ENHANCEMENTS TO AN IMAGE RETRIEVAL SYSTEM WITH LONG TERM MEMORY LEARNING AND SHORT-TIME FEEDBACK

Stephen Adams

University of Minnesota, Morris

600 E 4th St.

Morris, MN 56649

adams601@morris.umn.edu

Abstract

This paper presents modifications to the image retrieval framework that was presented in [1]. These modifications attempted to make the established frameworks learning and feedback methods more precise. Experimental results are presented to show the effectiveness of our changes.

1. Introduction

Content based image retrieval (CBIR) systems have been researched and proposed for the past decade [1]. The challenge to create a CBIR system that can effectively return images based on their content is to reduce the gap between the low-level features of an image and the high level of human perceptions. [1] even goes so far to state that it is almost impossible to create similarity measures based on low level features to perform this task.

To reduce this gap [1] introduced a relevance feedback (RF) system. The idea of an RF system is to incorporate a user's perception into the query and to allow the user to evaluate the retrieval results. This is referred to as short term learning.

This system still suffers from three primary drawbacks. These are: the incapability of capturing semantic, scarcity, and imbalance of feedback examples. These challenges are addressed by [1] by incorporating long term learning.

Long term learning stores feedback that the user provides. This feedback helps improve later searches. The system proposed in [1] uses a "Fuzzy Semantic Relevance Matrix" (FSRM) to store the similarity between images. The FSRM is updated after a query to better represent the relationship between images.

The system proposed in [1] combines these short and long term learning methods for increased precision. In this paper modifications are proposed to both of these methods to try and increase [1]'s precision.

2. Related Work

This work is based off of the CBIR system proposed by [1].

3. Proposed Approach

This paper will present three modifications to [1]'s system. The first modification changes the long term learning procedure. The other two changes modify the short term learning method proposed by [1].

The FSRM is initialized with basic data about the low level features of an image. These low level features are taken from the method outlined in [2].

When a query is submitted to the system proposed in [1] the first step is to perform an initial retrieval. The initial retrieval is performed by looking through the FSRM three times selecting images on each pass. The first pass

through the FSRM selects images that are more similar to the query image than a predetermined threshold this set of images is called R_0. Once the first pass has been completed the second pass goes through R_0 and finds any image in the matrix that's similarity to an image in R_0 is greater than the threshold this set is known as R_1. Finally similar to how R_1 was found, R_2 are the images that are more similar to an image in R_1 than the threshold value.

After the initial image retrieval happens the user will label each image returned as either positive or negative. Positive images are those that feature what the user is looking for, negative images do not. Now the system will attempt to find more images to return. An image is returned if its average similarity to the positive images is greater than the threshold and the average similarity to the negative images is less than the threshold. These images are then returned to the user for labeling. The user can continue labeling and getting different results as long as they wanted. Once the user is finished labeling the system will perform its long term learning.

The long term learning process will update the FSRM to make future queries more accurate. The long term learning updates the similarity of each of the positive images to all of the images returned in the following way:

If the image j is positive then

$Sim_{i,j} = Sim_{i,j} + \alpha(1- Sim_{i,j})$

If the image j is negative then

$Sim_{i,j} = Sim_{i,j} – \beta* Sim_{i,j}$

Alpha is a coefficient known as the positive feedback coefficient and Beta is the negative feedback coefficient. The coefficients are preset and static. Our modification to the long term learning changes that.

3.1 New Long term Learning Approach

Our new long term approach hoped to adapt both α and β to how close or far away the results are. Instead of a static value the average similarity between all of the positive images and the query image is used for the value of α. Beta is calculated in a similar way except the negative results are used instead.

3.2 New Short Term Learning Approach #1

Our first short term method aimed filter out more of the images included in the short term feedback. We still move through all of the remaining images as before calculating an image's positive and negative averages. Instead of just using the threshold as a comparison the negative average is subtracted from the positive average. Then that difference is compared to a new threshold. Having this new threshold allows us to be much more precise, and really start to affect what is returned.

3.3 New Short Term Learning Approach #2

The second short term learning approach was inspired by the initial image retrieval. This method is quite a bit different than the original method. Once again the method goes through all of the remaining images. Then the similarities between each of the remaining images and the positive and negative images are looked up. Then these similarities are subtracted from the similarity between each positive and negative image and then the query image. These differences are then compared to 0.1. For the positive differences these differences must be less than or equal to one. For negative differences they need to be greater than or equal to 0.1. When these conditions are met that value is given a one, if not they are given a zero. Then the average of these ones and zeros are taken. If this average is greater than the threshold the image is added to the next set of results.

4. Experimental Results

For our experiment we used an image database of 2000 images. There were 20

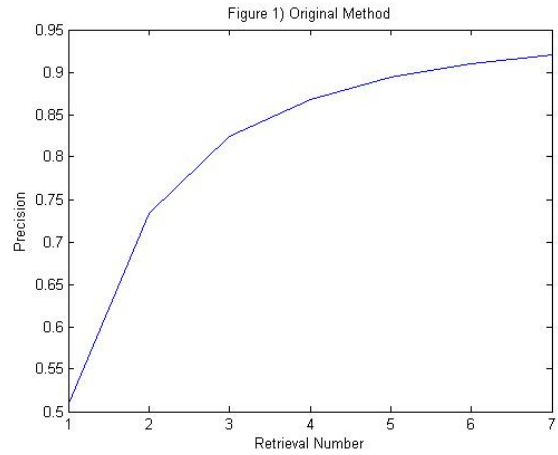
categories of images that we could query. Each image that was queried would have 6 short term training sessions. At each step only the 25 most similar positive images were kept.

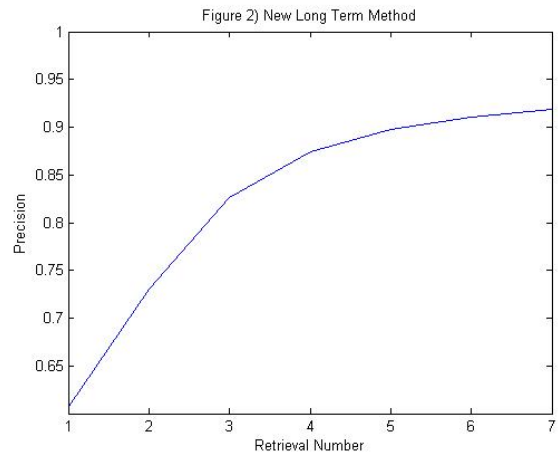
To avoid consecutive images being queried through the database a static random order was used. A list of the numbers from one to two thousand was generated and this list became the order that each image was referenced.

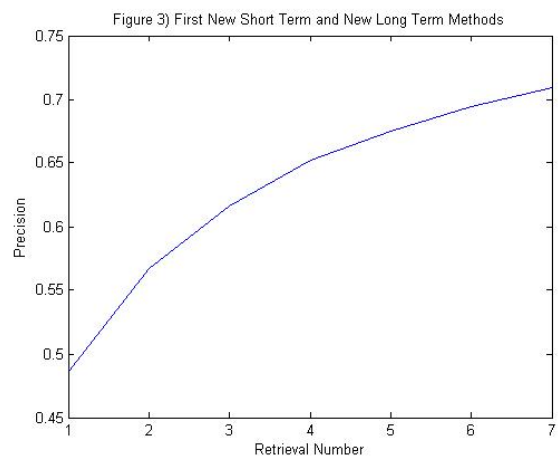
We measured the successfulness of a query based on its precision. Precision is the number of positive images retrieved divided by the total number of images retrieved.

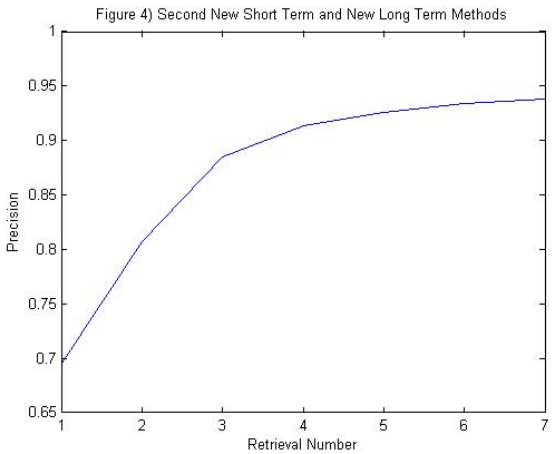
Here in figures one through four our results are plotted. These figures show the initial retrieval's precision followed by the precision after each of the six short term learning retrievals.

Figure one shows the original method from [1] without any of our changes. As you can see its precision tops out a little above 90 per cent. In figure two the results of our new Long term method show a small increase in precision. This small increase made us think that bigger increases where possible with a better short term learning procedure. This is what is shown in figures three and four. Both of these two figures both have the new long term methods running with one of the new short term methods. The first short term method has the opposite effect of what we wanted. When we look at the other method on the other hand the results are more positive. Though there is only a small improvement over the new long term method the adaptability of the second short term method was not explored fully.









5. Conclusion and Future Work

Though these initial results didn't return drastic improvements over the original method it is still possible these methods could be modified slightly for more improvements. The second short term method has a threshold which could be adjusted to increase precision. This is one thing that could be explored more in future work.

Since precision is already reaching a little above 90 per cent precision further improvement in that area will be difficult. After trying to improve the second short term method future work should move in a different direction.

The FSRM is a very large matrix. One of the best ways to improve this method could be to reduce the space required for the FSRM. This method does have room for improvement that will hopefully be accomplished soon.

6. Sources

1. Zhen Sun, Zhe-Ming Lu, and Hai-Jun Jin, "Image Retrieval with Long-Term Memory Learning and Short-Time Relevance Feedback." Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing, 2007.

2. Scott Fechser, Ran Chang, and Xiaojun Qi, "Inter-query Semantic Learning Approach to Image Retrieval," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'10), pp. 1246-1249, March 14-19, Dallas, Texas, USA, 2010.