syntactic relations between the nouns in the groups and the adjectives in the text. For the same brand feature, the importance parameter of the collected perceptual imagery vocabulary is calculated through word frequency. Therefore, for brand feature i of a sample, N adjectives with higher frequency of occurrence are selected as perceptual imagery words, and the corresponding perceptual imagery can be expressed as:

$$K_i = T_i^T \cdot W_i = \sum_{n=1}^{N} (t_{in} w_{in}) \tag{5}$$

where K_i denotes the perceptual imagery of the sample brand feature j. T_i denotes the vector consisting of imagery importance parameters. W_i is the vector of imagery words. w_{in} and t_{in} are the n th perceptual imagery vocabulary and the corresponding importance parameter of sample brand feature j, respectively. T_i is calculated using word frequency, then:

$$T_i = \left[\frac{tf_{W_{i1}}}{\sum_{n=1}^{N} tf_{W_{in}}}, \frac{tf_{W_{i2}}}{\sum_{n=1}^{N} tf_{W_{in}}}, \ldots, \frac{tf_{W_{iN}}}{\sum_{n=1}^{N} tf_{W_{in}}} \right]^T \tag{6}$$

where $tf_{w_{in}}$ denotes the frequency of occurrence of word w_{in}.

II. E. Word vector-based parameterization of perceptual imagery

After analyzing the text, the overall brand imagery and the local brand imagery can be obtained, and it can be seen from Eq. (5) that both of them have a similar structure, which is the product of the keywords and the weights, i.e:

$$K_i = T_i^T \cdot W_i = \sum_{n=1}^{N} (t_{in} w_{in}) \tag{7}$$

where w_{in} is the word, which is converted to a real vector to participate in the computation through the word vector model obtained in Section 2.2. The perceptual imagery vocabulary is downgraded through factor analysis to reduce the dimensionality of the word vectors involved in the calculation of the perceptual imagery parameters while retaining the information about the semantic differences between the words. Assuming that the dimensionality of the word vectors after dimensionality reduction is m, equation (7) can be expressed as:

$$K_i = \sum_{n=1}^{N} (t_{in} S(w_{in})) = \sum_{n=1}^{N} \begin{bmatrix} t_{in} S_1(w_{in}) \\ t_{in} S_2(w_{in}) \\ \ldots \\ t_{in} S_m(w_{in}) \end{bmatrix} \tag{8}$$

where $S(w_{in})$ denotes the word vector of word w_{in} after dimensionality reduction, and $S_m(w_{in})$ denotes the value of the word in dimension m after dimensionality reduction. Since multiple imagery can be superimposed on each other, the summation of imagery adjectives is the summation of the corresponding dimensions of their word vectors to obtain the values of the sample perceptual imagery in each dimension, i.e.:

$$K_i = \begin{bmatrix} \sum_{n=1}^{N} t_{in} S_1(w_{in}) \\ \sum_{n=1}^{N} t_{in} S_2(w_{in}) \\ \ldots \\ \sum_{n=1}^{N} t_{in} S_m(w_{in}) \end{bmatrix} \tag{9}$$

where the perceptual imagery K_i of brand i is represented by a vector of real numbers. Existing studies often take all possible values of the perceptual imagery parameter as the perceptual imagery space, and the tendency of the brand on each word as the perceptual imagery parameter. In this study, the parameterized description of perceptual imagery is realized by the above equation, and the word vector space S is the perceptual imagery space.

III. Brand demand characterization

III. A. Demand element modeling

In order to extract specific and comprehensible brand demand information from the structured brand semantic strings obtained by segmentation processing, and to clarify the brand demand elements and their characteristic relationships contained in brand online reviews, it is necessary to construct a brand demand element model to extract and analyze brand demand characteristics. Demand refers to the audience's expectations or requirements for the brand or service, which is the intersection of the audience's usage scenario and brand characteristics. Brand is the specific expression and satisfaction of demand, and brand features, emotional features, contextual features, and behavioral features are important components of the demand factor model. Brand characteristics mainly contain information related to function, modeling, and structure. Emotional polarity can be divided into three categories: positive, neutral and negative. Behavioral characteristics mainly consider the emitter of the behavior. Situational characteristics are mainly the events and locations where the events occur.

In the brand demand element model, associating brand features with emotional features can reflect the audience's satisfaction with a certain product feature, and combining the behavioral and contextual features under this brand feature can reveal the brand demand under a specific usage scenario, which can make the abstract demand features more concrete. Therefore, analyzing the brand behaviors and usage contexts corresponding to brand features under different emotional polarities can help further clarify brand demands and guide subsequent design and research.

III. B. Demand feature word extraction and screening

In order to analyze the brand behaviors and usage contexts corresponding to the brand feature words under different emotional polarities, it is necessary to extract the four types of demand elements from the brand semantic string data obtained after the lexical segmentation process according to the brand demand element model, and screen out the brand demand features with higher effectiveness to further clarify the brand demand. With the help of Language Technology Platform (LTP) to delineate the lexical nature of the participle, carry out lexical annotation, and categorize the semantics with different lexical natures into the four types of features of the brand demand element model. Partial annotation results of the feature words of the four categories of brand demand elements, brand, behavior, emotion and context, in the semantic data. There are differences in the meaning and usage of the feature words in different situations. Therefore, experts are invited to conduct manual proofreading review during categorization, and reference to expert opinions for further identification and labeling to reduce the bias of feature word labeling and categorization, in order to improve the rationality and accuracy of categorization.

In order to rank and filter the four types of word-splitting results separately by word frequency, feature weights are assigned using the word frequency-inverse document frequency (F_T-F_{ID}) method. Document frequency (F_D) is based on the statistical method for feature selection, and feature screening is performed by setting the upper and lower thresholds.F_D refers to the ratio of the number of documents t containing the particular feature, to the number of documents n used, which is defined as:

$$F_D = \frac{t}{n} \tag{10}$$

F_{ID} is used to measure the importance of a feature word in the whole text collection. Its calculation formula is:

$$F_{ID}(t,n) = \log\frac{N}{Nt + 0.01} + 0.01 \tag{11}$$

where N denotes the total number of documents used and N_t denotes the number of documents containing the feature t. The smaller the F_{ID}, the higher the importance of the vocabulary in the text collection.

F_T-F_{ID} is suitable for large-scale text processing and is used to weight important words with low word frequency in a document to improve the accuracy of text processing. Definition:

$$F_T - F_{ID} = F_T(t,d)F_{ID}(t,n) \tag{12}$$

III. C. Brand Feature Clustering

In order to sort the filtered brand feature words according to word frequency, in which there are words with similar meanings that still lack accuracy and formality in semantic expression, it is necessary to cluster the brand feature words. After clustering, the center word of each type of semantics, i.e. the core semantics, is the most written and concise description of brand features, which can accurately reflect the needs of the brand. Explanation semantics plays a semantic complementary role.

After word clustering of brand features, words describing the same brand features are combined into a feature group consisting of a core feature and an explanatory feature. Then similar brand feature words are combined and decomposed step by step to get the brand feature structure of financial central enterprises.

III. C. 1) Affective feature polarity clustering

Sentiment feature words reflect the brand's affirmative, negative or neutral attitude towards its own brand features, expressing the brand's emotional inclination towards the brand, which is the key object for extracting and analyzing the relationship between brand demand features. Affective feature polarity clustering can be more efficient and accurate to understand the outstanding problems reflected by the brand, which is of great significance for clarifying the brand demand subsequently.

III. C. 2) Brand Demand Characteristics Relationship Extraction and Construction

In order to obtain specific brand requirements, the word co-occurrence model was used to analyze the relationship between the four types of feature words. The higher the frequency of two feature words appearing in the same text, the closer the relationship between them. By calculating the co-occurrence degree between feature words, the co-occurrence frequency between feature words can be obtained, and then the co-occurrence relationship matrix is constructed and normalized. Afterwards, the word contribution mapping is drawn based on the co-occurrence degree, so as to visualize the degree of closeness between each feature word of the financial central enterprise brand.

The node with more connected edges in each functional family is the feature word that has a greater degree of co-occurrence for that functional family and represents the theme of that functional family. The formula (13) for the co-occurrence of feature word W_i is:

$$G(W_i) = \sum_{(W_i,W_i)\in E(G)} S(W_i,W_j) \tag{13}$$

where $E(G)$ denotes the set of all edges in Figure G. $S(W_i,W_j)$ denotes the average word co-occurrence of word W_i and word W_j.

The visualization of feature word associations can be used to map brand features to emotional polarity, usage scenarios, and usage behaviors one by one, in order to learn about brand needs.

IV. Brand design driven by “visual” + “semantic” models

IV. A. Construction of the semantic model

IV. A. 1) Feature extraction

After the text has been segmented, it needs to be converted into a feature vector that can be processed by machine learning algorithms. Feature extraction can be performed using the TF-IDF algorithm, a commonly used text feature extraction algorithm that calculates the weight of each word based on its frequency of occurrence in the document and in the entire anthology, thus converting the text data into vector form.

IV. A. 2) Text classification

Text classification can be performed using Support Vector Machine (SVM) algorithm [17]. SVM is a classical supervised learning algorithm whose core idea is to find an optimal hyperplane to separate different classes of data. In text classification, we represent each text sample as a feature vector and then use SVM algorithm to categorize different categories of text.

IV. A. 3) Thematic models

Topic modeling is a method of transforming textual data into a probabilistic model that identifies potential topics from textual data and assigns textual data to those topics. In topic modeling, each topic is represented by a word distribution vector, and each document is represented by a distribution vector of topics. The core idea of topic modeling is to transform text data into a probabilistic model, and discover potential topics in the text through model learning, and then perform classification, clustering and other operations on the text.

IV. A. 4) Sentiment analysis

Sentiment analysis is a method of determining and analyzing the sentiment of a text, which can automatically identify the emotional tendencies expressed in the text. Sentiment analysis can be divided into two methods based on sentiment dictionary and machine learning. Among them, the sentiment dictionary-based method classifies each word as positive, negative or neutral by constructing a sentiment dictionary, and then calculates the sentiment score of the text based on the sentiment polarity of the words and the contextual information. Machine learning-based methods, on the other hand, classify the text as positive, negative or neutral by training a classification model.

The advantage of text semantic mining is that it can dig deeper into the meaning and potential relationships behind the text to better understand user needs and feedback. By analyzing a large number of user comments and feedback, brand design and experience can be better optimized.

IV. B. Construction of the "vision" model

As a theory of visual grammar, Michelian Imagery not only has a wide range of applications in the field of image recognition and classification, but also plays an important role in the field of data mining. Data mining refers to the process of analyzing and mining a large amount of data in order to find the hidden relationships, patterns and laws in it. In data mining, image data, as a rich form of information, has a rich amount of information and expressive power, so the application of Michelle's theory of imaging in image data mining can not only improve the accuracy and efficiency of data mining, but also mine the laws and associations in the data from a deeper level.

First, we need to perform data preprocessing and feature extraction in order to convert the image data into a vector form that can be processed by machine learning algorithms. Data preprocessing includes operations such as image normalization, grayscaling, denoising, etc. to ensure the reliability and consistency of the data. Feature extraction refers to the extraction of meaningful features from an image for subsequent model training and classification. We used a pre-trained model on the LaMem dataset, where images are fed into the model and high-dimensional feature vector representations are extracted through the computation of multiple convolution and pooling layers.The LaMem dataset assigns an image memorability score to each image, which indicates how much of this image is retained in human memory.

After feature extraction is complete, we select high quality images with high image memorability scores for subsequent data mining and classification. By organizing and then segmenting the image with MemNet, its memorability score is calculated for each sub-region to find out the sub-region with the highest score.

Finally, we input the feature vectors into the classifier for classification. The subregions with the highest memorability scores are identified and the designer extracts the parts that can be used as design elements. The graph is then connected to the "semantic" model. The conceptual framework of the "visual" + "semantic" dual-model model is shown in Figure 1.

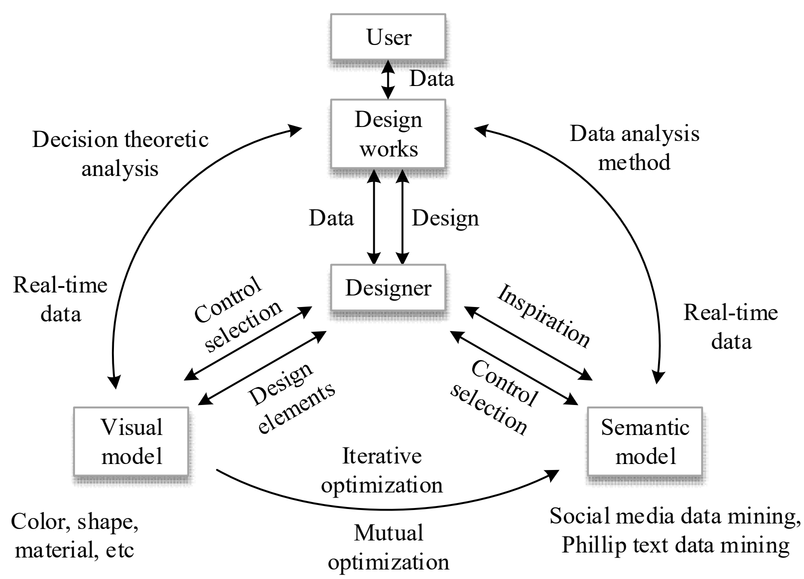

Figure 1: "Semantic"+"visual" dual mode design model framework

V. Brand Design Analysis

V. A. Brand Demand Analysis

This paper takes W, a financial central enterprise, as a research object and analyzes its brand design in depth. The brand image of W enterprise is mined after data mining technology. Data mining of the results after word division and de-duplication can get the lexical nature (e.g. adjective, adverb, noun, etc.), number of articles and frequency of occurrence of each word after word division. After processing, the word frequency statistics are obtained, including: word, lexical nature (noun, adverb, adjective, verb, etc.), number of times, number of articles, word frequency, etc. The results (partial) of the word frequency statistics are shown in Table 1.

Table 1: Word frequency statistics (part)

Word	Frequency	TF-IDF	Word	Frequency	TF-IDF	Word	Frequency	TF-IDF
National credit	10449	0.01629	Public interest	2628	0.00963	TOP 500	1668	0.00684
Financial security	5008	0.01224	Low-carbon operation	2460	0.00937	Credit AAA	1637	0.00677
Prudent operation	4802	0.01211	Anti-fraud	2350	0.00918	Industry TOP	1633	0.00642
Professional	4385	0.01198	Knowledge popularization	2308	0.00893	Customer satisfaction	1620	0.00633
Authoritative	4379	0.01136	Red gene	2179	0.00892	Value	1614	0.00539
Heritance	4158	0.01114	Identification	2168	0.00889	Cross-border	1586	0.00535
Strategy	4088	0.01109	Main visual system	2034	0.00823	Digital	1574	0.00532
National trust	3533	0.01076	IP image	1926	0.00819	Crisis dealing	1535	0.00521
Global layout	3136	0.01063	History	1872	0.00799	Public opinion	1471	0.00504
Technology	3047	0.01054	Brand	1869	0.00792	Reputation	1433	0.00503
Service	2920	0.01026	Slogan	1845	0.00775	Fintech	1413	0.00499
ESG	2803	0.01008	Brand image	1744	0.00735	Digital RMB	1409	0.00497
Revitalization	2753	0.00997	Media tone	1726	0.00726	Intelligent	1377	0.00463
Inclusive	2718	0.00981	Visual specification	1719	0.00707	Ecosystem	1270	0.00445
Green finance	2650	0.00975	Brand ceremony	1704	0.00687	Open	1173	0.00423

Table 2: Processed theme word statistics (part)

Word	Frequency	TF-IDF	Word	Frequency	TF-IDF	Word	Frequency	TF-IDF
National credit	14520	0.02687	Slogan	1819	0.00852	Open	637	0.00522
Fintech	8100	0.02178	Intelligent	1676	0.00827	Brand	601	0.00465
Financial security	7145	0.01862	Digital	1665	0.00827	IP image	590	0.00445
National trust	5307	0.01224	Service	1517	0.00765	Ecosystem	518	0.00424
Inclusive	4129	0.01153	Low-carbon operation	1475	0.00756	Revitalization	507	0.00381
Prudent operation	4050	0.01142	Media tone	1425	0.00756	Knowledge popularization	488	0.00359
Green finance	3591	0.01125	Heritance	1414	0.00742	Governance	463	0.00354
Professional	3091	0.01069	TOP 500	921	0.00716	ESG	455	0.00323
Public interest	2900	0.01038	Credit AAA	833	0.00697	Digital RMB	445	0.00299
Identification	2850	0.01035	Reputation	774	0.00687	Meta	377	0.00258
Authoritative	2739	0.00969	Cross-border	755	0.00655	Technology	363	0.00253
History	2715	0.00958	Public opinion	731	0.00647	Data	345	0.00245
Strategy	2572	0.00944	Global layout	729	0.00614	Anti-fraud	324	0.00207
Brand image	2151	0.00901	Customer satisfaction	685	0.00611	Visual specification	318	0.00159
Value	1859	0.00878	Red gene	670	0.00593	Future	261	0.00134

It should be noted that although the Micro-Word Cloud tool can basically complete the text word segmentation and word frequency statistics of data, as well as count the recurring nouns in users' online comments, it is insufficient in handling synonymous words and does not have an automatic filtering function, which may result in the appearance of some synonymous words. Moreover, due to the different language habits of users who leave comments, the same functional feature of smart speakers may be described in different words. Therefore, it is possible that in the comment data, different nouns or noun phrases describe the same brand feature. For example, the "appearance" and "shape" of Enterprise W refer to the same attribute. The "price" and "price range" of a speaker both refer to the feature of "price", etc. Therefore, the obtained high-frequency word list needs to be manually corrected based on the product features and the contextual semantic analysis of user comments. The high-frequency words such as "sound quality" and "sound effect", "national credit" and "national trust", "professionalism" and "authority" are combined into a group of high-frequency words, and the TF-IDF values of the combined high-frequency words are calculated. And manually delete the words in the key words that are useless for guiding brand innovation. The statistical table of the frequency of the processed subject words and the TF-IDF value is shown in Table 2.

As shown in Table 2, the processed list of theme words is the effective information extracted from users' online comments after various processes of data mining, and these theme words contain users' needs and expectations for various characteristics of W Enterprise, a financial centralized enterprise, which can be used to guide the direction of brand design according to the theme words obtained from these mining. However, due to the different categories and directions of expression of these lemmas, it is necessary to map and categorize the lemmas according to the brand characteristics before in-depth design guidance can be provided.

V. B. Brand Design Effectiveness

Using this paper's brand design method based on data mining to design the brand of financial central enterprise W, design evaluation of the design program is conducted to verify the effectiveness of this paper's method. The indicators of design evaluation should reflect the degree of satisfaction of the audience, and here we draw on the brand design evaluation indicators in several academic literatures and the brand demand information in Chapter 3 of this paper to comprehensively determine the evaluation indicators of the brand. The brand design evaluation index system of financial central enterprises is shown in Table 3.

Table 3: Evaluation index system for financial SOE brand designing

Target layer	Primary index	Secondary index
Financial SOE brand designing	IP factor	Brand image attractiveness
		Brand image recognition
		Brand style characteristics
	Aesthetic factor	Color aesthetics
		Brand fineness
		Brand cuteness
		Detail fineness
	Mental factor	Story association
		Emotional resonance
		Interactivity
	Functional factor	Functional implementation
		Usage value
		Brand symbolism
	Social factor	Brand social image representation
		Brand social responsibility embodiment
		Social acceptance

Table 4: Evaluation index weight for financial SOE brand designing

Target layer	Primary index	Weight	Secondary index	Weight
Financial SOE brand designing	IP factor	0.1842	Brand image attractiveness	0.3756
			Brand image recognition	0.3615
			Brand style characteristics	0.2629
	Aesthetic factor	0.1811	Color aesthetics	0.2258
			Brand fineness	0.2846
			Brand cuteness	0.2134
			Detail fineness	0.2762
	Mental factor	0.1883	Story association	0.3915
			Emotional resonance	0.3752
			Interactivity	0.2333
	Functional factor	0.2186	Functional implementation	0.3141
			Usage value	0.3542
			Brand symbolism	0.3317
	Social factor	0.2278	Brand social image representation	0.3185
			Brand social responsibility embodiment	0.3168
			Social acceptance	0.3647

The entropy value method is utilized to calculate the weights of the indicators at all levels in the brand design evaluation index system of financial central enterprises constructed above, and the results are shown in Table 4.

Publicize the designed brand image of W enterprise to the audience and collect the audience's opinions through online questionnaires. A total of 500 questionnaires were distributed and 487 were retrieved. Among them, 472 were valid, with an effective rate of 94.4%. The Likert five-point scale was used as the basis for data calculation, where 1 to 5 represent "very dissatisfied", "satisfied", "average", "satisfied" and "very satisfied" respectively. After calculation, the evaluation results of the W brand design of financial central enterprises are shown in Table 5.

From Table 5, it can be found that the overall rating of the brand design of the financial centralized enterprise W based on the data mining method is 4.39, which has gained a more satisfactory feedback among the audience. Among the first-level evaluation indicators, the scoring results from high to low are: spiritual factors (4.44), IP factors (4.43), aesthetic factors (4.41), social factors (4.37), and functional factors (4.32), and the scores of all the first-level indicators are above 4.30. Among the secondary indicators, the scoring results range from 4.17 to 4.60, with all secondary indicators scoring more than 4.15. The lowest evaluated score is brand loveliness, and the highest scoring secondary indicator is brand sophistication.

Table 5: Evaluation results of W financial SOE brand designing

Target layer	Primary index	Score	Secondary index	Weight
Financial SOE brand designing (4.39)	IP factor	4.43	Brand image attractiveness	4.40
			Brand image recognition	4.59
			Brand style characteristics	4.24
	Aesthetic factor	4.41	Color aesthetics	4.25
			Brand fineness	4.60
			Brand cuteness	4.17
			Detail fineness	4.53
	Mental factor	4.44	Story association	4.58
			Emotional resonance	4.28
			Interactivity	4.48
	Functional factor	4.32	Functional implementation	4.54
			Usage value	4.26
			Brand symbolism	4.19
	Social factor	4.37	Brand social image representation	4.58
			Brand social responsibility embodiment	4.24
			Social acceptance	4.31

VI. Conclusion

Big data mining methods have significant advantages in the construction of brand design demand model for financial central enterprises. By applying the technical means of text big data acquisition, word vectorization, keyword extraction and syntactic relationship analysis, the brand demand characteristics implied in the user comments are effectively parsed. The experimental results show that in the brand design evaluation of financial central enterprises W, the social factor has the highest weight of 0.2278, followed by the functional factor of 0.2186, reflecting the importance of the social responsibility attribute and functional utility of financial central enterprises' brands in the audience's mind. The brand design effect evaluation shows that the design solution based on the data mining method obtains an overall satisfaction score of 4.39, with the spiritual factor scoring the highest at 4.44, followed by the IP factor at 4.43, indicating that the design has successfully shaped a brand image with distinctive spiritual connotations and identifying features. In the secondary index evaluation, the brand image recognition and brand social image performance both received a high score of 4.58, highlighting the success of the design solution in brand recognition and social image building. The “visual” + “semantic” dual-mode driven brand design model innovatively combines textual semantic mining with visual image analysis, realizes the effective transformation from data to design, provides scientific methods and practical paths for brand design of financial central enterprises, and has important reference value for improving the relevance and effectiveness of brand design of financial central enterprises.

References

[1] Naughton, B. (2017). The current wave of state enterprise reform in China: A preliminary appraisal. Asian Economic Policy Review, 12(2), 282-298.

[2] Bai, C. E., Lu, J., & Tao, Z. (2006). The multitask theory of state enterprise reform: Empirical evidence from China. American Economic Review, 96(2), 353-357.
[3] Wang, L., & Yung, K. (2011). Do state enterprises manage earnings more than privately owned firms? The case of China. Journal of Business Finance & Accounting, 38(7-8), 794-812.
[4] Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. Information systems and e-business management, 16(3), 479-491.
[5] Wang, W. Y. C., & Wang, Y. (2020). Analytics in the era of big data: The digital transformations and value creation in industrial marketing. Industrial Marketing Management, 86, 12-15.
[6] Xu, M., Zhang, Y., Sun, H., Tang, Y., & Li, J. (2024). How digital transformation enhances corporate innovation performance: The mediating roles of big data capabilities and organizational agility. Heliyon, 10(14).
[7] Melewar, T. C., Dennis, C., & Foroudi, P. (Eds.). (2021). Building corporate identity, image and reputation in the digital era. Routl
[8] Rindell, A., & Strandvik, T. (2010). Corporate brand evolution: corporate brand images evolving in consumers' everyday life. European Business Review, 22(3), 276-286.
[9] Jo Hatch, M., & Mirvis, P. H. (2010). Designing a positive image: Corporate branding and social responsibility. In Positive design and appreciative construction: From sustainable development to sustainable value (pp. 35-55). Emerald Group Publishing Limited.
[10] Foroudi, M. M., Mahdavi, F., & Foroudi, P. (2021). Corporate Brand Design Management from Different Perspectives. Corporate Brand Design, 15-40.
[11] Raiko, D., & Cherepanova, V. (2019). Creation of the image and brand of the enterprise on the basis of marketing as a philosophy of management. Baltic Journal of Economic Studies, 5(2), 191-205.
[12] Fahmi, D. Y., & Zulbainarni, N. (2021, March). Mining social media (twitter) data for corporate image analysis: a case study in the indonesian mining industry. In Journal of Physics: Conference Series (Vol. 1811, No. 1, p. 012107). IOP Publishing.
[13] Malyshenko, K., Malyshenko, V., Anashkina, M., & Anashkin, D. (2020). Informal evaluation of corporate image based on text mining. In SHS Web of Conferences (Vol. 89, p. 03009). EDP Sciences.
[14] Okoro, H. (2025). Corporate Data Mining and Manipulation in Marketing. In Impacts of Leakage, Whistleblowing, and the Rise of Propaganda (pp. 231-260). IGI Global Scientific Publishing.
[15] Zhan Zerui. (2025). Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis: A Case of Food Reviews. ITM Web of Conferences,70.
[16] Lu Jiaxin. (2025). Text vectorization in sentiment analysis: A comparative study of TF-IDF and Word2Vec from Amazon Fine Food Reviews. ITM Web of Conferences,70.
[17] Huiyan Li. (2020). Text recognition and classification of english teaching content based on SVM. Journal of Intelligent & Fuzzy Systems,39(2),1-11.