Centrum voor Wiskunde en Informatica

REPORTRAPPORT

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA

Addressing special structure in the relevance feedback learning problem through aspect-based image search

M.J. Huiskes

REPORT PNA-R0410 DECEMBER 2004

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Addressing special structure in the relevance feedback learning problem through aspect-based image search

ABSTRACT

In this paper we focus on a number of issues regarding special structure in the relevance feedback learning problem, most notably the effects of image selection based on partial relevance on the clustering behavior of examples. We propose a simple scheme, aspect-based image search, which directly addresses these issues. The scheme additionally allows for natural simulation of the relevance feedback process. By means of simulation we analyze retrieval performance, sensitivity to feature errors, and demonstrate the value of taking into account partial relevance for a database of decoration designs.

2000 Mathematics Subject Classification: 68T10, 68Q32
1998 ACM Computing Classification System: H.3.3, I.2.6
Keywords and Phrases: content-based image retrieval; machine learning; relevance feedback; feature selection

Addressing special structure in the relevance feedback learning problem through aspect-based image search

M.J. Huiskes
Centre for Mathematics and Computer Science (CWI)
Kruislaan 413, 1098 SJ Amsterdam
Mark.Huiskes@cwi.nl

Abstract

In this paper we focus on a number of issues regarding special structure in the relevance feedback learning problem, most notably the effects of image selection based on partial relevance on the clustering behavior of examples. We propose a simple scheme, aspect-based image search, which directly addresses these issues. The scheme additionally allows for natural simulation of the relevance feedback process. By means of simulation we analyze retrieval performance, sensitivity to feature errors, and demonstrate the value of taking into account partial relevance for a database of decoration designs.

1. Introduction

As image content interpretation is both user- and task-dependent, content-based image retrieval (CBIR) revolves to an important extent around the task of *interactively* reaching an understanding of what a user is looking for. The most natural type of interaction is generally to ask the user to provide feedback regarding the relevance of results directly in terms of the presented images: by analyzing indicated relevant (positive) and irrelevant (negative) example images, the system may achieve an improved result in the next round. Direct feedback in terms of images is particularly convenient given that, unlike for text documents, relevance of images can truly be determined "at-a-glance". Recent reviews of the state-of-the-art of relevance feedback in CBIR are given in [14] and [13].

As it is clear that the importance of image features representing the image content will differ from query to query, much research has been aimed at *feature re-weighting* (e.g. [7], [8]). The aim is to use relevance feedback data to determine feature weights, in for instance similarity metrics or relevance measures, in accordance with the importance of the associated features to the user in that query. For example, [8] update weights of different feature classes by using the inverse variance of the positive examples, thereby giving higher weights to features for which the positives are relatively close together. Many variants of this approach have been proposed (e.g. [2], [5]) typically based on the idea of assigning higher weights to features in which positives cluster, while negatives remain separated.

In many recent approaches the feedback images are taken as training samples and are used to train a classifier or other learners for predicting the (ir)relevance of the database images. Typically two classes or levels of relevance are assumed; extensions to more levels are sometimes straightforward (e.g. [8]), but may incur the cost of a less natural interaction with the user. Examples of such learning approaches to relevance feedback are: support vector machines [11], boosting [10], decision trees [3], and nearest neighbors [12].

In the following we discuss some issues regarding the special nature of the relevance feedback learning problem. In particular we focus on the effects of example selection by *partial relevance* on the clustering behavior of examples. This will be especially important for retrieval systems for which after initial query specification typically no representatives of the desired target class are available, but only images that are relevant in some, partial, sense. Next we discuss a simple scheme that addresses this special structure, and introduce a novel simulation approach which is used to explore the feasibility and various measures of performance of the scheme.

Throughout this article we use a concrete application to illustrate concepts and procedures: a retrieval system for on-demand delivery of decoration designs, e.g. wallpaper or textile pattern designs; for the simulation study we use a commercial image database of tie designs. As main scenario we take a customer who is initially presented with a random selection of designs, say on a large screen in a store or through a web interface. By clicking the images he may express his preferences and dislikes, thereby iteratively guiding the system through the database based on his aesthetic appreciation and the designs he has in mind.

2. Special structure in the relevance feedback learning problem

As a learning problem we cannot treat relevance feedback analysis as a standard two-class, relevant versus irrelevant, classification problem; we mention the following issues:

Small sample learning problem. It has often been recognized (e.g. [14]) that the relevance feedback problem is a *small sample* learning problem. The number of example images depends on the willingness of the user to cooperate but is generally small, say at most 10 examples per feedback cycle, whereas the dimension of the feature space is large (often higher than 100).

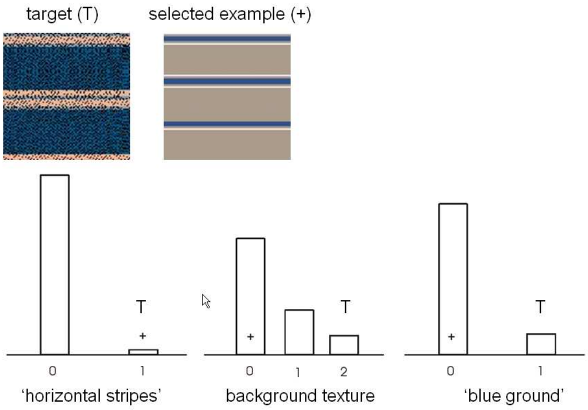

Figure 1: Shown are a target image (representing a simple user query for images of this type) and an image that the user has selected as a positive example. Also shown are histograms of database values for three (hypothetical) features: "presence of horizontal stripes", a feature measuring some characteristic of ground texture, and "presence of blue ground". The plus sign indicates the feature values of the example image; the T symbol indicates target values desired by the user. The example image is selected based on a single feature, viz. the possession of horizontal stripes. No positive feedback for other features was intended; as a consequence, such features will receive feedback on values that are (approximately) random draws from the feature value distributions. This often leads to misleading evidence, as is illustrated by the two other features shown here.

The small sample sizes typically encountered for relevance feedback disqualify many of the standard learning methods unless special measures are taken (e.g. [11]). As an aside we note that feature heterogeneity is often high as well and is to be taken into account: we must deal with a combination of numerical and discrete features, and with various high dimensional feature spaces with their own similarity metric (e.g. MPEG-7 descriptors, [4]). In our case, binary variables representing additional semantic categories obtained through manual annotation are also available.

Example selection by partial relevance. When a user selects an image as feedback he generally does so based on *partial* relevance of the image. This means that he finds one or a few aspects in that image to be of relevance; however, not all salient aspects present in the image need to be relevant, nor need all aspects of interest be present in the image.

For features other than those by which an image was chosen, which is often the large majority, the feedback received is thus to a large extent random: positive feedback is given for feature values, for which no such feedback was intended. As a consequence positive examples often give *misleading evidence*. This is illustrated in Figure 1.

Examples will tend to cluster at feature values that are most common in the database, thus interfering with the identification of the proper regions of relevance. This is related to the next issue.

Skewness of feature value distributions. Features often have value distributions that are highly skewed. This is particularly the case for features directly marking the presence of specific salient properties. As examples one may think of binary features such as "has-colored-stripes", "contains-a-paisley-motif" or "is-a-tartan". For many such features, the great majority of images will not possess the aspect thus leading to highly skewed value distributions. Also, if we take a feature measuring yellow-ness, say divided into three classes: "no yellow", "some yellow" and "very yellow", then by far most of the database images will be in the first class and, relatively, very few will be in the last.

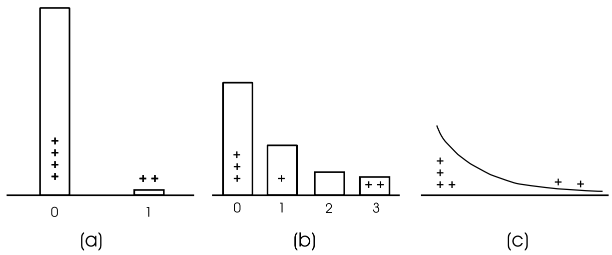

Figure 2: Diagram shows examples of feature value histograms (a, b), and a density function (c), and potential feedback for (a) a binary feature; (b) a discrete feature; and (c) a continuous feature. For each case, two positive examples were chosen based on the feature shown; the remaining examples are chosen based on other features.

The clustering effect of example selection by partial relevance is amplified by the skewness in feature value distributions as illustrated in Figure 2. Even though negative examples may counteract the misleading clustering of positives to some extent, learning methods will generally be influenced by the relatively small fraction of feature values for which feedback was actually intended, and the unintended concentration of positive examples. This also holds for many feature re-weighting approaches as they are usually based on the variation or clustering behavior of example feature values.

Not a two-class classification problem. Given the partial relevance issue described above it is already clear that viewing relevance feedback learning as a standard two class classification problem is an oversimplification. Images may be relevant in some aspects but not relevant in others, Also, if we do divide the images in two distinct classes, viz. those that are (fully) relevant and those that are not, we run into problems. First, true representatives of the former class are generally hard to come by initially, and second, the latter class will be very diverse and hard to represent with few examples.

Our main conclusion is that given that examples are often selected based on partial relevance, we must beware of its effects. The way we go about this will be by defining natural units of relevance, *aspects*, and their importance or *saliency* based on their frequency of occurrence in the database. These will allow us to take into account feedback on feature values only if it is sufficiently strong in comparison to feedback expected from neutral user behavior. In the following we discuss a simple scheme that implements these ideas.

3. Aspect-based image search

Features measure a variety of image quantities, where given a certain search task, some will matter to image relevance and others will not (neutral features). When a feature matters we should find out which feature *values* influence relevance positively, and which negatively. Note that only for neutral features, any feature value has (approximately) the same effect on image relevance, i.e. no effect. For "relevant features", not only will there be feature values that lead to higher perceived relevance, but there must always also be feature values that make images *less* relevant. We will not analyze the relevance of features as a whole, separately from the relevance assignment to feature values, but rather directly analyze the relevance of an image having feature values satisfying certain conditions or belonging to a certain set. We consider for instance the influence of "high complexity", where "high" is defined as a range of complexity feature values.

To be more precise, we will treat images as collections of "aspects", where we understand an aspect simply as a property which an image either has or has not, and for which we intend to resolve its effect on perceived relevance as a unit. Aspects can thus be explicitly defined in terms of a condition on feature values, but also live solely in the "eye of the beholder". In the following we will mainly consider the former type of aspects: derived binary features that model a specific perceptual quality.

There are two main reasons why we choose to employ aspects as an intermediary conceptual layer between the features and relevance estimates. First, it provides a simple and convenient framework for modeling partial relevance. Each aspect can be considered as either *neutral*, relevance enhancing (*positive*, or simply *relevant*) or relevance inhibiting (*negative*). In this way we can model a search task as a collection of positive, neutral and negative aspects. The relevance of an image can then be evaluated directly in terms of the aspects in this collection ("I appreciate this, but I do not like that.") Second, it allows us to associate a frequency of occurrence in the database to such "unit of relevance": for every aspect, we define the *aspect image frequency* p_{db} as the fraction of images in the database that possess the aspect. In the next section this measure will be related to the saliency of the aspect, allowing us to detect meaningful clustering and to discern neutral from positive and negative aspects. Construction of aspects is discussed in section 5.2.

As an illustrative example, suppose a user is interested in finding tie designs that: (i) have a blue background; (ii) have simple round motifs that are relatively far apart; and (iii) have high contrast between motifs and ground. Depending on the available features, we can translate this to requirements in terms of aspects. Some aspects are clearly positive, e.g. blue-ness of the ground should be high, dominant motif shape should be round, and relative amount of background should be high. Aspects in opposition to relevant aspects are negative, e.g. the user does not want squares, or a ground that is red. Additional negative aspects may come up during the feedback process, e.g. a user may decide that he is in fact not interested in yellow motifs. Other aspects are neutral, for example the user may not care about the pattern in the ground: it may be plain or have some texture.

4. Selection and relevance ranking

In the following we assume feedback example selection is facilitated by presenting images in clickable selection displays, each consisting of a grid of a fixed number of, say 50, thumbnail images. The number of images inspected per cycle may be larger as the user can leaf through the selection displays, or "reset" for a new random selection. Additional selection displays may be available, for instance offering "most-informative-images" (e.g. [14]). The sequential ordering of the images is either random in the first cycle (or based on an initial query mechanism, e.g. keyword-based), or by relevance ranking in subsequent cycles. The positive examples and (negative) counterexamples are collected in positive and negative *example sets*.

At each cycle of the feedback process the user updates the examples in the example sets by either: (i) selecting new images as positive or negative examples adding them to their respective sets; (ii) by removing images from the sets, i.e. the sets are preserved unless specific images are no longer deemed representative enough and are deleted explicitly.

In the aspect-based image search approach we use the feedback data available at the end of each cycle foremost to establish the effect (neutral, positive or negative) of the various aspects. The main idea on how to go about this is the following: as the user selects an image as feedback example based on one or a few positive or negative aspects, possession of the other aspects will approximately follow the distribution of aspect possession in the database. We are interested in finding those aspects for which the user has actively selected more examples with that aspect than may be expected to arise by chance only, i.e. as a side product of selection by other aspects. As for each aspect we know its associated aspect image frequency p_{db}, denoting the fraction of images in the database possessing that aspect, we can model the probability of the number of examples that would arise for a neutral aspect.

Taking this approach has the benefit that feature selection and, ultimately, relevance assignment is based not only on clustering behavior of positives and negatives, but may also be compared to clustering behavior of random database images. This leads to a natural emphasis on salient[1] aspects, effectively giving higher weights to those example feature values of the image that are more rare in the database. In addition, by taking into account feature value distributions, we are not dependent on negative examples to down-weight positives that cluster at aspects with low saliency. This means negatives can be used to indicate which aspects are not desired, but are not required for the sole purpose of getting sufficient data for classification.

4.1. Independence hypotheses

Let n^+ (n^-) be the total number of positive (negative) images selected, which we take to be fixed, and N^+ (N^-) be the number of positive (negative) examples that possess the aspect. For each aspect, we consider two hypotheses, H_0^+ and H_0^-, stating that the aspect behaves as if it were neutral to the user in regard to the accumulation of positive (resp. negative) examples. Under these hypotheses we model aspect possession of an example image as a Bernoulli variable with probability p_{db}; consequently, the number of positives and negatives with given aspect can be modeled as binomial variables with probability parameter p_{db}:

$$N^+ \sim B(n^+, p_{\mathrm{db}}), \quad \text{and} \quad N^- \sim B(n^-, p_{\mathrm{db}}). \qquad (1)$$

We intend to select aspects as positive or negative, only if there is sufficiently strong evidence supporting this decision relative to the independence hypotheses. We do so by first assessing the probabilities of finding the same or a higher number of example images with the given aspect as in the current example sets. If we select only those aspects for which these probability values are below a certain threshold, p_0^+ (resp. p_0^-), we limit the probability of the error of erroneously deciding that the aspect is not neutral.

More formally, we define two p-values associated with the respective hypotheses

$$p^+(N^+) = \sum_{i=N^+}^{n^+} \binom{n^+}{i} p_{\mathrm{db}}^i (1-p_{\mathrm{db}})^{(n^+-i)}, \qquad (2)$$

with $p^-(N^-)$ defined analogously. See also Figure 3.

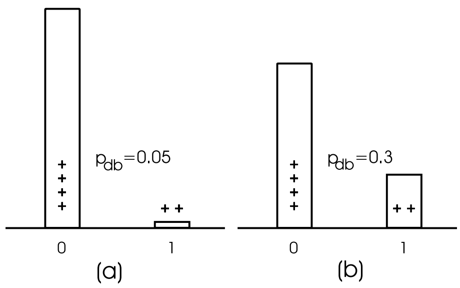

Figure 3: Shown are two aspects, a and b, with their respective aspect image frequencies p_{db}, and 6 positive feedback examples (+'s) for both aspects: 2 examples with aspect, 4 without. For aspect a the number of examples with aspect is unlikely if we assume the aspect is neutral to the user, indicating that this may be a positive aspect. For aspect b the feedback data does not contradict the independence hypothesis. Associated p-values are 0.033 and 0.58 , respectively.

4.2. Adaptive p-values

The p-values p_0^+ and p_0^- are probabilities of erroneously selecting an aspect as positive or negative when it is neutral.

[1]Note that saliency, in the sense of how rarely the aspect occurs in the database, is inversely related to image frequency. In analogy to "inverse document frequency" well-known in information retrieval we may quantify the importance or saliency of an aspect a through an inverse image frequency $\mathrm{iif}(a) = \log(1/p_{\mathrm{db}}(a))$, see [9].

When we reduce these values, thereby raising the number of examples required for selection, we increase the probability of missing actual positive and negative aspects.

As evidence is expected to accumulate in subsequent feedback cycles, we use the following dynamic p-value strategy. For the positive aspects we start with a relatively large p-value, say 0.05, in order not to miss relevant aspects when evidence is still relatively weak. After a number of feedback cycles (e.g. 3) evidence can be expected to have accumulated and the p-value is reduced, to say 0.001, in order to increase precision by avoiding false positive aspects. For negative aspects we take a small p-value (0.005) from the beginning, as negative feedback is necessary only when a certain aspect starts to accumulate in the display of highest ranking images, at which point sufficient examples will be available (see section 5). To monitor evidence accumulation more accurately, explicit user involvement is required e.g. by letting the user indicate fully relevant examples.

4.3. Relevance ranking

Let M be the aspect matrix with columns of boolean variables indicating if images have a given aspect or or not.

We can, trivially, determine N_j^+ and N_j^- from the image index sets S^+ and S^- of positive and negative examples, using sums $\sum_{i=1}^{n^+} M(S_i^+, j)$ and $\sum_{i=1}^{n^-} M(S_i^-, j)$ respectively, giving the two p-values, $p^+(N^+)$ and $p^-(N^-)$ by (2).

Let A^+ be the index set of accepted enhancing aspects, and A^- be the index set of accepted inhibiting aspects, then the relevance rel_i for image i is defined by $\mathrm{rel}_i = \sum_j M(i, A_j^+) - \sum_j M(i, A_j^-)$.

Note that, of course, the decision of taking into account an aspect need not be so black-or-white, and a variety of weighting schemes could be devised to obtain softer boundaries.

5. Simulation

5.1. Setup

Simulation of user behavior can proceed directly in terms of the aspects. We use the setup outlined in Figure 4.

The *aspect database simulation* component determines user and system aspects. The *user aspects* are the aspects as they are perceived by the user; they guide his interaction with the system, and are also used for *query generation*. The *system aspects* on the other hand are based on the features computed for the images; analysis of the feedback data is based on these aspects. The aspect database simulation and testing scenarios adopted are discussed in further detail in section 5.2.

The *interface simulation* and *user feedback simulation* together make up the user interaction simulation of the feedback cycle (see section 5.3). The output of this process are the positive and negative example sets, which are transformed by the *feedback analysis* component into a new relevance ranking of the database images.

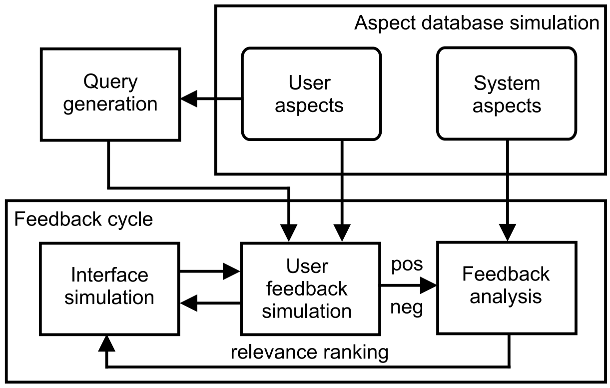

Figure 4: Components and their relations for an aspect-based simulation system (see text).

In the following we will be concerned mainly with analyzing two issues. First, as reliable and robust computation of perceptually interesting features is often difficult, we are interested in the effect of aspect errors on retrieval performance. Second, we are interested whether retrieval by aspect-based relevance feedback proceeds *regularly*. This will be discussed in more detail in section 5.3.

5.2. Aspect database simulation

We test retrieval performance for a database of decoration designs. We have designed and selected a variety of features suitable for representing the global appearance of designs; these include features for: color (e.g. dominant colors, saturation), texture, complexity and periodicity. In addition several features have been computed based on the decomposition of designs into figure and ground, e.g. relative amount of background, background texture, various properties of motifs (e.g. size, number, variation) and their spatial organization. Finally a set of 42 manually annotated semantic category labels (e.g."floral", "geometric") were also available.

Construction of aspects varies by feature type. Binary and discrete features can be converted directly into aspects. For single dimensional numerical features we use quantization. This can be done manually by inspection, or automatically. We have taken an automatic approach based on a grouping mechanism: we take a redundant group of aspects, defined at a number of locations and scales, and consider only the aspect with the smallest p-value of the group as a candidate for selection. High aspect redundancy is feasible as computational costs per aspects are very small.

For higher dimensional feature spaces our preferred solution is to take an exemplar or case-based approach. For instance, we have selected a number of simple example shapes as prototype shapes, and defined a "simple-motif" aspect by marking shapes that are close enough to one of the prototypes based on the similarity metric of the MPEG-7 contour shape descriptor ([1]). Another approach constructs data-driven aspects by mining for image clusters in feature spaces, where aspects again follow from cluster membership.

Numerical features are computed for a database of 1018 images that are representative in variety for a much larger (commercial) database. From the features, a total of 350 aspects were derived as described above. To obtain a test set of a size suitable for testing, we have used a sampling method for multivariate binary variables outlined in [6] to extend the aspect matrix to a total of 50000 images. The method simulates new images by sampling aspect values such that the overall aspect image frequencies and their correlations remain as in the original set of real images.

Simulation of aspect errors proceeds by defining two error probabilities, p_{10} and p_{01}, respectively for type 1 errors of not detecting an aspect in an image, and type 2 errors of assigning an image an aspect it does not possess. Based on these values we consider 5 testing scenarios: (I) no errors: user and system aspects are the same; (II) no type 1 errors; type 2 error rate such that number of images with aspect increases by 25%; (III) reversely, no type 2 errors, type 1 rate leading to a reduction in number of images that are assigned the aspects by a factor of 25%; (IV) equal errors $p_{10} = p_{01} = 5\%$; and (V) equal errors $p_{10} = p_{01} = 10\%$.

5.3. User interaction model

The query generation component determines a set of positive and negative aspects, called the target aspects; the user is assumed to treat the remaining aspects as neutral. Images that have the positive aspects, and do not have the negative aspects, are the *target images*.

User behavior with respect to new selection displays is determined by the partition of aspects, into positive, neutral and negative aspects, resulting from the query generation. For the selection of positive examples, we assume the user first ranks the display images and images already in the positive example set based on their number of positive aspects; additionally, a penalty factor (set to 2) is used for possession of negative aspects. He then chooses the positive example set based on this ranking: first, all target images are selected; next, further images are added in the order of decreasing ranking until there are no more images with positive score, or representation of the aspects is already sufficiently strong, i.e. each aspect has a sufficient number of representing images (e.g. 3). For selection of negative examples we assume the user actively pursues selection of examples for a negative aspect only if the selection display manifests a strong accumulation of images with such aspect. Once a negative aspect has become "active", the user selects examples avoiding positive aspects as much as possible.

If the user detects that he cannot improve his example sets by means of the selection display of highest ranked images, he may opt to leaf through more displays to find additional useful examples for aspects that have not yet sufficiently accumulated: either by proceeding down the ranking, or by choosing a new random selection from the database.

As we are interested if searching proceeds regularly, and we take the view that visiting additional displays interrupts the natural flow of a search, we monitor how often this is required per search as a measure of the irregularity of that search: for each search we define the number D_s of special displays that is visited (after at least one image has been found for each aspect).

A search ends after at most 10 cycles, or usually, at convergence: example sets cannot be improved and all positive aspects have sufficiently accumulated (at least 10 images out of 50).

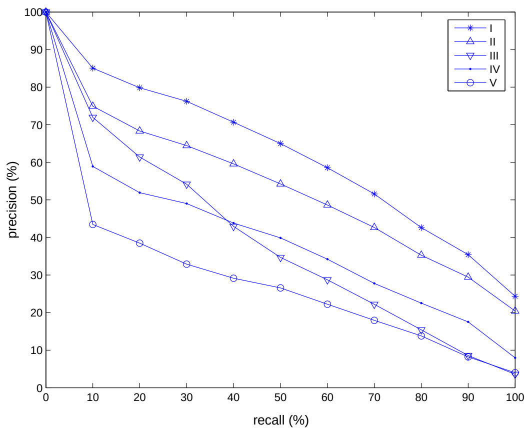

Figure 5: Precision-recall graphs for scenarios I through V (see section 5.2).

5.4. Results

Figure 5 shows the precision-recall graphs for the 5 scenarios outlined in section 5.2 based on aspects for 50000 images and 250 simulations. To obtain sufficiently many target images, query generation was based on 3 positive aspects, and 5 inhibiting aspects.

Performance deteriorates depending on the error types. Table 1 shows additional statistics for the five scenarios. For scenario I the average number of displays that is visited to find additional examples is only .21, i.e. about one display per 5 queries. This shows searches generally proceed regularly.

Scenario	T50	T100	1T	5T	D_s
I	67	74	2.6(3)	3.9(12)	.21
II	60	67	2.7(4)	3.8(20)	.10
III	46	54	2.9(10)	4.2(20)	.24
IV	48	55	2.6(11)	4.0(32)	.32
V	34	40	2.8(26)	4.5(45)	.39

Table 1: Statistics for scenarios I through V (see section 5.2). T50: average percentage of target images with ranking within first 50 highest ranked images (1 display); T100: average percentage of target images within first 100 highest ranked images (2 displays); 1T: average number of cycles required to get one target image on first display given such target is found; in braces is the percentage of experiments in which no such target was found; 5T: average number of cycles required to get 5 target images on first display given 5 targets are found; D_s: average D_s-value per query as defined in section 5.3.

Table 2 demonstrates the effectiveness of the two-stage p-value strategy described in section 4.2. If we take a fixed p-value $p = 0.05$, precision decreases substantially as may be expected as the probability of erroneously selecting positive aspects is rather high. On the other hand if the p-value is set to $p = 0.001$ from the beginning, precision is the same as for the two-stage p-value, but the average regularity of the search is reduced, i.e. the user needs to visit more displays to find suitable example images.

Strategy	T50	T100	1T	5T	D_s
two-stage	67	74	2.6(3)	3.9(12)	.21
$p = 0.05$	37	42	2.5(8)	3.5(35)	.07
$p = 0.001$	67	73	2.6(7)	3.9(13)	.53

Table 2: Statistics comparing performance of three different p-value strategies. For an explanation of the captions, see Table 1.

6. Conclusion

Relevance feedback by example selection based on partial relevance is natural user behavior that must be accommodated for in the design of retrieval systems. The aspect-based approach proposed here does so by making sure feedback on feature values is accepted only once evidential support that the user actually intended this feedback is sufficiently strong.

First simulation results confirm the feasibility of this approach. Generally few positive examples are required, and there is a regular progression to the target class. A further interesting property is the lack of need of negative examples solely for obtaining sufficient data for classification.

Selection of positive and negative aspects is based on a comparison of user behavior to the case that he were neutral to the given aspect. What constitutes neutral user behavior will depend on the search *context*, e.g. which database is used, which subset the user is interested in, or the peculiarities in the search behavior of a given user.

The use of context-conditional aspect collections and context-dependent aspect image frequencies offer an excellent opportunity to adapt the search system to user context. This will be explored in future work. Further work will be directed at more extensive simulations and detailed comparison to other learning methods. Also we intend to study generalizations such as fuzzy aspect possession, and alternative relevance ranking schemes.

References

[1] Miroslaw Bober, F. Preteux, and Whoi-Yul Yura Kim. Shape descriptors. In B.S Manjunath, Philippe Salembier, and Thomas Sikora, editors, *Introduction to MPEG-7 – multimedia content description interface*, pages 231–260. John Wiley and Sons, Ltd., Chichester, England, 2002.

[2] G. Ciocca and R. Schettini. Using a relevance feedback mechanism to improve content-based image retrieval. In D.P. Huijsmans and A.W.M Smeulders, editors, *Visual information and information systems*, pages 107–114. Springer Verlag, 1999.

[3] S.D. MacArthur, C.E. Brodley, and C.R. Shyu. Relevance feedback decision trees in content-based image retrieval. *IEEE Workshop on content-based access of image and video libraries*, pages 68–72, 2000.

[4] B.S. Manjunath and T. Sikora. Overview of visual descriptors. In B.S Manjunath, Philippe Salembier, and Thomas Sikora, editors, *Introduction to MPEG-7 – multimedia content description interface*, pages 231–260. John Wiley and Sons, Ltd., Chichester, England, 2002.

[5] J. Peng, B. Bhanu, and S. Qing. Probabilistic feature relevance learning for content-based image retrieval. *Computer Vision and Image Understanding*, 75(1/2):150–164, 1999.

[6] B.F. Qaqish. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika*, 90(2):455–463, 2003.

[7] J.J. Rocchio Jr. Relevance feedback in information retrieval. In G. Salton, editor, *The SMART retrieval system: experiments in automatic document processing*, pages 313–323. Prentice-Hall, 1971.

[8] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Trans. Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[9] K. Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972.

[10] K. Tieu and P. Viola. Boosting image retrieval. *International Journal of Computer Vision*, 56(1/2):17–36, 2004.

[11] S. Tong and E. Chang. Support vector machine active learning for image retrieval. *Proc. of 9th ACM Int'l Conf. on Multimedia*, pages 107–118, 2001.

[12] P. Wu and B.S. Manjunath. Adaptive nearest neighbor search for relevance feedback in large image databases. *Proc. of 9th ACM Int'l Conf. on Multimedia*, pages 89–97, 2001.

[13] H. Zhang, C. Zheng, M. Li, and Z. Su. Relevance feedback and learning in content-based image search. *World Wide Web: Internet and web information systems*, 6:131–155, 2003.

[14] X.S. Zhou and T.S. Huang. Relevance feedback in image retrieval: a comprehensive review. *ACM Multimedia Systems Journal*, 8(6):536–544, April 2003.