Learning Task-Specific Similarity

by

Gregory Shakhnarovich

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

Author .....
Department of Electrical Engineering and Computer Science
September 30, 2005

Certified by .....
Trevor J. Darrell
Associate Professor
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Learning Task-Specific Similarity

by

Gregory Shakhnarovich

Submitted to the Department of Electrical Engineering and Computer Science
on September 30, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

The right measure of similarity between examples is important in many areas of computer science. In particular it is a critical component in example-based learning methods. Similarity is commonly defined in terms of a conventional distance function, but such a definition does not necessarily capture the inherent meaning of similarity, which tends to depend on the underlying task. We develop an algorithmic approach to learning similarity from examples of what objects are deemed similar according to the task-specific notion of similarity at hand, as well as optional negative examples. Our learning algorithm constructs, in a greedy fashion, an encoding of the data. This encoding can be seen as an embedding into a space, where a weighted Hamming distance is correlated with the unknown similarity. This allows us to predict when two previously unseen examples are similar and, importantly, to efficiently search a very large database for examples similar to a query.

This approach is tested on a set of standard machine learning benchmark problems. The model of similarity learned with our algorithm provides and improvement over standard example-based classification and regression. We also apply this framework to problems in computer vision: articulated pose estimation of humans from single images, articulated tracking in video, and matching image regions subject to generic visual similarity.

Thesis Supervisor: Trevor J. Darrell
Title: Associate Professor

Acknowledgments

I would like to thank my advisor, Trevor Darrell, for his influential role in this thesis and in my graduate career. Being Trevor's student has been fun in many ways; he has struck the perfect balance in his role as an advisor. On the one hand, he has given me a great deal of independence in pursuing my ideas and lots of encouragement. On the other hand he has provided thorough advice (on things academic and not) and, at times, challenging the nonsense I would produce. Not less importantly, Trevor has succeeded in building a really great research group, where things go smoothly professionally and socially.

Bill Freeman, Mike Collins and Shimon Ullman have been great thesis readers, and their careful and challenging feedback has made the thesis much better than it would otherwise be. I am also thankful for their patience with my ever shifting deadlines. My gratitude also goes to Tommi Jaakkola for many helpful conversations about learning, and for the enriching experience of TA-ing his machine learning course, and to Piotr Indyk, whose work has inspired much of the work described in the thesis and who was always willing to meet and talk. John Fisher has been an excellent colleague and friend, and I thank him for his appreciation of my Soviet jokes (one in particular).

My work and thinking has also been benefited by a number of collaborations outside of MIT. A summer internship at MERL with Paul Viola taught me many things about vision, learning and scientific intuition. Paul has contributed some of the ideas in the core of this thesis, and has collaborated with Trevor and me on the work presented in Chapter 4. I have also benefited from collaboration with Baback Moghaddam, Liu Ren, Jessica Hodgins, Paul Viola and Hanspeter Pfister; Liu led much of the effort in the last stages of the project described in Section 4.5.

During the five years I have spent at MIT, I have learned a great deal from courses and talks, but probably more so from the many seemingly random conversations with fellow students, staff and faculty. Naturally, the most frequent victims have been my officemates. I am very grateful to Leonid Taycher, Kristen Grauman and Ariadna Quattoni for many interesting conversations, for being kind and helpful in all matters, and for humoring me by listening to my ramblings and tolerating the (often bad) jokes. The other members of Trevor's research group have made my time here more productive and enjoyable. I would like to especially mention Mario Christoudias, David Demirdjian, Louis-Philippe Morency, Ali Rahimi, Kate Saenko

and Mike Siracusa who were always willing to talk about research, philosophy, politics and food. The work described in Section 5.3 is a collaborative effort with David, Leonid and Kristen. Many other people at MIT have been a privilege to know. I would especially like to thank Biswajit Bose, Polina Golland, Alex Ihler, Kevin Murphy, Bryan Russell, Erik Sudderth, Marshall Tappen, Kinh Tieu, Antonio Torralba and Lior Wolf for many useful and fun conversations. I would also like to acknowledge Rahul Sukhtankar, Yan Ke and Derek Hoiem from CMU for helpful conversations and for bringing to my attention the observation leading to the semi-supervised setup in Section 3.2.4.

Last but not least I would like to thank Misha, Olga, Gabi and Dorel who have provided the emotional backup (and occasionally nutrition and shelter) one needs in the journey through graduate school. My very special thanks of course go to my mother Lena, who I hope will be proud. Karen, nothing I could write here about how grateful I am to you for everything would be expressive enough!

Contents

List of Figures

List of Tables

List of Algorithms