

Research Methods in computer science

Spring 2017

Lecture 14

Omprakash Gnawali
March 6, 2017

Agenda

HW6 live grading

Research Conference Updates

Sampling/Dataset Bias

HW8

Sampling Bias

“In statistics, sampling bias is a bias in which a sample is collected in such a way that some members of the intended population are less likely to be included than others. It results in a biased sample, a non-random sample of a population (or non-human factors) in which all individuals, or instances, were not equally likely to have been selected. If this is not accounted for, results can be erroneously attributed to the phenomenon under study rather than to the method of sampling.” -- wikipedia

Dataset Bias in Object Recognition Research

Unbiased Look at Dataset Bias [CVPR 2011]

Object Recognition Research

Dataset is a set of pictures of objects

Run algorithm to recognize/identify objects

Compute accuracy or other metrics

What are potential dataset bias?

Sampling, Capture, Negative Set

How to reduce selection bias
in visual object recognition datasets?

Research that uses online/social media data

Research: how people communicate, spread information, discuss, decide, etc.

What are some potential bias in the dataset?

What are the implications?

Datasets and consequences

How Vector Space Mathematics Reveals the
Hidden Sexism in Language

[https://www.technologyreview.com/s/602025/
how-vector-space-mathematics-reveals-the-
hidden-sexism-in-language/](https://www.technologyreview.com/s/602025/how-vector-space-mathematics-reveals-the-hidden-sexism-in-language/)

Exercise

Think of a big-data application

Identify a dataset on which you want to do
“machine learning”

Describe potential bias there in the
application

HW8 –Introduction Patterns

1. For each of the ten papers that you have identified as related to your research, write down the question the author is trying to answer in each paragraph.
2. Do interesting statistics and visualization to describe how introductions are written in your sub-discipline or research area based on the data you collected in part1.