Writing a paper & Picking Projects

CS 197 | Stanford University | Kanishk Gandhi
cs197.stanford.edu

Slides adapted from Michael Bernstein


AI AND
HUMANS

Writing a paper & Picking Projects

CS 197 | Stanford University | Kanishk Gandhi
cs197.stanford.edu

Slides adapted from Michael Bernstein

Today’s goals

We have a bunch of things we tried, some of them worked, some of them didn’t — how do we write a paper about this?

Introducing the concept of model papers and how to use them

How do I pick projects to work on, going forward?

Writing A Paper


Scene Graph Prediction with Limited Labels
How do we get here?

The common misunderstanding

work
work
work
coffee
work
work
imposter syndrome
work


arXiv:1904.11622v2 [cs.CV] 20 Aug 2019

Scene Graph Prediction with Limited Labels

Vincent S. Chen, Paroma Varma, Ranjay Krishna, Michael Bernstein, Christopher Ré, Li Fei-Fei
Stanford University
{vincentsc, paroma, ranjaykrishna, msb, chrismre, feifeili}@cs.stanford.edu

Abstract

Visual knowledge bases such as Visual Genome power numerous applications in computer vision, including visual question answering and captioning, but suffer from sparse, incomplete relationships. All scene graph models to date are limited to training on a small set of visual relationships that have thousands of training labels each. Hiring human annotators is expensive, and using textual knowledge base completion methods are incompatible with visual data. In this paper, we introduce a semi-supervised method that assigns probabilistic relationship labels to a large number of unlabeled images using few labeled examples. We analyze visual relationships to suggest two types of image-agnostic features that are used to generate noisy heuristics, whose outputs are aggregated using a factor graph-based generative model. With as few as 10 labeled examples per relationship, the generative model creates enough training data to train any existing state-of-the-art scene graph model. We demonstrate that our method outperforms all baseline approaches on scene graph prediction by 5.16 recall@100 for PREDCLS. *In our limited label setting, we define a complexity metric for relationships that serves as an indicator ($R^2 = 0.778$) for conditions under which our method succeeds over transfer learning, the de-facto approach for training with limited labels.*

1. Introduction

In an effort to formalize a structured representation for images, Visual Genome [27] defined **scene graphs**, a formalization similar to those widely used to represent knowledge bases [13, 18, 56]. Scene graphs encode objects (e.g. `person`, `bike`) as nodes connected via pairwise relationships (e.g., `riding`) as edges. This formalization has led to state-of-the-art models in image captioning [3], image retrieval [25, 42], visual question answering [24], relationship modeling [26] and image generation [23]. However, all existing scene graph models ignore more than 98% of relationship categories that do not have sufficient labeled instances (see Figure 2) and instead focus on modeling the

Limited labels
Unlabeled images
Our semi-supevised method
Probabilistic training labels
Eat: 0.35
Eat: 0.7
Eat: 0.85
Eat: 0.3
Any existing scene graph model

Figure 1. Our semi-supervised method automatically generates probabilistic relationship labels to train any scene graph model.

few relationships that have thousands of labels [31, 49, 54].

Hiring more human workers is an ineffective solution to labeling relationships because image annotation is so tedious that seemingly obvious labels are left unannotated. To complement human annotators, traditional text-based knowledge completion tasks have leveraged numerous semi-supervised or distant supervision approaches [6, 7, 17, 34]. These methods find syntactical or lexical patterns from a small labeled set to extract missing relationships from a large unlabeled set. In text, pattern-based methods are successful, as relationships in text are usually **document-agnostic** (e.g. `<Tokyo - is capital of - Japan>`). Visual relationships are often incidental: they depend on the contents of the particular image they appear in. Therefore, methods that rely on external knowledge or on patterns over concepts (e.g. most instances of `dog` next to `frisbee` are `playing with` it) do not generalize well. The inability to utilize the progress in text-based methods necessitates specialized methods for visual knowledge.

In this paper, we automatically generate missing relationships labels using a small, labeled dataset and use these generated labels to train downstream scene graph models (see Figure 1). We begin by exploring how to define **image-agnostic** features for relationships so they follow patterns across images. For example, `eat` usually consists of one object consuming another object smaller than itself, whereas `look` often consists of common objects: `phone`, `laptop`, or `window` (see Figure 3). These rules are not dependent on raw pixel values; they can be derived from image-agnostic features like object categories and relative spatial positions between objects in a relationship. While such rules are simple, their capacity to provide supervision for unannotated relationships has been unexplored. While image-agnostic

1

Why is this a misunderstanding?

Research papers are complex documents, with too many degrees of freedom to “just write”. Being strategic will save time and avoid dead ends.

…so what do we do instead?

There are many genres

Even within areas, there exist many different genres of paper. Each genre is typically built around the claim you are making, and implies a structure to the sections and to the writing. For example:

We solve a problem: articulate the problem, explain what causes that problem and what others have done to deal with it, detail your approach, and prove that you make progress on the problem

We measure an outcome: explain that nobody has bothered understanding how a phenomenon behaves, explain how to create a study that sheds light, and report the outcomes of it

We introduce a technique: articulate a problem as above, but focus the narrative on the technique you've created, since it will generalize

Genres imply structure

Common "We Solve A Problem" structure:

But, this will vary by area!

Introduction: overview and thesis

Related Work: situate your contribution relative to prior research

Approach: describe your approach and important implementation details

Evaluation: test whether your approach succeeds at its stated goals

Method

Results

Discussion: reflect on limitations, implications, and future work

Conclusion: summarize and restate your contribution

“Which genre is our project?”

You can often derive the appropriate genre in the same way that you derived the evaluation — what is the thesis and claim that you are supporting?

But this may be challenging until you’ve read a large number of papers. So instead…

Model papers

A model paper is a paper that you can use as a model or template for constructing your paper.

You should be able to structure your paper in the same way as your model paper

Follow its general flow of argument in the introduction

Use similar section and subsection heading organization

Create figures, tables, and graphs that fulfill the same function as theirs

Apply the same general proportions, e.g., number of pages per section

Selecting your model paper

Model paper != nearest neighbor paper

The model paper should be a paper that makes the same type of argument as yours. It should be in the same genre as you seek.

Often the nearest neighbor paper will make a similar form of argument, but not necessarily

Often the nearest neighbor paper will be a well-written paper, but not necessarily

Find your model paper and share it with your TA for a thumbs up before writing.

From model to paper

Start by reverse-outlining the model paper.

How does it structure its argument into sections?

What is the main expository goal of each section? What is its sub-thesis?

What role does each figure play?

From model to paper

Next, build a mapping from their outline to yours.

Translate each section and sub-section heading into what the equivalent heading is for you

Translate each sub-thesis into what the equivalent sub-thesis is for you

Translate each figure into what the equivalent figure is for you

What if it doesn't quite fit?

Model papers should be templates, not straightjackets. You will probably need to adapt your mapping slightly from what your model paper does.

e.g., you require a slightly different evaluation structure or visualization than them

e.g., you're drawing on a different literature than them, and need to explain something that they didn't

You can play with the genre — just don't discard the genre. Check with your TA for any substantial changes that you want to make.

Assignment 5: draft paper

Work together with your team to write a draft paper. This should be a complete draft in the template format of your research, and include reviewable drafts of every section.

"Can we include text we already wrote?" Absolutely! + tweaks

"Do we need the results of our evaluation?" Yes, but you can continue to update your results through the final deadline.

"What if our project doesn't work out?" Still write up the report. Negative results can be valuable. Unpack in Discussion what it was about your idea or assumptions that wasn't borne out.

After this, Assignment 6 will be a draft talk.

Picking Projects

Where do research ideas come from?

A common mindset: riffing

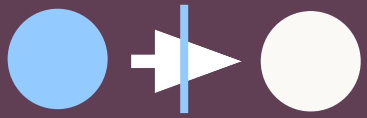
Ye Olde Riffing Recipe, let the researcher cook:

Read a bunch of papers

Pick a paper you really like

Ask yourself: how could I extend this to another domain, or make progress on one of its challenging assumptions, or otherwise extend it?

This is a process for generating a one-paper bit flip



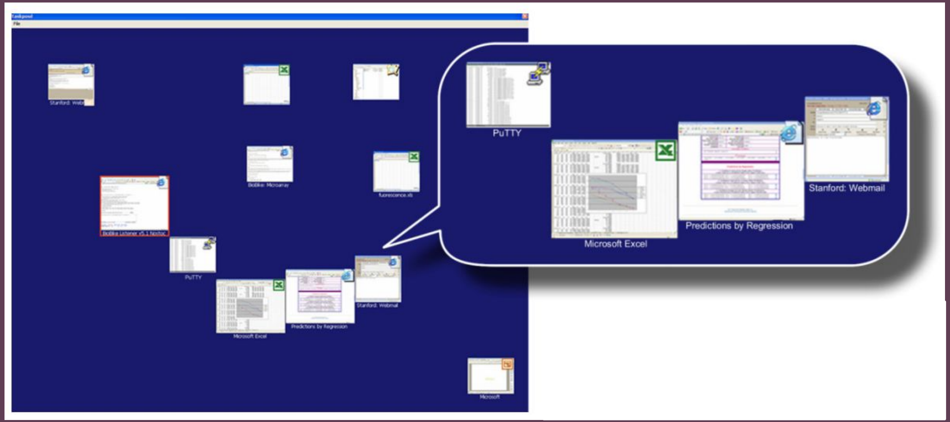
Riffing is often a good starting point for a first independent project

It places focus on execution, and gives you most of the inputs, outputs, and constraints—the assumptions—up front

Even for experienced researchers

Lots of work on task-centric workspaces

MSB: "But tasks can have fuzzy boundaries!"

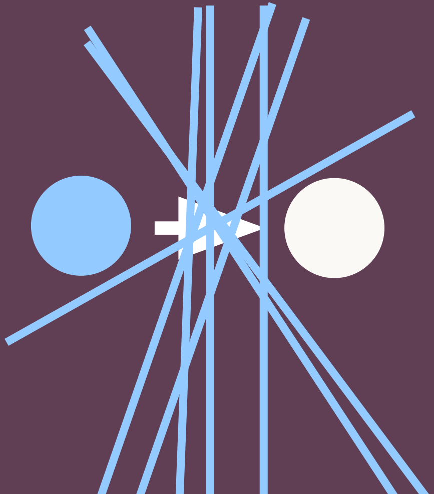


MSB (Michael Bernstein)

What are the risks here?

It's not clear that all bit flips are worthwhile.

A misappropriated quote: *"Your scientists were so preoccupied with whether or not they could that they didn't stop to think if they should."*



"Salami Science": possibility of incremental work when we don't view the field's assumptions broadly

What we mean when we say “incremental”

Research and science are not neutral: they embed values

Incrementally is a push back against minor adjustments to models that don’t build substantial theory

What we mean when we say science isn't neutral

Science and Technology Studies (STS) establishes that what counts as a contribution, or as major vs. incremental, or even what counts as Computer Science, is socially constructed by elites in the field.

Not so long ago, HCI and Ethics were not seen as legitimate CS

Also not so long ago, CS itself was not seen as a legitimate field

Objection to creating a CS department at Stanford, via Leo Guibas:
"We don't have a department of Refrigerator Science!"

Thanks to Jingyi Li!

So what should we do instead of only riffing on papers?

Desert Metaphor

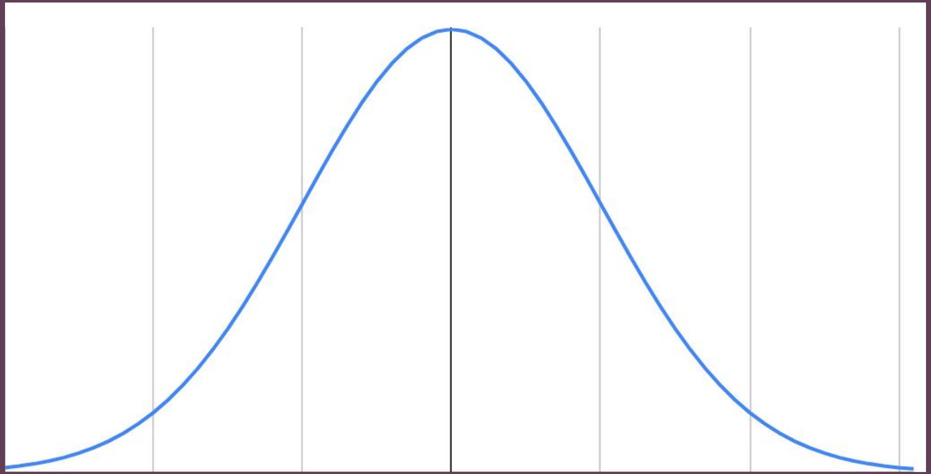
Is this a big rock?

Do I have an angle on it?

"If you want to have a good idea, you must have many ideas."

– Nobel Prize winning chemist Linus Pauling

“If you want to have a good idea, you must have many ideas.”



$2 \cdot \sigma$ = 95% of samples

$3 \cdot \sigma$ = 99.7% of samples

Some Strategies and Stories

Rage-based research

When a pattern or underlying assumption in the field starts to dig at you until you decide to prove that it's wrong.

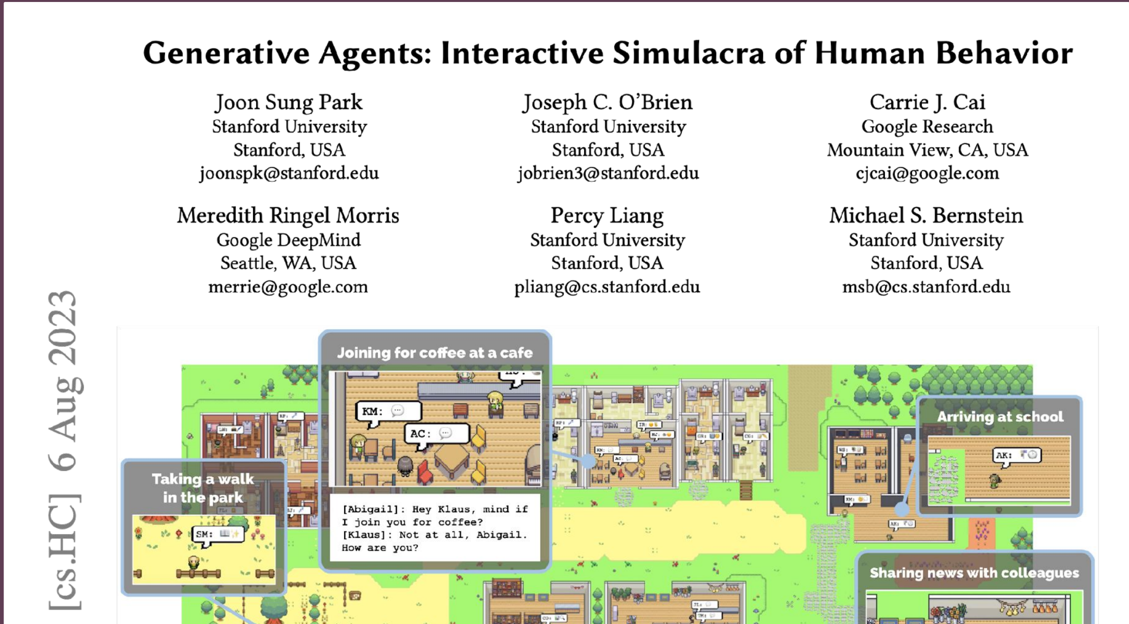
Understanding Social Reasoning in Language Models with Language Models

Kanishk Gandhi * **J.-Philipp Fränken** * **Tobias Gerstenberg** **Noah D. Goodman**
Stanford University
{kanishk.gandhi, jphilipp}@stanford.edu

Abstract

As Large Language Models (LLMs) become increasingly integrated into our everyday lives, understanding their ability to comprehend human mental states becomes critical for ensuring effective interactions. However, despite the recent attempts to assess the Theory-of-Mind (ToM) reasoning capabilities of LLMs, the degree to which these models can align with human ToM remains a nuanced topic of exploration. This is primarily due to two distinct challenges: (1) the presence of inconsistent results from previous evaluations, and (2) concerns surrounding the validity of existing evaluation methodologies. To address these challenges, we present a novel framework for procedurally generating evaluations *with* LLMs by populating causal templates. Using our framework, we create a new social reasoning benchmark (**BigToM**) *for* LLMs which consists of 25 controls and 5,000

When new tools reopen old problems


Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

Meredith Ringel Morris
Google DeepMind
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu

[cs.HC] 6 Aug 2023

When you see a new north star

Social Contract AI: Aligning AI Assistants with Implicit Group Norms

Jan-Philipp Fränken, Sam Kwok[†], Peixuan Ye[†], Kanishk Gandhi

Dilip Arumugam, Jared Moore, Alex Tamkin

Tobias Gerstenberg, Noah D. Goodman
Stanford University
jphilipp@stanford.edu

Abstract

We explore the idea of aligning an AI assistant by inverting a model of users' (unknown) preferences from observed interactions. To validate our proposal, we run proof-of-concept simulations in the economic *ultimatum game*, formalizing user preferences as policies that guide the actions of simulated players. We find that the AI assistant accurately *aligns* its behavior to match standard policies from the economic literature (e.g., selfish, altruistic). However, the assistant's learned policies lack robustness and exhibit limited *generalization* in an out-of-distribution setting when confronted with a currency (e.g., grams of medicine) that was not included in the assistant's training distribution. Additionally, we find that when there is *inconsistency* in the relationship between language use and an unknown policy (e.g., an altruistic policy combined with rude language), the assistant's

When you see a new north star

Searching for Computer Vision North Stars

Li Fei-Fei & Ranjay Krishna

Computer vision is one of the most fundamental areas of artificial intelligence research. It has contributed to the tremendous progress in the recent deep learning revolution in AI. In this essay, we provide a perspective of the recent evolution of object recognition in computer vision, a flagship research topic that led to the breakthrough data set of ImageNet and its ensuing algorithm developments. We argue that much of this progress is rooted in the pursuit of research "north stars," wherein researchers focus on critical problems of a scientific discipline that can galvanize major efforts and groundbreaking progress. Following the success of ImageNet and object recognition, we observe a number of exciting areas of research and a growing list of north star problems to tackle. This essay recounts the brief history of ImageNet, its related work, and the follow-up progress. The goal is to inspire more north star work to advance the field, and AI at large.

Pulling the thread on a weird result

Jury Learning: Integrating Dissenting Voices into Machine Learning Models

Mitchell L. Gordon
Stanford University
Stanford, USA
mgord@cs.stanford.edu

Michelle S. Lam
Stanford University
Stanford, USA
mlam4@stanford.edu

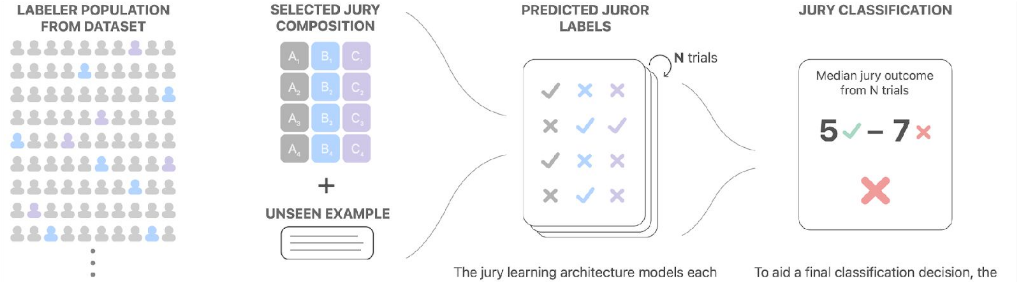
Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Kayur Patel
Apple Inc.
Seattle, USA
kayur@apple.com

Jeffrey T. Hancock
Stanford University
Stanford, USA
hancockj@stanford.edu

Tatsunori Hashimoto
Stanford University
Stanford, USA
tatsu@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu



In this dataset, the labeler population consists of labelers

The decisionmaker composes a jury to rule on input examples (here, they

The jury learning architecture models each individual labeler in the dataset. Jury learning then samples labelers to fill the

To aid a final classification decision, the model surfaces the median jury outcome over multiple trials (each with re-sampled

Playing a hunch: “Hey, would it be possible to…”

Strategic Reasoning with Language Models

Kanishk Gandhi **Dorsa Sadigh** **Noah D. Goodman**
Stanford University
kanishk.gandhi@stanford.edu

Abstract

Strategic reasoning enables agents to cooperate, communicate, and compete with

Pulling the thread on a weird result

Eliciting Compatible Demonstrations for Multi-Human Imitation Learning

Kanishk Gandhi, Siddharth Karamcheti, Madeline Liao, Dorsa Sadigh
Department of Computer Science, Stanford University
{kanishk.gandhi, skaramcheti, madelineliao, dorsa}@stanford.edu

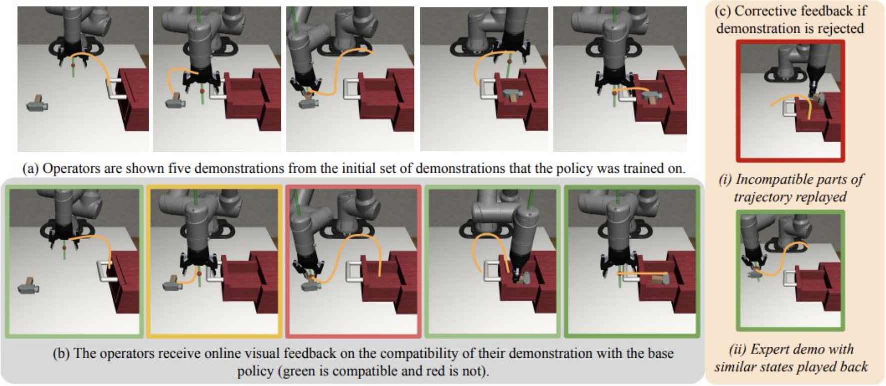


Figure 3: The three phases of our active elicitation interface spanning the initial *prompting* phase (a), the subsequent *demonstration* phase with live feedback (b), and finally, the *feedback* phase (c).

Which approach do I apply?

This is a skill you develop through mentorship — it’s highly contingent, and depends on the problem and solution space that you’re navigating.

My suggestion: try on different hats around the problems you’re interested in, and see what works.

One final note:

people >> projects

Writing a paper & Picking Projects