The Design of Images Database by Using Bootstrapping

Chin-Hwa Kuo, Nai-Lung Tsao, Yi-Fan Chen, and Yun-Long Lin

Department of Computer Science and Information Enginerring

Tamkang University, Taiwan

chkuo@mail.tku.edu.tw, beaktsao@mail2000.com.tw, emeth@ms31.url.com.tw, cloudlin@mail.tku.edu.tw

ABSTRACT

In this paper, we have designed a database that can automatically classify images; for the purpose of sorting through a large number of images more conveniently and thus save manpower and resources. This database is characterized by high level features to image classifying. Its features include: extending a keyword through bootstrapping construction. Whereas common ways of extending a keyword deal with its definition, bootstrapping construction allows expansion through associative extension. This type of keyword expansion mechanism is capable of classifying images in ways that WordNet cannot. Aside from using bootstrapping construction to expand keywords and to classify images, we have also added a discriminative feature metric to increase the precision and recall rates of image classifying to our standards. Through the use of bootstrapping construction, the user could greatly increase the precision and accuracy of grouping images while constructing the image database.

1: INTRODUCTIONS

Through the rapid development of information technology and heightened interest in the Internet in recent years, the Internet has become one of the most commonly used medium for sharing multimedia information. Among this information, images and graphics have always been extensively used on the Internet for various purposes, for instance: entertainment, online education, etc. Since the progress of Internet technology has allowed information sharing to become more convenient, in effect, the amount of images and graphics on the Internet has increased. Evidently, a mechanism for sorting through this immense amount of information is needed, accounting for the research and development of image databases.

The construction of an image database can be described as dividing a database into categories, then putting images into their respective classifications according to certain characteristics or features. Manually constructing such a database may ensure precision; however, it does require a significant amount of manpower. If a set of rules can let a computer automatically and correctly classify these images, a dearth of manpower would no longer be a problem.

Therefore, to be able to help the user efficiently construct an image or group while constructing the image database, we have use "bootstrapping". First we create categories, to focus on the few manually inputted keywords. Then these keywords are used as the search query, and new keywords are obtained from various web pages containing the original query. By repeating this cycle, a collection of image group names can be collected, which will help the next steps in image grouping. To improve the accuracy of image database, we use the discriminative feature metric to run the image database grouping. The purpose of this reason is to find out if bootstrapping keyword collection helps make the construction of an image database easier.

2: RELATED WORK

Bootstrapping [6, 12] refers to enabling the system to run on its own from the beginning to the end of preparation. The term comes from the idiom "to lift oneself by ones' own bootstraps", which means: the boot program must complete a task on its own. The basis of bootstrapping construction runs two factors in circulation: data and rules. It can find the rules in a minimal amount of known data. Using these rules, it then finds unknown data in which the same rules apply, and find a new set of rules. The process is repeated until enough data is collected. Fig 1 shows the concept of bootstrapping.

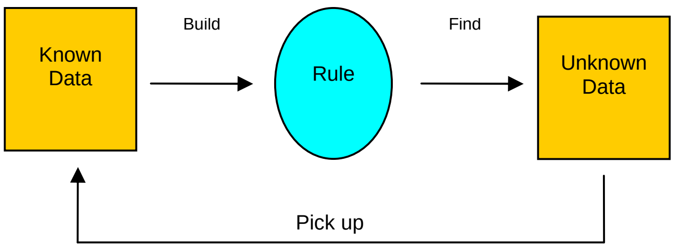

Fig 1 the concept of bootstrapping

Computer engineers have applied this concept to their design of a calculating method, hoping that from this concept, a solution to automatic or semi-automatic related problem can be addressed. Blum and Mitchell suggested a co-training method [2]. This is a cooperative type of work and automatic method. This basic concept behind this method lies in the construction of an independent condition. In a controllable mode, two automatic or bootstrapping construction processes can cooperatively work and find a better result in the problem.

In the information search, to be able to reach the information collected by the automatic or semi-automatic methods, many researchers have used this type of method to do data collection. The method described above is used as a solution to how to collect all the untagged information, to use these two classifiers to exchange new tag data. Feng and Chua [5] used the concept of this cooperative method to design their own bootstrapping construction calculation to collect for image data annotations and information. First, they used two independent classifiers, which have tagged data as their basis for expansion. They use SVN technology to undergo grouping. These two classifiers are different in that their information of interest are produced differently. One is from a color histogram, the other from texture and shape features. The designer used bootstrapping construction to first select the right conditions and ballot mechanism, and then put untagged data through these two classifiers to obtain keywords and annotations. Then the status of the classifiers are updated, letting the classifiers be able to recognize more data. This type of method can greatly save time and human labor.

On the other hand, Adami, Avesani, and Sona [1] have applied the bootstrapping method to categorizing document. First the author used SOM as a basis, designed a categorizing mode called TaxSOM. Then, the system will based on this categorizing mode, use bootstrapping construction to automatically collect untagged documents and put them in the appropriate document category. This way, a more systematic and automatic way of grouping documents can be achieved. After that, through the manual correction of an expert on the results, a qualified and complete categorization of images is done. This type of categorization uses a renewable training classifier, enabling the classifier to obtain better results next time it runs.

3: SYSTEM FRAMEWORK

In this section, the framework of bootstrapping construction for image sorting database will be introduced, and the sub system will be discussed in greater detail. Fig 2 shows the flows of data among each mechanisms and subsystems.

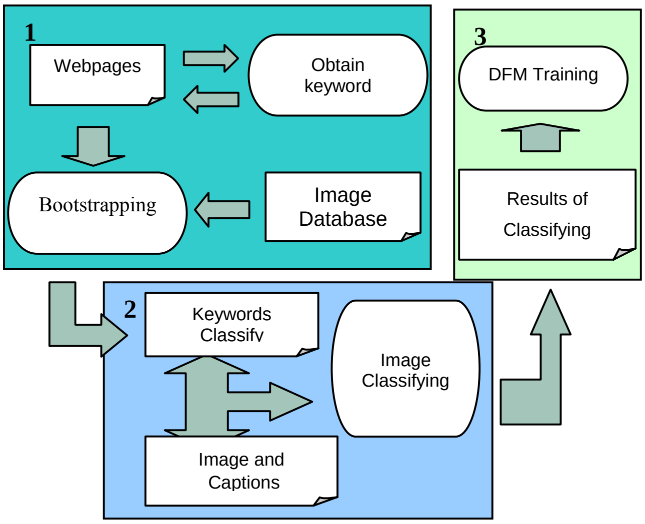

Fig 2 data flows in each subsystems

3.1: Image Classifying and Sub System Development

To ensure the precision of high level features and images and classifications, we have used a subsystem to insert keywords for the classifications. Aside from increasing the number of keywords, we must acknowledge their correct usage, as it is highly related to calculating the similarity value.

The database comprises two parts: classification database and bootstrapping construction (keyword expansion) system:

1. **Classification Database:**
   This is where data about image classifications is stored, the main data being classificatory labels/names and manually inserted keywords for the first classificatory labels.

2. **Bootstrapping Construction - Keyword Expansion System:**
   Although manually inserted keywords are indeed highly accurate for classificatory labels, they will not be enough to cover all the classifications. Furthermore, different users will have their own tone. This automatic system will solve these problems that occur in manual insertion.

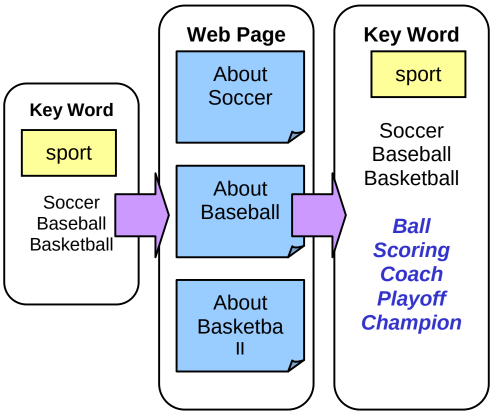

Fig 3 bootstrapping construction process

A keyword is manually inserted in the first step, shown in Fig 3, as the first keyword. Then, term frequency and invert document frequency (TFIDF) [11] run through the database, which currently stores 300,000 individual web pages, filtering out web pages without the keyword. The remaining web pages with the keyword are known as related web pages. Finally, the keyword within the related web pages that occurs with the highest frequency becomes the new Keyword, and the process is repeated, until no new keywords can be found, or until the number of keywords reaches a certain limit.

Bootstrapping Construction uses an associative method to do keyword expansion. The following is an example:

sport (by WordNet)
$\Rightarrow$ play、game、contest
sport (by bootstrapping construction)
$\Rightarrow$ basketball、baseball、soccer

If a picture of a baseball is needed, the only keywords in its captions will be related to baseball. Using WordNet [3, 8], the expanded keywords: play, game, and contest, cannot be used to find the picture. Bootstrapping construction, with its "associative" method, does not have this problem.

3.2: Primary Image Classifying System

This system aims to calculate the similarity value between the image and a classification, and from these calculations, do an initial classifying. Its purpose is to increase the number of image retrievals. Increasing the precision of image classifying will be the main purpose of the next system.

The procedure involved in classifying is calculating the similarity value between the image's captions and the keywords in the classification. This is done by applying the cosine rule, and calculating the angle between two vectors, shown in equation 1:

$$sim(d_j, q) = \frac{\vec{d_j} \bullet \vec{q}}{|\vec{d_j}| \times |\vec{q}|} = \frac{\sum_{i=1}^{t} w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^{t} w_{i,j}^2} \times \sqrt{\sum_{j=1}^{t} w_{i,q}^2}}$$

Equation 1

This is the well known vector space mode. The keyword i and searched document d_j uses $W_{i,j}$ to represent the similarity value, whereas the keyword i and search query q uses $W_{i,q}$ to represent its similarity values. The vectors $\vec{q}$ and $\vec{d}_j$ to respectively represent the search query q and searched document d_j.

When the similarity value is greater than a certain value, the image can be placed in the classification. Because of this, after an initial run of classifying, an image may be placed in more than one classification.

3.3 Image Training Subsystem

The classifying results from keyword expansion appear to have a high recall rate, but a high precision rate is also important. The main goal here is to increase the precision rate while making sure that the recall rate does not drop.

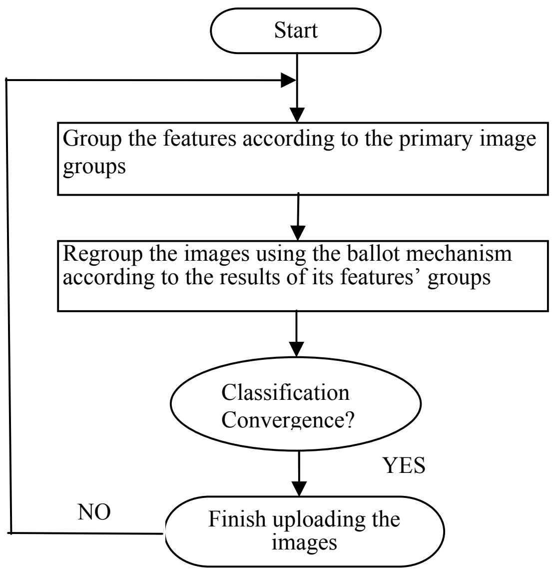

Fig 4 image training process

Fig 4 shows the image training process. The two main steps involved in this process are: using the discriminative feature metric to classify keywords that appear, then using the ballot mechanism to classify the images. Classifying technology like KNN rule [13] can be done by a person, according to the text [4], special features [7], and other manually exhausting classifying procedures. The discriminative feature metric method [9] focuses on classifying documents. The idea behind the method is to label the classifications according to different features, prior to classifying the data, then put the data into its respective classification.

Thus this process is divided into two procedures: the first being feature Classifying, the second being image classifying:

1. **Feature Classifying**
   From the results of the initial run, we can decide which classify each image belongs to. Because every image has keywords from the caption, the results of the initial classifying decide the number of times the keyword appears in its classification.

   The process of labeling classifications according to special features is done using the following equations:

$$DFM(f_i) = \log \frac{g_{in}(f_i)}{g_{out}(f_i)}$$

$$g_{in}(f_i) = \max(g(f_i, c_1), g(f_i, c_2), ..., g(f_i, c_k))$$

$$g_{out}(f_i) = \frac{\sum_j g(f_i, c_j) - g_{in}(f_i)}{k-1}$$

Equation 2

   In equation 2, f_1 represents the first feature in the first round. $g_{in}(f_i)$ represents the maximum number of times f_1 occurs in each classification. On the other hand, $g_{out}(f_i)$ represent the average number of times f_i occurs in other classifications. DFM(f_i) represent the similarity degree between f_i itself and the classification in which f_i occurs the maximum number of times.

   From the above equation, we find the classification the feature belongs to, as well as the similarity degree between the feature and its classification. Next, using these two pieces of known data, the classification in which the data belong to will be determined.

2. **Image Classifying**
   First, a threshold will be set to filter out DFM(f_i) values that are too small, to ensure that the feature and classification would have a certain level of similarity. Next, a ballot mechanism decides the classification in which each piece of data belongs to. Even if all the features have the same similarity value, the data will belong to the classification in which its feature belongs to.

   Before all the data is put in their respective classifications, new classifying results will occur. Therefore, all the classifications the features are in and DFM(f_i) value will change accordingly. This sub system will run through the same process until the classifications cease to change, meaning that the process has terminated.

4: SYSTEM IMPLEMENTATION, DISCUSSION AND APPLICATIONS

In regards to image grouping, the main thesis of this paper is to investigate the effect of the combination of the keywords expansion system of bootstrapping construction and discriminative feature method on image training system, to observe whether or not it can help improve image grouping. We used images taken from CNN webpage and images uploaded by a student from Tamkang University's English Department, with a total of 1271 images with English annotations. From the two following experiments, the thesis of this paper can be verified by using the differences between the results of the experiments.

4.1: The Effects of Bootstrapping Construction (keyword expansion) on Image Classifying

The experimental classification will be the keywords resulting from the bootstrapping construction step (after keywords have been extended). The control classification will be the manually constructed keywords inserted at the start. These two classifications will be compared to support whether or not this method is helpful to image classifying.

Again, the topic "sport" is used as the initial classifying. Different threshold values between (0.1 ~0.9) will be used to test the image's precision rate as well as its recall rate, to signify the differences between the two classifications.

Table 1 various threshold values and their respective rates

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Recall	96%/49%	92%/26%	78%/15%	46%/14%	23%/12%	15%/11%	11%/8%	7%/7%
Precision	71%/80%	70%/68%	67%/55%	57%/53%	65%/100%	65%/100%	62%/100%	100%/100%
F-measure	82/61	80/37	72/23	51/22	34/22	24/20	18/15	13/13

The blue figures in table 1 are the results of bootstrapping construction keyword expansion. The red figures are the results of not using bootstrapping construction. The data evidently shows that as the threshold value decreases, the precision rate does not alter much, but there is a great increase in the recall rate. F-measure [10] is defined as the sum of two pieces of data and the significance value, after evaluating the importance of recall rate and precision rate.

$$F = \frac{1}{\alpha \cdot \frac{1}{P} + (1 - \alpha) \cdot \frac{1}{R}}$$

Equation 3

In equation 3, the threshold of the precision rate (P) is the value α, while the threshold of the recall rate is 1-α. In different situations, the F values calculated from adjusting α would have different values.

Because the purpose of this step is to increase the recall rate of image classifying, the α value in this equation is set as 0.3, to signify the importance in recall rate. From the data collected in the experiment, the results are shown in the following graph.

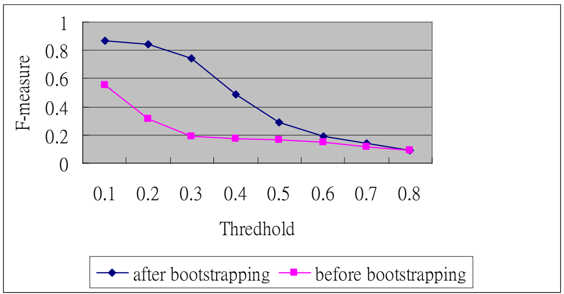

Fig 5 keyword expansion and F-measure

It is evident from Fig. 5 that after lowering the threshold, the F-measure of the experimental classification is significantly higher than the control classification. Therefore, using bootstrapping construction to extend keywords indeed increases the recall rate of image classifying.

4.2: Discriminative Feature Metric Training System and Its Effects on Image Classifying

The purpose of the image training system is to eliminate images that do not belong to a classification, even when the recall rate is high enough. In other words, it serves to increase the precision rate.

The experiment uses five image classifications that have already gone through bootstrapping construction and keyword expansion, to do image training. Different TFIDF Barrier values are used to compare the effect of Discriminative Feature Metric to do Image Training on precision rate and recall rate.

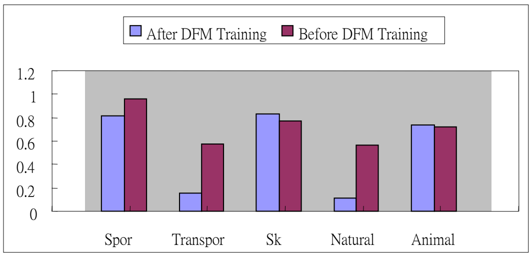

Fig. 6 shows the recall rate of six classifications before and after they have gone through training.

In Fig6, the recall rates of transport and natural were significantly lower after training. A probable cause of this phenomenon is that the recall rate before training was not very high, so during training, only a few keywords were obtained. Therefore after training, the recall rate of images decreased. This problem can perhaps be solved using a modified language database source.

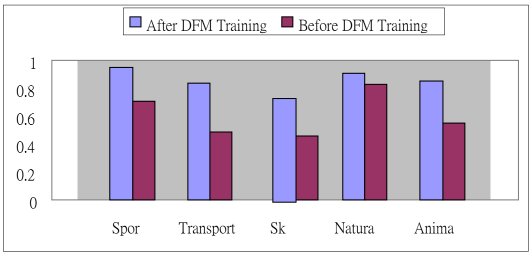

Fig.7 precision rates of six classifications before and after training

Fig.7 shows the precision rate of five classifications before and after training. From this graph, we can see that after training, regardless of the recall rate and precision rate at the start, the precision rate lies in a high zone.

This experiment clearly shows that after training, the recall rate can be sacrificed to increase the precision rate. Overall, this action would allow better results in image classifying.

5: CONCLUSION AND FUTURE RESEARCH

In this proposal, bootstrapping construction increased the number of keywords in each classification, allowing for greater keyword expansion to include all keywords that may be used. The Cosine Rule was then applied to do an initial classifying, so that the classifying results would maintain a high recall rate. The discriminative feature metric was then used to increase the precision rate of image classifying, to reach the standard of the image database. Using these two mechanisms and the image database, the precision rate of image classifying can be raised, and the original high recall rate can be maintained.

In future research, bootstrapping construction can be applied to increasing keywords in webpage material such as online news. Because of this, keywords extended after classifying are more similar to vocabulary used in the news. Therefore, news sites are more appropriate for finding images. If a large amount of documents can be obtained in the future, so that the sources of bootstrapping construction is not limited to only news websites, these documents can be used to expand bootstrapping construction, and even further increase recall and precision rate. On the other hand, discriminative feature metric uses the similarity of every individual keyword. In reality, every feature does not necessarily hold the same significance to each classification. Therefore, there should be an evaluation on this significance, in addition to the ballot mechanism. This would change and increase the precision rate of image classifying.

6: ACKNOWLEDGEMENT

This research is a part of project supported by National Science Council under grant number NSC93-2520-S-032-001. We really appreciate their support.

7: REFERENCES

[1] Adami, G., Avsani, P., and Sona, D., "Bootstrapping for Hierarchical Document Classification," In P*roceedings of the twelfth international conference on Information and knowledge management*, pp295-302, 2003.

[2] Blum, A., Mitchell, T., "Combined Labeled Data and Unlabeled Data with Co-Training," *In Proceedings of the 11th Annual Conference on Computational Learning Theory,* 1998.

[3] Chiang, Y.T., The Designing of a Semantic Image Retrieval System, *Master*'s *Thesis in Department of Computer Science and Information Engineering of Tamkang University, Taiwan,* 2002.

[4] Dumais, S., Platt, J., Heckerman D., and Sahami, M., "Inductive learning algorithms and representations for text categorization," *Proceedings of the 7th International Conference on Information and Knowledge Management*, Bethesda, MD, pp. 148-155, 1998.

[5] Feng, H.M., Chua, T.S., "A Bootstrapping Approach to Annotating Large Image Collection," ACM *SIGMM International Workshop on Multimedia Information Retrieval,* pp55-62, 2003

[6] Jones, R., McCallum, A., Nigam, K., and Riloff, E., " Bootstrapping for Text Learning Tasks," In *IJCAI-99 Workshop on Text Mining: Foundations, Techniques and Applications*, 52-63, 1999.

[7] Kim, K. I., Jung, K.C, Park, S.H., and Kim, H.J., "Support vector machines for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp.1542-1550, 2002.

[8] Lan, Y.H., Building Semantic Indexing for Images Retrieval Systems, Master*'s Thesis in Department of Computer Science and Information Engineering of Tamkang University, Taiwan*, 2004.

[9] Liu, X., Gong, Y.H., Xu, W., and Zhu. S.H., "Document clustering with cluster refinement and model selection capabilities," In *Proceedings of ACM SIGIR 2002*, Tampere, Finland, Aug. 2002.

[10] Manning, C., Schütze, H., Foundations of Statistical Natural Language Processing, 3rd ed, pp267-307, 2000.

[11] Manning, C., Schütze, H., Foundations of Statistical Natural Language Processing, 3rd ed, pp539-543, 2000.

[12] Thelen, M., Riloff, E.,"A Bootstrapping Method for Learning Semantic Lexicons Using Extraction Pattern Contexts," In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2002.

[13] Yu, C., Ooi, B.C., Tan, K.L., and Jagadish, H.V., "Indexing the Distance: An Efficient Method to KNN Processing," VLDB, 2001.