M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 359
Also appears: ACM Computing Surveys

Computer learning of subjectivity*

Rosalind W. Picard

MIT Media Lab; 20 Ames St., Cambridge, MA 02139
picard@media.mit.edu; http://www.media.mit.edu/~picard/

1 Subjectivity in multimedia retrieval and annotation

"When I say 'city scene, " said Arthur, "I want a lot of skyscrapers, a skyline; the prototypical city – you know, cars, smog."

"When I think of a city," said Beth, "I think of beautiful cities – with nicely landscaped green areas around each building, and colorful people milling about."

Consider how many variations there are on the query "Computer, find a good city scene." Researchers working in multimedia information retrieval (including browsing and annotation) are well aware that the *theme in the user's head* is subjective, and therefore will rarely match the stored annotations. Moreover, the theme changes as the user browses through data, so that ultimately it cannot be characterized by a static learning system, but must be continuously learned.

The earliest systems designed for multimedia retrieval [1], and those that have become commercially available, either ignore subjectivity and use fixed algorithms, or require the user to adjust parameters. For example, in user-adaptive systems, before making a query, the user specifies what percent of the query is color and *which* colors, what percent is texture or shape and *which* textures or shapes, etc. The human is stuck with the task of transcoding the theme in his or her head into a set of slider and control knob positions.

Consequently, for years, I have advocated the importance of finding models of images and sounds that have "semantic control knobs." For example, the Wold model for retrieving perceptually similar visual patterns has knobs corresponding to periodicity, directionality, and randomness [2].

*Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

However, rarely can models with semantic control knobs be found. Even when they exist, it is an effort to know how to optimally set them. Moreover, usually a person does not make a single query, but a succession of queries, with slight variations each time. Therefore, she not only needs to know how to set the knobs when initiating a query session, but also how to adjust them with each new query.

What I have described is the current trend in content-based retrieval and annotation systems, and it needs to change. The system must recognize that the user's goals evolve while they browse; subjectivity, mood-dependence, and fickleness are to be expected. Furthermore, a system that tracks the evolving goals of a subjective human will also be helpful for the difficult but common query sessions best described as "I'll know it when I see it."

2 Learning to deal with subjectivity

Can a computer system deal intelligently with a user's subjectivity? It will be difficult, but it can be achieved. Let me briefly outline a strategy and highlight some recent progress.

First, consider how humans deal with subjectivity, such as how you learn your friend's preferences in movies: (1) you have common language and knowledge allowing you to share opinions, (2) you observe his or her choices, and (3) you *learn*. A criterion of success is the ability to accurately predict movies your friend likes, not perfectly, but much better than chance.

The strategy I propose for computers is similar. Specifically, they will need to (1) share some of the common sense of the user, (2) observe and model the user's actions, and (3) learn from these interactions. Although these three goals comprise immense problems, and much research remains to be done, there has been substantial progress toward each.

Lenat has perhaps done the most work toward the first goal, developing common sense reasoning systems. A brief description of his latest effort related to information retrieval is in [3]. His system works with text, allowing a request for "somebody wet" to retrieve an annotated shot such as "Garcia finishing a marathon."

The second goal, based on experiments by Minka and Picard, appears to be best satisfied by a sys-

tem that works with a *society of models* [4]. The society of models is an interacting set of heterogeneous models – parametric, non-parametric, semantic, or even user-provided. The models communicate via common *groupings* of data, perhaps provided by a hierarchical clustering according to the model parameters, or by a text thesaurus, or even by a user who clusters data subjectively under the label, "these look good together." The society of models enables the system to model a wide variety of user's actions, fulfilling the second goal.

The third goal, learning, works with the entire system to improve performance as the user interacts with it, typically by giving it positive and negative examples of what he wants, and feedback on its performance. If the user has taught the computer how to label and retrieve images of trees, and now shows it unlabeled photos that include trees, then the learned system should now be faster at finding and labeling the trees.

Other systems in information retrieval (IR) have also attempted to implement learning with methods such as relevance feedback. The system we have built, "FourEyes," [4] differs from relevance feedback methods such as [5] in that FourEyes operates directly on features computed from images, as opposed to operating on attributes provided by humans. FourEyes also uses a nonlinear learning algorithm, vs. the linear weighting algorithms commonly used in the IR community. The most novel aspect of FourEyes' learning algorithm however, is a dynamic bias. Different biases, or sets of weights, have been shown to give significant improvement in learning performance [4].

FourEyes achieves several levels of learning subjectivity in retrieval. When FourEyes sees a problem it has seen before, it automatically switches to a bias that it learned for that problem. When it sees a significantly new problem, then it learns a new bias. It therefore behaves differently over time, depending on what it has been exposed to. FourEyes has three stages that learn at different rates, from interactive-speed online learning, to longer-term evaluative offline learning. FourEyes learns groupings from a subjective human trying to retrieve, segment, or annotate multimedia data. The groupings that it learns ultimately form a representation of low-level *perceptual* common sense [6] for a user, especially as the user attaches semantic labels (annotations) to the data.

To summarize, subjectivity – here described as the evolving theme in the user's head – is expressed as the user interacts with the information retrieval system during a succession of queries. These evolving inputs of the user can be tracked, modeled, and used to retrieve data consistent with changing requests, by use of a society of models coupled with a learning algorithm. The system we have built achieves subjectivity at several levels – especially by finding "similar" groupings where similarity is defined by the user's current subjective theme. Additaionlly, it learns a dynamic bias, allowing it to develop different behaviors to different problems it sees. The result is a subjective learning system which can collect information across multiple users. The results should assist not only in the improvement of multimedia information retrieval systems, but also in furthering basic understanding of subjectivity in information retrieval.

References

[1] B. Furht, S. W. Smoliar, and H.-J. Zhang, *Video and Image Processing in Multimedia Systems.* Kluwer Academic Publishers, 1995.

[2] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *Submitted for Publication*, 1995. Also appears as MIT Media Laboratory Perceptual Computing TR#320.

[3] D. B. Lenat, "Artificial intelligence," *Scientific American*, pp. 80–82, Sept. 1995.

[4] T. P. Minka and R. W. Picard, "Interactive learning using a 'society of models'," *Submitted for Publication*, 1995. Also appears as MIT Media Lab Perceptual Computing TR#349.

[5] G. Jung and V. Gudivada, "Distributed adaptive attribute-based image retrieval," in *Digital Image Storage and Archiving Systems, Photonics East*, SPIE, October 1995.

[6] R. W. Picard, "Toward a visual thesaurus," in *Springer-Verlag Workshops in Computing*, 1995. To appear; Also appears as MIT Media Lab Perceptual Computing TR #358.