

Research Methods in computer science

Spring 2019

Lecture 14

Omprakash Gnawali

March 4, 2019

Agenda

HW6 Live Grading

Conference Organization

Experiments and metrics

Reading Papers

HW7

Types of Experiments

Model / Analysis

Simulations

Testbed (Real world ^{lite})

“Real world”

Which one to use when?

Metric

Why do we want to measure?

What to measure?

Most of the time we measure improvements

Eigenfaces for Recognition

[Turk '91]

“We have developed a near-real-time computer system that can locate and track a subject’s head, and then recognize the person by comparing the characteristics of the face to those of known individuals.”

Scenarios and metrics from [Turk '91]

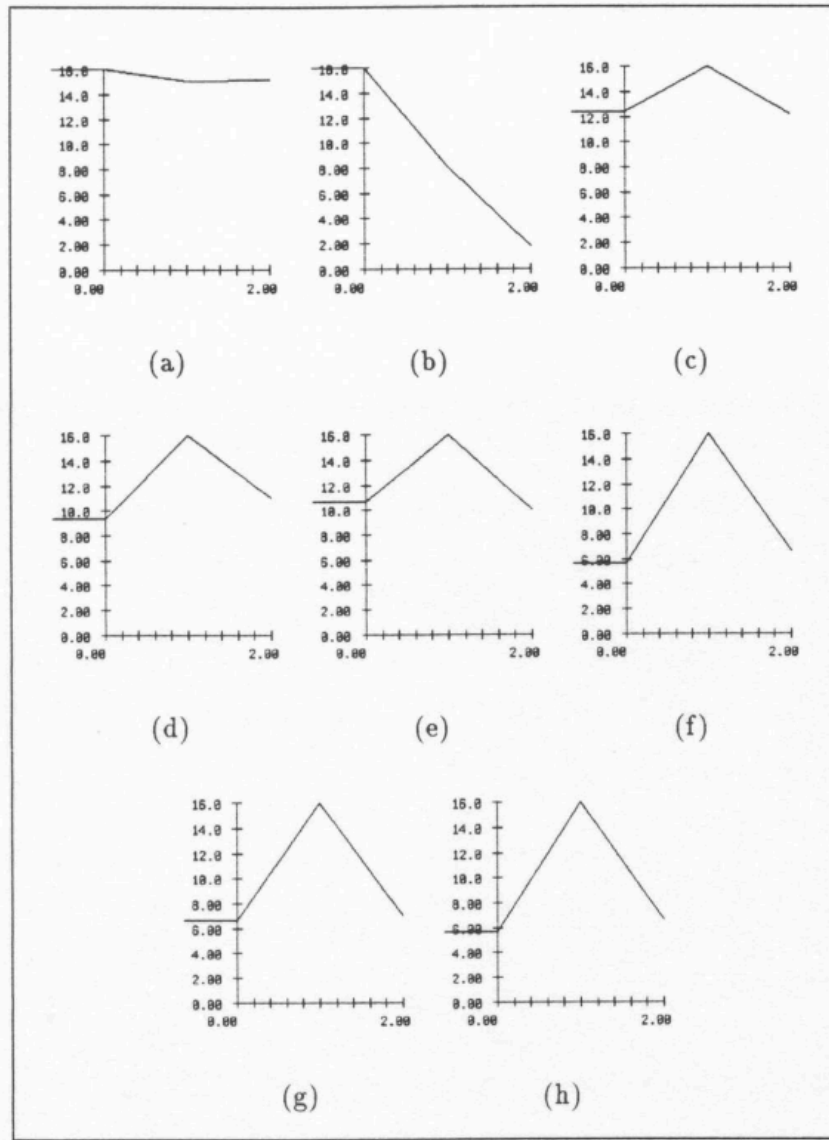


Figure 9. Results of experiments measuring recognition performance using eigenfaces. Each graph shows averaged performance as the lighting conditions, head size, and head orientation vary—the y-axis depicts number of correct classifications (out of 16). The peak (16/16 correct) in each graph results from recognizing the particular training set perfectly. The other two graph points reveal the decline in performance as the following parameters are varied: **(a)** lighting, **(b)** head size (scale), **(c)** orientation, **(d)** orientation and lighting, **(e)** orientation and size (#1), **(f)** orientation and size (#2), **(g)** size and lighting, **(h)** size and lighting (#2).

The Anatomy of a Large-Scale Hypertextual Web Search Engine

[Brin and Page '98]

What hypothesis, scenarios, and metrics should we expect to see in this paper?

5 Results and Performance

The most important measure of a search engine is the quality of its search results. While a complete user evaluation is beyond the scope of this paper, our own experience with Google has shown it to produce better results than the major commercial search engines for most searches. As an example which illustrates the use of PageRank, anchor text, and proximity, Figure 4 shows Google's results for a search on "bill clinton". These results demonstrates some of Google's features. The results are clustered by server. This helps considerably when sifting through result sets. A number of results are from the whitehouse.gov domain which is what one may reasonably expect from such a search. Currently, most major commercial search engines do not return any results from whitehouse.gov, much less the right ones. Notice that there is no title for the first result. This is because it was not crawled. Instead, Google relied on anchor text to determine this was a good answer to the query. Similarly, the fifth result is an email address which, of course, is not crawlable. It is also a result of anchor text.

All of the results are reasonably high quality pages and, at last check, none were broken links. This is largely because they all have high PageRank. The PageRanks are the percentages in red along with bar graphs. Finally, there are no results about a Bill other than Clinton or about a Clinton other than Bill. This is because we place heavy importance on the proximity of word occurrences. Of course a true test of the quality of a search engine would involve an extensive user study or results analysis which we do not have room for here. Instead, we invite the reader to try Google for themselves at <http://google.stanford.edu>.



[Brin and Page '98]

Storage Statistics	
Total Size of Fetched Pages	147.8 GB
Compressed Repository	53.5 GB
Short Inverted Index	4.1 GB
Full Inverted Index	37.2 GB
Lexicon	293 MB
Temporary Anchor Data (not in total)	6.6 GB
Document Index Incl. Variable Width Data	9.7 GB
Links Database	3.9 GB
Total Without Repository	55.2 GB
Total With Repository	108.7 GB

Web Page Statistics	
Number of Web Pages Fetched	24 million
Number of Urls Seen	76.5 million
Number of Email Addresses	1.7 million
Number of 404's	1.6 million

Table 1. Statistics

[Brin and Page '98]

Why did the authors
decide to report these
measurements?

Metrics/Experiments?

Accurately Initializing Real Time Clocks to Provide Synchronized Time in Sensor Networks

CTP: An Efficient, Robust, and Reliable Collection Tree Protocol for Wireless Sensor Networks

On the Effectiveness of Energy Metering on Every Node

Surviving Sensor Network Software Faults

Metrics from Classification Research

Classification Accuracy

Logarithmic Loss

Area Under ROC Curve

Confusion Matrix

Classification Report

Precision

Recall

F1-Score

Partly from <https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/>

Metrics from Regression Research

Mean Absolute Error

Mean Squared Error

R^2

Partly from <https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/>

Metrics from Systems Research

Reliability

Latency

Coverage

Energy

Group Activity

Experiment Design

Metric Selection

Experiments

What experiments are useful?

Critical for the main arguments of the paper

What experiments are not useful?

Pointless experiments that generate pointless numbers, graphs, and tables

Types of Experiments

From the “context” perspective

Controlled

Uncontrolled

There are other perspectives to be covered in future lectures

A new algorithm that translates English text to Spanish.

A new wireless networking technology.

A new algorithm that can identify the person in an image.

HW7 – Intro & Related Work

Write a draft Introduction and Related Work for your paper. It should be about 2 pages in length. Use the template to be used in the conference.