Research on Image Emotional Semantic Retrieval Mechanism Based on Cognitive Quantification Model

Tian Liang, Liqun Zhang(✉), and Min Xie

Institute of Design Management, Shanghai Jiao Tong University,
Shanghai, China
zhanglliqun@gmail.com

Abstract. In the wake of the development of first-person engagement and crowdsourcing content creation, images are given abundant subjective dimensions of information, especially emotional ones. This research tried to purpose an approach for the image emotional semantic retrieval based on cognitive quantification model by using tags. In this research "Daqi", a typical Chinese emotional experience, is taken as an example to construct an emotional quantification model of it through semantic association analysis and statistical data analysis. The results of verification experiments indicated that it is practical and effective to rank images and recommend tags in image emotional retrieval system based on cognitive model. It is foreseeable that the theory of this research can be applied to other social digital resources, like music or video.

Keywords: Image emotional semantic retrieval · Cognitive quantification · Image annotation

1 Introduction

With the rapid development of Web2.0, first-person engagement and crowdsourcing content creation have boomed as new paradigms of interactions. Digital media resources, especially images, are given abundant subjective and expressive dimensions of cognitive contents, which advocates researches for retrieval of the emotional information. Human perception and understanding of image emotional information are operations mainly on the semantic level. However, "Semantic gap" between low-level image features and high-level emotional semantic can be hardly to bridge completely.

This research aims to purpose an approach for the image retrieval of emotional content based on tags. Tags are image description added by users directly, so image emotional semantic retrieval can be implemented based on text retrieval technology without the extraction of image information. To increase the number of tags, the channels of tag generation are expanded by including the relevant user interactive behaviors.

In this research, a cognitive quantification model of their emotional qualities or of their reception by users is constructed to organize and manipulate social image resources. Meanwhile, the model is applied to emotional semantic tag recommendation,

A. Marcus and W. Wang (Eds.): DUXU 2017, Part III, LNCS 10290, pp. 115–128, 2017.
DOI: 10.1007/978-3-319-58640-3_10

which is beneficial to improve the efficiency of image annotation and the validity of image recommendation.

This research creatively proposes a mechanism for the emotional semantic retrieval of images. The mechanism has the following advantages,

- Retaining the user's subjective view of images maximally by using user-generated tags, which ensures the credibility of retrieval and is more lightweight.
- Binding the behaviors and views of users in image retrieval based on behavior psychology, which expands tag sources and provides more data for the modeling.
- Mining the potential emotional semantic of pictures based on the cognitive quantification model, which improve the effectiveness of image emotional semantic retrieval. Meanwhile diverse associated tags can be recommended according to the relevant weight in the model, which further improves the emotional image annotation.

The rest of the paper is organized as follows. An overview of the tag-based image emotional semantic image retrieval mechanism is presented in Sect. 2. Section 3 discussed the details of methods used in the mechanism. To verify the rationality of the proposed mechanism, experiments with small sample size have been done. The process and analyses are presented in Sect. 4. Section 5 is the summary and prospect.

2 Concepts and Methods

2.1 Concepts

Image Tag. Tags are the keywords added by users to describe the image contents. In particular, tags are not only the labels, but also can be the keywords in the titles, comments and so on. As tags are image descriptor, it is easy to recommend images directly based on text retrieval technology and there is no need to extract and analyze information of images.

An image can be added multiple tags, and a tag is also used to describe multiple images. The user is the creator of annotation behaviors, creating an association between images and tags.

At present, annotation behaviors are as follows,

- Adding the title or labels when uploading images;
- Adding the labels or grouping when collecting images;
- Making comments on images.

The tags generated by the above behaviors are "explicit tags". In fact, the above annotation behaviors are non-essential and costly behaviors for users, who is lacking of motivation. A large portion of the users only view images without leaving a tag.

However, studies have shown that the behaviors of users to retrieve images can reflect how much users agree with the retrieval results, revealing the relevance between the retrieval keywords and images. Using the data of users' browsing behaviors, "implicit tags" can be made. The details are discussed in Sect. 2.2.

Emotional Quantification Model. Image semantics has several levels. Emotional semantics lies on the highest level of abstract semantics, which can be defined as the semantics described intensity and type of feelings, moods, affections or sensibility evoked in humans by viewing images. It is usually represented in adjective form, romantic, brilliant etc.

Constructing an image emotional computational model usually involves three parts,

- extracting image perceptual features that can stimulate users' emotions;
- establishing the emotional recognition mechanism to bridge the semantic gap between low-level visual features and high-level emotional semantics;
- constructing the model to represent image emotional semantics that meet the needs of users' query.

Visual identity and machine learning are main methods in the first and the second parts. They are aimed to build an association between images and their emotional semantics that can be retrieved easily, which can be implemented just by "tags".

Based on tags, this paper is focused on the construction of the model to quantify image emotions which users search for.

In general emotional semantic models, the specific emotion is split and associated with the six basic dimensions of the emotion, anger, disgust, fear, joy, sadness, and surprise. The models relying on these six basic dimensions are not enough in emotional fine grain to represent complex emotional semantics or to distinguish between the various emotions clearly.

Learning from this idea, in this paper the emotional semantics are represented by more flexible and more targeted "emotional dimensions" which contain a variety of "emotional elements" associated with a certain relationship. The emotional semantic quantification model is expressed by "emotional dimensions", and the emotional dimension is extracted from "emotional elements".

2.2 Methods

Users' Retrieval Behavior. As mentioned in the previous section, it is non-essential and costly for users to add image tags and most users only browse images without leaving a tag. Studies have shown that users' behaviors to retrieve images can reflect the degree of users' recognition on the search results, in the image retrieval system, that is, the relevance of the search keywords and images.

Using the data of users' retrieval behaviors, it can be predicted whether the images are associated with the search keyword, and if so, the "search keyword" can be added to the images as an "implicit tag".

In users' retrieving images, the operations of generating "implicit tags" are as follows.

- click to view the image after retrieving;
- download/save the image after retrieving;
- snapshot the image after retrieving.

Among them, we remain neutral on the operation of clicking to view, because it is impossible to exclude behaviors that users click images to view due to curiosity and so on rather than recognition.

Combining the behaviors generating "explicit tags" mentioned in the previous section, the relationship between tags generated by users' behaviors ("explicit tags" and "implicit tags") and images is divided into three level, related_1, neutral_0.5 non-relevant_0, as follows (Fig. 1).

BEHAVIOR	TAG	RELEVANCE
add titles when uploading images	keywords in titles	1
add labels when uploading images	label	1
add labels when collecting images	label	1
add to groups when collecting images	group name	1
make comments on images	keywords in comments	1
download/save images after retrieval	search keywords	1
snapshot images after retrieval	search keywords	1
click images to view after retrieval	search keywords	0.5
other non-relevance behaviors	none	0

Fig. 1. Relevance of behavior and tag

A user may have more than one annotation for an image, but the above behaviors are not cumulative in relevance degree up to 1, that is, as long as there is a strong annotation behavior (relevance_1), the tag is added to the image by the user.

Tag Clustering Analysis. Clustering is a common data analysis tool and a basic algorithm for data mining. The essence of clustering analysis is to divide data into several clusters according to the relevance. Therefore, it has high similarity within clusters and big difference between clusters.

Tag clustering can be used to find semantic-related labels in social annotation systems, Begdman et al. The principle tag can be mostly represented by identifying the subject of the cluster. If the clusters constitute the special emotion, the principle tags of them is "emotional elements".

The semantic relevance of two tags can be obtained by relying on the semantic knowledge databases, such as Wordnet (for English) and CSC (for Chinese), to build a semantic correlation matrix (Fig. 2).

Tag \ Tag	t_1	t_2	t_3	…	t_n
t_1	1	0.2	0.4	…	0.3
t_2	0.2	1	0.5	…	0.8
t_3	0.4	0.5	1	…	0.7
…	…	…	…	1	…
t_n	0.3	0.8	0.7	…	1

Fig. 2. Semantic correlation matrix of tags

Based on the semantic relevance coefficient in the matrix to build N-dimensional space, the Euclidean distance formula (1) can be used to calculate the spatial distance of two tags. The closer, the more similar tags can be considered.

$$Euclid(1,2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{1}$$

The shortest two clusters are merged into a large cluster until all small clusters are merged into a large cluster. The whole process can be shown in a form of a tree structure. Any number of semantic groups can be got through hierarchical clustering analysis (Fig. 3).

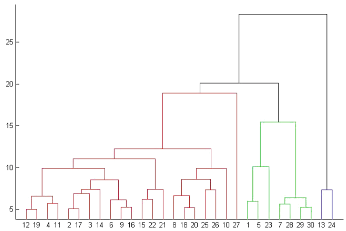

Fig. 3. Hierarchical clustering

Factor Analysis of Emotional Cognition. The main purpose of factor analysis is to reduce dimension by transferring lots of indicators into several comprehensive indicators under little information lost.

As tags increases, redundancy and uniqueness should be considered when performing image matching in the emotional space. Using factor analysis, an orthogonal emotional space can be constructed not only to retain the majority of original indicators meaning, but also to ensure the simplification of the model.

At the same time, the weight of each emotion dimension is allocated according to the contribution rate of each factor, rather than the artificial judgment, which makes the model more objective and reasonable.

Creating a tag-image matrix S = $\{s_{im}\}$ based on website image-tag database() (Fig. 4).

Tag / Image	T_1	T_2	T_3	…	T_m
I_1	s_{11}	s_{12}	s_{13}	…	s_{1m}
I_2	s_{21}	s_{22}	s_{23}	…	s_{2m}
I_3	s_{31}	s_{32}	s_{33}	…	s_{3m}
…	…	…	…	…	…
I_i	s_{i1}	s_{i2}	s_{i3}	…	s_{im}

Fig. 4. A tag-image matrix

s_{im} is the score of image I_i on tag T_m, determined by the number of T_m on I_i. Since the number of tags on different images are in a different order of magnitude, it needs to be standardized. For an image I_a and the number of tags N = $\{n_1, n_2, \ldots, n_m\}$, its score s_{ai}

$$s_{ai} = n_i / \max(n_i), \quad i = 1, 2, \ldots, m$$

After factor analysis, factors F = {F1, F2, …, Fn}, that is the "emotional dimensions", and their variance contribution rate A = {a1, a2, …, an} can be got. The emotion Y can be represented by emotional cognitive factors F, as (2).

$$Y = a_1F_1 + a_2F_2 + \ldots + a_nF_n \quad (2)$$

In addition, we obtain a factor load coefficients matrix of tags T and factors F. The rotation factor load coefficient matrix $B = \{b_{mn}\}$ can be obtained by using Varimax to rotate the initial factor load matrix. The rotation method can keep the factors orthogonal to each other, but the variance difference of each factor is maximized, so it is convenient to explain the factor.

Quantization models of each emotional dimension can be obtained.

$$F_i = b_{i1}T_{p1} + b_{i2}T_{p2} + \ldots + b_{im}T_{pm} \quad (i = 1, 2, \ldots, n) \quad (3)$$

(3) into (2), we can get an emotional cognitive quantification model of Y

$$Y = c_1T_{p1} + c_2T_{p2} + \ldots + c_mT_{pm} \quad (4)$$

$$C = AB \quad (5)$$

3 Image Emotional Semantic Retrieval System

In the image emotional semantic retrieval systems, it is an important part to find most appropriate images for given tags and find the most appropriate tags for given images. That is, need to find the most appropriate match with each other tag-image pairs.

The system has three main functional modules,

First, recommend the relevant images according to users' search terms.

Second, recommend the relevant tags for the images that users agree with.

The third, based on users' feedback on the recommended results, expand the tag-image database.

Given an input tag T_a, the recommended image set $I_R = \{I_{r1}, I_{r2}, \ldots, I_{ri}\}$, and the recommended tag set for image $T_r = \{T_{r1}, T_{r2}, \ldots, T_{ri}\}$, the flow of the system is divided into the following steps (Fig. 5).

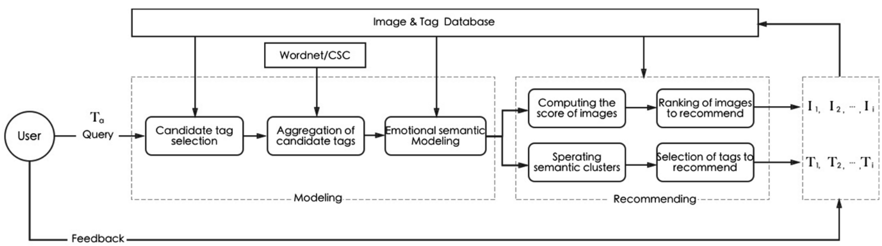

Fig. 5. Image emotional semantic retrieval system

3.1 Emotional Semantic Modeling

Candidate Tag Selection. Given an initial tag T_a, the tag set $T = \{T_1, T_2, \ldots, T_i\}$ where all tags associated with it are included is collected based on the coherence principle. The more images any two tags are annotated on by users at the same time, the more there are cognitive links with them. In order to avoid the tag noise, it is needed to set a threshold, usually an empirical value. There are experimental evidences found that the value set 10 can make the best performance.

Through synonyms to merger these coherence tags, get the candidate tag set $T_c = \{T_{c1}, T_{c2}, \ldots, T_{ci}\}$

Aggregation of Candidate Tags. Using the external semantic knowledge database Wordnet (for English) or CSC (for Chinese), establish a semantic association matrix according to the semantics relevance of T_c, shown as below. The correlation value is 0–1, the higher the value is, the higher the relevance is.

Then, cluster the candidate tags by the clustering algorithm to generate several original semantic clusters. A representative label is selected as a representative label to represent each cluster of the class, which forms emotional elements set $T_p = \{T_{p1}, T_{p2}, \ldots, T_{pm}\}$.

Emotional Semantic Modeling. Based on the website image-tag database, making factor analysis on T_p, to get the emotional dimension set $F = \{F_1, F_2, \ldots, F_n\}$ and its variance contribution rate $A = \{a_1, a_2, \ldots, a_n\}$, and then T_a emotional cognitive model can be expressed as follow.

$$Y = a_1F_1 + a_2F_2 + \ldots + a_nF_n \tag{6}$$

According to the rotation factor load factor matrix $B = \{b_{mn}\}$, each factor F quantization model can be expressed as follow.

$$F_i = b_{i1}T_{p1} + b_{i2}T_{p2} + \ldots + b_{im}T_{pm} \quad (i = 1, 2, \ldots, n) \tag{7}$$

Combining the above two expressions to obtain the emotional semantic model of T_a with tag elements T_p,

$$Y = c_1T_{p1} + c_2T_{p2} + \ldots + c_mT_{pm} \tag{8}$$

3.2 Images Ranking and Tags Selection

Ranking of Images to Recommend. The value of each image on T_p is s_{mi}. The emotional value of each image on tag T_a can be calculated by using the formula (8). Recommended image are sorted by the emotional value from high to low.

Selection of Tags to Recommend. Each image has the highest score tag T_{p1} that is included in the emotional dimension F, where there are other tags T_p $\{T_{p2}, T_{p3}, \ldots, T_{pi}\}$. Tags are recommended according to the weight in the formula (7).

Feedback Collection. According to the user's browsing behavior, collect user's various types of annotation activities, which can generate explicit tags or implicit tags to enrich and expand the image-tag library data.

4 Simulation Experiment

4.1 Experiment Setting

"Daqi" was set as the special emotional term, and 25 testers (12 males and 13 females) were invited to participate in the experiment. 150 images from appliances, furniture, transportation, construction, utensils, jewelry and other fields were selected as exciter. Testers were asked to grade the correlation between images and some terms related to "Daqi", non-relevance 0, neutral 0.5, relevance 1. Using the experimental data to get the "Daqi" emotional semantic quantification model, some validation tests were made.

4.2 Experiment Process

Construct Emotional Quantification Model. Through the literature, online comments, and website etc., more than 140 adjectives appearing at the same time as "Daqi" were collected. Then, merge these terms into 45 terms through synonyms. Researchers, according to their own professional experience and cognition. A semantic correlation matrix (Fig. 6) of these 45 emotional terms were built on researchers' professional experience.

	concise	grand	elegant	solemn	funky	large size	delicate	strong	purely	symmetry	unite	smooth	generous	full	onceptio
concise	1	0.3	0.4	0.2	0.7	0.6	0.2	0.2	0.8	0.5	0.5	0.6	0.7	0.4	0.3
grand	0.3	1	0.5	0.7	0.4	0.9	0.4	0.5	0.4	0.6	0.6	0.5	0.6	0.6	0.5
elegant	0.4	0.5	1	0.7	0.5	0.4	0.7	0.3	0.7	0.5	0.6	0.7	0.8	0.7	0.7
solemn	0.2	0.7	0.7	1	0.4	0.5	0.6	0.4	0.5	0.7	0.7	0.6	0.7	0.6	0.8
funky	0.7	0.4	0.5	0.4	1	0.5	0.3	0.4	0.8	0.5	0.6	0.6	0.6	0.5	0.4
large size	0.6	0.9	0.4	0.5	0.5	1	0.4	0.5	0.5	0.6	0.6	0.6	0.5	0.5	0.6
delicate	0.2	0.4	0.7	0.6	0.3	0.4	1	0.2	0.4	0.5	0.6	0.5	0.6	0.7	0.5
strong	0.2	0.5	0.3	0.4	0.4	0.5	0.2	1	0.4	0.4	0.5	0.5	0.5	0.6	0.4
purely	0.8	0.4	0.7	0.5	0.8	0.5	0.4	0.4	1	0.6	0.6	0.7	0.6	0.6	0.6
symmetry	0.5	0.6	0.5	0.7	0.5	0.6	0.5	0.4	0.6	1	0.8	0.7	0.5	0.6	0.5
unite	0.5	0.6	0.6	0.7	0.6	0.6	0.6	0.5	0.6	0.8	1	0.7	0.5	0.6	0.5
smooth	0.6	0.5	0.7	0.6	0.6	0.6	0.5	0.5	0.7	0.7	0.7	1	0.6	0.7	0.6
generous	0.7	0.6	0.8	0.7	0.6	0.5	0.6	0.5	0.6	0.5	0.5	0.6	1	0.6	0.6
full	0.4	0.6	0.7	0.6	0.5	0.5	0.7	0.6	0.6	0.6	0.6	0.7	0.6	1	0.6
conception	0.3	0.5	0.7	0.8	0.4	0.6	0.5	0.4	0.6	0.5	0.5	0.6	0.6	0.6	1
strong	0.2	0.7	0.3	0.6	0.2	0.6	0.3	0.7	0.4	0.4	0.5	0.4	0.4	0.3	0.4
rounded	0.6	0.5	0.7	0.6	0.5	0.5	0.6	0.4	0.6	0.6	0.6	0.7	0.5	0.8	0.5
rough	0.3	0.8	0.1	0.4	0.7	0.7	0.1	0.8	0.4	0.3	0.4	0.4	0.4	0.4	0.5
delicate	0.7	0.3	0.6	0.3	0.6	0.4	0.7	0.2	0.7	0.5	0.5	0.5	0.4	0.4	0.4
peaceful	0.8	0.4	0.6	0.6	0.8	0.5	0.5	0.3	0.8	0.6	0.7	0.6	0.6	0.5	0.7
free and easy	0.7	0.5	0.4	0.4	0.8	0.7	0.4	0.5	0.5	0.5	0.5	0.7	0.7	0.5	0.6
structured	0.7	0.5	0.5	0.5	0.5	0.5	0.6	0.4	0.5	0.7	0.7	0.6	0.5	0.6	0.6
hale	0.5	0.6	0.3	0.5	0.5	0.6	0.2	0.8	0.4	0.6	0.5	0.4	0.5	0.4	0.4
simple and elegant	0.9	0.3	0.7	0.5	0.8	0.4	0.4	0.3	0.8	0.7	0.5	0.5	0.6	0.5	0.6
since muddy	0.3	0.8	0.4	0.7	0.4	0.8	0.3	0.8	0.4	0.5	0.5	0.5	0.6	0.6	0.5
rigorous	0.4	0.5	0.6	0.6	0.4	0.5	0.7	0.6	0.5	0.7	0.8	0.6	0.6	0.6	0.5
vivid	0.4	0.3	0.6	0.5	0.5	0.5	0.7	0.5	0.6	0.5	0.5	0.6	0.6	0.7	0.6
uninhibited	0.4	0.7	0.3	0.6	0.4	0.7	0.4	0.7	0.4	0.5	0.5	0.6	0.5	0.5	0.5
tall	0.5	0.9	0.3	0.6	0.3	0.9	0.4	0.7	0.3	0.5	0.5	0.5	0.5	0.5	0.5
heavy	0.3	0.8	0.3	0.7	0.5	0.7	0.4	0.6	0.3	0.6	0.5	0.4	0.4	0.5	0.5

	strong	rounded	rough	delicate	peaceful	e and ea	structured	hale	e and ele	nce mudd	rigorous	vivid	ninhibitec	tall	heavy
concise	0.2	0.6	0.3	0.7	0.8	0.7	0.7	0.5	0.9	0.3	0.4	0.4	0.4	0.5	0.3
grand	0.7	0.5	0.8	0.3	0.4	0.5	0.5	0.6	0.3	0.8	0.5	0.3	0.7	0.9	0.8
elegant	0.3	0.7	0.1	0.6	0.6	0.4	0.5	0.3	0.7	0.4	0.6	0.6	0.3	0.3	0.3
solemn	0.6	0.6	0.4	0.3	0.6	0.4	0.5	0.5	0.5	0.7	0.6	0.5	0.6	0.6	0.7
funky	0.2	0.5	0.7	0.6	0.8	0.8	0.5	0.5	0.8	0.4	0.4	0.5	0.4	0.3	0.5
large size	0.6	0.5	0.7	0.4	0.5	0.7	0.5	0.6	0.4	0.8	0.5	0.5	0.7	0.9	0.7
delicate	0.3	0.6	0.1	0.7	0.5	0.4	0.6	0.2	0.4	0.3	0.7	0.7	0.4	0.4	0.4
strong	0.7	0.4	0.8	0.2	0.3	0.5	0.4	0.8	0.3	0.8	0.6	0.5	0.7	0.7	0.6
purely	0.4	0.6	0.4	0.7	0.8	0.5	0.5	0.4	0.8	0.4	0.5	0.6	0.4	0.3	0.3
symmetry	0.4	0.6	0.3	0.5	0.6	0.5	0.7	0.6	0.7	0.5	0.7	0.5	0.5	0.5	0.6
unite	0.5	0.6	0.4	0.5	0.7	0.5	0.7	0.5	0.5	0.5	0.8	0.5	0.5	0.5	0.5
smooth	0.4	0.7	0.4	0.5	0.6	0.7	0.6	0.4	0.5	0.5	0.6	0.6	0.6	0.5	0.4
generous	0.4	0.5	0.4	0.4	0.6	0.7	0.5	0.5	0.6	0.6	0.6	0.6	0.5	0.5	0.4
full	0.3	0.8	0.4	0.4	0.5	0.5	0.6	0.4	0.5	0.6	0.6	0.7	0.5	0.5	0.5
conception	0.4	0.5	0.5	0.4	0.7	0.6	0.6	0.4	0.6	0.5	0.5	0.6	0.5	0.5	0.5
strong	1	0.2	0.6	0.2	0.2	0.4	0.6	0.7	0.3	0.7	0.6	0.4	0.6	0.7	0.6
rounded	0.2	1	0.4	0.5	0.6	0.6	0.5	0.2	0.5	0.6	0.5	0.6	0.4	0.5	0.4
rough	0.6	0.4	1	0.1	0.2	0.6	0.5	0.7	0.4	0.7	0.4	0.4	0.8	0.6	0.6
delicate	0.2	0.5	0.1	1	0.7	0.4	0.5	0.3	0.7	0.4	0.5	0.6	0.4	0.4	0.3
peaceful	0.2	0.6	0.2	0.7	1	0.6	0.5	0.4	0.7	0.5	0.6	0.4	0.4	0.5	0.5
free and easy	0.4	0.6	0.6	0.4	0.6	1	0.5	0.6	0.4	0.6	0.3	0.6	0.8	0.5	0.4
structured	0.6	0.5	0.5	0.5	0.5	0.5	1	0.5	0.5	0.6	0.7	0.6	0.5	0.5	0.5
hale	0.7	0.2	0.7	0.3	0.4	0.6	0.5	1	0.3	0.7	0.6	0.5	0.7	0.6	0.6
simple and elegant	0.3	0.5	0.4	0.7	0.7	0.4	0.5	0.3	1	0.4	0.5	0.6	0.4	0.5	0.3
since muddy	0.7	0.6	0.7	0.4	0.5	0.6	0.6	0.7	0.4	1	0.5	0.6	0.8	0.7	0.7
rigorous	0.6	0.5	0.4	0.5	0.6	0.3	0.7	0.6	0.5	0.5	1	0.6	0.5	0.5	0.6
vivid	0.4	0.6	0.4	0.6	0.4	0.6	0.6	0.5	0.6	0.6	0.6	1	0.6	0.4	0.3
uninhibited	0.6	0.4	0.8	0.4	0.4	0.8	0.5	0.7	0.4	0.8	0.5	0.6	1	0.6	0.4
tall	0.7	0.5	0.6	0.4	0.5	0.5	0.5	0.6	0.5	0.7	0.5	0.4	0.6	1	0.5
heavy	0.6	0.4	0.6	0.3	0.5	0.4	0.5	0.6	0.3	0.7	0.6	0.3	0.4	0.5	1

Fig. 6. A semantic correlation matrix

Through the clustering analysis, ward method, 16 related emotional terms, emotional elements, on behalf of each cluster respectively were obtained. They were Quality, Generous, Uniform, Smooth, Solemnly, Full, Rounded, Elegant, Simple, Artless, Pretty, Delicate, Angular, Hard, Huge, Uninhibited.

25 testers were asked to score the degree of correlation between 150 stimuli images and the 16 emotional terms, respectively. The result of factor analysis (Fig. 7) on the experimental data is as follow. KMO is 0.789 and the data is suitable for factor analysis.

Communalities

	Initial	Extraction
Quality	1.000	.872
Generous	1.000	.834
Uniform	1.000	.767
Smooth	1.000	.771
Solemnly	1.000	.894
Fulled	1.000	.789
Rounded	1.000	.910
Elegant	1.000	.774
Simple	1.000	.813
Artless	1.000	.895
Pretty	1.000	.768
Delicate	1.000	.867
Angular	1.000	.876
Hard	1.000	.767
Huge	1.000	.822
Uninhibited	1.000	.897

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.538	34.614	34.614	5.538	34.614	34.614	4.362	27.262	27.262
2	4.131	25.816	60.429	4.131	25.816	60.429	2.694	16.840	44.102
3	1.923	12.018	72.447	1.923	12.018	72.447	2.458	15.361	59.464
4	.989	6.179	78.627	.989	6.179	78.627	2.399	14.994	74.457
5	.735	4.596	83.223	.735	4.596	83.223	1.402	8.766	83.223
6	.511	3.197	86.420						
7	.423	2.642	89.061						
8	.338	2.115	91.176						
9	.318	1.990	93.166						
10	.245	1.531	94.696						
11	.232	1.450	96.146						
12	.163	1.019	97.165						
13	.145	.909	98.074						
14	.124	.775	98.849						
15	.095	.594	99.443						
16	.089	.557	100.000						

Extraction Method: Principal Component Analysis.

Fig. 7. Communalities and total variance explained

To ensure a reasonable explanation, we choose the factor combination of which contribute is up to 83.223%. There are five factors in it. And the degree of extraction of each emotional term can reach more than 75%. Get the expression (9) of the five factors i, as follows (Fig. 8).

$$Y = 0.42F_1 + 0.31F_2 + 0.14F_3 + 0.07F_4 + 0.06F_5 \tag{9}$$

Component Matrix[a]

	Component				
	1	2	3	4	5
Quality	.907	-.041	.037	.208	-.053
Generous	.852	.048	.279	.159	-.060
Smooth	.812	.276	.029	-.160	-.088
Uniform	.780	.172	.272	-.080	.219
Fulled	.729	.057	-.309	-.323	.232
Elegant	.709	.489	.084	.117	-.112
Delicate	.654	-.261	-.198	.503	-.280
Solemnly	.632	-.305	-.267	.235	.525
Hard	.255	-.730	.387	.108	-.084
Angular	-.200	-.730	.484	.256	.055
Pretty	-.202	.728	.284	.318	.122
Huge	.485	-.686	.133	-.262	.174
Rounded	.518	.684	-.349	-.146	-.173
Simple	.179	.629	.604	-.073	-.126
Uninhibited	.380	-.595	.383	-.429	-.262
Artless	-.187	.578	.664	-.055	.288

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Fig. 8. Component matrix

According to the composition of the score coefficient matrix, "Daqi" emotional cognition model can be got, as follows.

$$
\begin{aligned}
Y = {} & 0.1T_1 + 0.12T_2 + 0.12T_3 + 0.17T_4 - 0.04T_5 + 0.11T_6 + 0.17T_7 + 0.15T_8 \\
& + 0.13T_9 + 0.03T_{10} + 0.01T_{11} + 0.03T_{12} - 0.13T_{13} - 0.04T_{14} + 0.01T_{15} \\
& + 0.08T_{16}
\end{aligned} \tag{10}
$$

Extract Emotional Dimensions. According to the rotation component matrix, five emotional dimensions can be made sure, and each emotional dimension inside the emotional composition is as follows (Fig. 9).

F_1 (Quality, Generous, Delicate, Elegant, Smooth, Uniform)
F_2 (Angular, Rounded, Hard, Full)
F_3 (Artless, Simple)
F_4 (Uninhibited, Huge, Pretty)
F_5 (Solemnly).

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Quality	.873	.068	-.101	.179	.250
Generous	.860	.011	.142	.219	.164
Delicate	.767	-.143	-.506	-.024	.039
Elegant	.755	.376	.235	-.077	.033
Smooth	.689	.445	.131	.248	.138
Uniform	.646	.176	.334	.254	.377
Angular	-.089	-.891	-.048	.268	.012
Rounded	.452	.821	.054	-.165	-.033
Hard	.256	-.639	-.160	.514	.056
Fulled	.388	.524	-.121	.307	.505
Artless	-.069	-.031	.924	-.187	-.024
Simple	.334	.167	.772	-.034	-.276
Uninhibited	.204	-.221	-.044	.894	-.080
Huge	.200	-.249	-.199	.715	.411
Pretty	.067	.071	.598	-.618	-.137
Solemnly	.422	-.069	-.304	.033	.786

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 13 iterations.

Fig. 9. Rotated component matrix

And the correlation matrix between the various emotional components is as follows (Fig. 10).

Correlation Matrix[a]

		Quality	Generous	Uniform	Smooth	Solemnly	Fulled	Rounded	Elegant	Simple	Artless	Pretty	Delicate	Angular	Hard	Huge	Uninhibited
Correlation	Quality	1.000	.815	.654	.669	.582	.554	.388	.654	.158	-.201	-.163	.656	-.094	.283	.415	.291
	Generous	.815	1.000	.652	.701	.456	.471	.384	.612	.322	.011	.005	.524	-.025	.289	.366	.343
	Uniform	.654	.652	1.000	.701	.377	.523	.367	.611	.336	.168	.030	.353	-.163	.141	.350	.231
	Smooth	.669	.701	.701	1.000	.342	.548	.641	.606	.288	-.006	.033	.377	-.353	-.013	.225	.237
	Solemnly	.582	.456	.377	.342	1.000	.549	.123	.235	-.256	-.318	-.317	.502	.032	.306	.464	.118
	Fulled	.554	.471	.523	.548	.549	1.000	.512	.439	.044	-.248	-.222	.325	-.372	.038	.314	.215
	Rounded	.388	.384	.367	.641	.123	.512	1.000	.643	.346	.043	.227	.194	-.790	-.476	-.256	-.229
	Elegant	.654	.612	.611	.606	.235	.439	.643	1.000	.471	.192	.203	.400	-.436	-.144	.033	-.016
	Simple	.158	.322	.336	.288	-.256	.044	.346	.471	1.000	.668	.465	-.166	-.228	-.125	-.323	-.046
	Artless	-.201	.011	.168	-.006	-.318	-.248	.043	.192	.668	1.000	.609	-.453	-.082	-.266	-.317	-.190
	Pretty	-.163	.005	.030	.033	-.317	-.222	.227	.203	.465	.609	1.000	-.248	-.266	-.454	-.582	-.482
	Delicate	.656	.524	.353	.377	.502	.325	.194	.400	-.166	-.453	-.248	1.000	.069	.310	.319	.220
	Angular	-.094	-.025	-.163	-.353	.032	-.372	-.790	-.436	-.228	-.082	-.266	.069	1.000	.622	.411	.400
	Hard	.283	.289	.141	-.013	.306	.038	-.476	-.144	-.125	-.266	-.454	.310	.622	1.000	.509	.630
	Huge	.415	.366	.350	.225	.464	.314	-.256	.033	-.323	-.317	-.582	.319	.411	.509	1.000	.707
	Uninhibited	.291	.343	.231	.237	.118	.215	-.229	-.016	-.046	-.190	-.482	.220	.400	.630	.707	1.000
Sig. (1-tailed)	Quality		.000	.000	.000	.000	.000	.000	.000	.048	.017	.043	.000	.164	.001	.000	.001
	Generous	.000		.000	.000	.000	.000	.000	.000	.000	.453	.481	.000	.398	.001	.000	.000
	Uniform	.000	.000		.000	.000	.000	.000	.000	.000	.039	.378	.000	.044	.069	.000	.007
	Smooth	.000	.000	.000		.000	.000	.000	.000	.001	.476	.367	.000	.000	.447	.009	.006
	Solemnly	.000	.000	.000	.000		.000	.099	.007	.003	.000	.000	.000	.371	.001	.000	.108
	Fulled	.000	.000	.000	.000	.000		.000	.000	.322	.004	.010	.000	.000	.346	.000	.012
	Rounded	.000	.000	.000	.000	.099	.000		.000	.000	.327	.008	.021	.000	.000	.003	.008
	Elegant	.000	.000	.000	.000	.007	.000	.000		.000	.022	.016	.000	.000	.066	.366	.435
	Simple	.048	.000	.000	.001	.003	.322	.000	.000		.000	.000	.041	.008	.095	.000	.315
	Artless	.017	.453	.039	.476	.000	.004	.327	.022	.000		.000	.000	.195	.002	.000	.023
	Pretty	.043	.481	.378	.367	.000	.010	.008	.016	.000	.000		.004	.002	.000	.000	.000
	Delicate	.000	.000	.000	.000	.000	.000	.021	.000	.041	.000	.004		.235	.000	.000	.010
	Angular	.164	.398	.044	.000	.371	.000	.000	.000	.008	.195	.002	.235		.000	.000	.000
	Hard	.001	.001	.069	.447	.001	.346	.000	.066	.095	.002	.000	.000	.000		.000	.000
	Huge	.000	.000	.000	.009	.000	.000	.003	.366	.000	.000	.000	.000	.000	.000		.000
	Uninhibited	.001	.000	.007	.006	.108	.012	.008	.435	.315	.023	.000	.010	.000	.000	.000	

a. Determinant = 1.053E-006

Fig. 10. Correlation matrix

4.3 Experimental Verification

Verification of Model Calculated Value. To ensure similar cognition comparisons, "appliances" are served as test. 10 appliance images (Fig. 11) were shown to be scored on the relevance with "Daqi" and we collected the scores data from 100 volunteers (48 male and 52 female). Take the average to rank the image and compare with the theoretical values of model (10) (Fig. 12).

Fig. 11. Appliance image

No.	TEST	CAMPUT
10	0.87	0.94
6	0.80	0.81
7	0.80	0.70
3	0.73	0.76
4	0.73	0.72
9	0.73	0.72
1	0.67	0.65
2	0.60	0.59
5	0.60	0.65
8	0.53	0.62

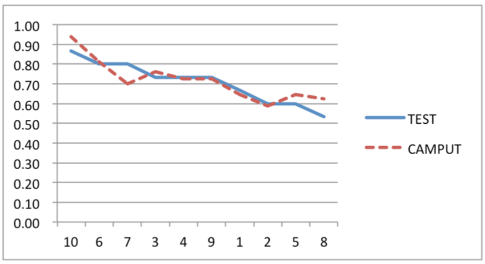

Fig. 12. Compare compute and test value

There are two ambiguities in the forecasting trend (No. 7 and No. 2), and the coincidence degree is as high as 80%. The model established can basically predict the emotional sensitivity of images.

Verification of Tag Recommended. Based on the emotional dimensions and the relevance between the terms, recommend the relevant label, some tags were selected to recommended, the maximum number 7, and asked 25 testers (12 male and 13 female) to choose the related ones from the recommended ones and calculated the use rate, the proportion of selected and provided tags.

Taking image No. 10 as an example, the highest-scored tags are Generous, Simple, and Rounded, in F1, F2 and F3 dimension, respectively. Combining the relevance matrix, as follows,

- Generous → F1 → Quality0.815, Smooth0.701, Uniform0.652, Elegant0.612, Delicate0.524
- Rounded → F2 → Full0.512
- Simple → F3 → Artless0.668

The recommended tags are as follow,

Generous, Rounded, Simple, Quality, Smooth, Artless, Uniform.

In this way, the recommended tags for the above 10 images and the average usage rate are as follows (Fig. 13).

NO	Tag max	Tag recommend	Use rate
1	Generous Simple Pretty	Generous Simple Pretty Quality Smooth Artless Uniform	0.94
2	Generous Hard	Generous Hard Quality Smooth Artless Uniform Angular	0.81
3	Quality	Generous Smooth Delicate Uniform Elegant	0.96
4	Smooth Delicate Hard	Smooth Delicate Hard Generous Uniform Quality	0.85
5	Simple Hard	Simple Hard Artless Angular	0.75
6	Quality Generous Simple	Quality Generous Simple Smooth Artless Uniform Elegant	0.97
7	Quality Simple Hard	Quality Simple Hard Generous Artless Uniform Angular	0.94
8	Uniform Hard	Uniform Hard Smooth Quality Generous Angular	0.90
9	Generous Simple	Generous Simple Quality Smooth Artless Uniform	0.88
10	Generous Rounded Simple	Generous Rounded Simple Quality Smooth Artless Uniform	0.96

Fig. 13. Average usage rate of 10 images

The average the adoption rate of the recommended tags is 89.6%. It is reasonable to recommend tags based on the emotional dimensions and emotional terms' relevance.

5 Conclusion

This paper initially envisages an image emotional semantic retrieval mechanism based on cognitive quantification model. Its core idea is to use semantic cognitive relevance of tags to divide some specific emotion into other relevant emotional dimensions and construct the emotional semantic cognition model.

At the same time, based on behavior psychology, tag generation channels are expanded by adding the users' retrieval behaviors which means "recognition", which provides more data for the modeling and make the model more representative.

As images need a lot of exposure to accumulate data to get a more accurate model, the idea of emotional semantic modeling is limited for cold-start images.

It is foreseeable that the theory of this research can be applied to other social digital resources, like music or video.

References

1. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM (2007)
2. Budanitsky, A., Hirst, G.: Evaluating wordnet-based measures of lexical semantic relatedness. Comput. Linguist. **32**(1), 13–47 (2006)
3. Cambria, E., Hussain, A., Havasi, C., Eckl, C.: Sentic computing: exploitation of common sense for the development of emotion-sensitive systems. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) Development of Multimodal Interfaces: Active Listening and Synchrony. LNCS, vol. 5967, pp. 148–156. Springer, Heidelberg (2010). doi:10.1007/978-3-642-12397-9_12
4. Fukumoto, T.: An analysis of image retrieval behavior for metadata type image database. Inf. Process. Manag. **42**(3), 723–728 (2006)
5. Hanbury, A.: A survey of methods for image annotation. J. Visual Lang. Comput. **19**(5), 617–627 (2008)
6. Li, X., et al.: Low-rank image tag completion with dual reconstruction structure preserved. Neurocomputing **173**, 425–433 (2016)
7. Liu, N., et al.: Multimodal recognition of visual concepts using histograms of textual concepts and selective weighted late fusion scheme. Comput. Vis. Image Underst. **117**(5), 493–512 (2013)
8. Liu, Y., et al.: A survey of content-based image retrieval with high-level semantics. Pattern Recogn. **40**(1), 262–282 (2007)
9. Lu, Y., et al.: A unified framework for semantics and feature based relevance feedback in image retrieval systems. In: Proceedings of the Eighth ACM International Conference on Multimedia. ACM (2000)
10. Matusiak, K.K.: Information seeking behavior in digital image collections: a cognitive approach. J. Acad. Librariansh. **32**(5), 479–488 (2006)
11. Müller, H., et al.: Learning from user behavior in image retrieval: application of market basket analysis. Int. J. Comput. Vis. **56**(1–2), 65–77 (2004)
12. Sánchez-Rada, J.F., Iglesias, C.A.: Onyx: a linked data approach to emotion representation. Inf. Process. Manag. **52**(1), 99–114 (2016)
13. Scherer, K.R.: What are emotions? And how can they be measured? Soc. Sci. Inf. **44**(4), 695–729 (2005)
14. Schmidt, S., Stock, W.G.: Collective indexing of emotions in images. A study in emotional information retrieval. J. Am. Soc. Inf. Sci. Technol. **60**(5), 863–876 (2009)
15. Schröder, M., Pirker, H., Lamolle, M., Burkhardt, F., Peter, C., Zovato, E.: Representing emotions and related states in technological systems. In: Cowie, R., Pelachaud, C., Petta, P. (eds.) Emotion-Oriented Systems, Cognitive Technologies, pp. 369–387. Springer, Heidelberg (2011)
16. Sun, A., et al.: Tag-based social image retrieval: an empirical evaluation. J. Am. Soc. Inf. Sci. Technol. **62**(12), 2364–2381 (2011)
17. Wang, W., He, Q.: A survey on emotional semantic image retrieval. In: 15th IEEE International Conference on Image Processing, ICIP 2008. IEEE (2008)
18. Wei-ning, W., et al.: Image retrieval by emotional semantics: a study of emotional space and feature extraction. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2006. IEEE (2006)
19. Wu, L., et al.: Tag completion for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 716–727 (2013)

20. Zha, Z.-J., et al.: Interactive social group recommendation for Flickr photos. Neurocomputing **105**, 30–37 (2013)
21. Zhang, H., Augilius, E., Honkela, T., Laaksonen, J., Gamper, H., Alene, H.: Analyzing emotional semantics of abstract art using low-level image features. In: Gama, J., Bradley, E., Hollmén, J. (eds.) IDA 2011. LNCS, vol. 7014, pp. 413–423. Springer, Heidelberg (2011). doi:10.1007/978-3-642-24800-9_38
22. Zimmermann, P., et al.: Affective computing – a rationale for measuring mood with mouse and keyboard. Int. J. Occup. Saf. Ergon. **9**(4), 539–551 (2003)