MODELS AND ALGORITHMS FOR GROUNDED, INTERACTIVE SEMANTICS

BY BRIAN MCMAHAN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Matthew Stone

and approved by

____

____

____

____

New Brunswick, New Jersey

June, 2017

Table of Contents

Chapter 1

Introduction

An important aspect of human communication is understanding how people use language to describe the world (Quine, 1960)

People use words to describe and refer to entities in the worlds. The nature of this relationship is a long-standing curiousity in the scientific community.

A perennial challenge in the scientific understanding of language is understanding how people relate words to the world.

Human communication crucially involves people using language to describe the world.

People categorize the world

Words are used to in the world.

The relationship between words and the world is a long-standing

In the process of communicating knowledge about the world, people use words to denote categories. Specifically, people can use words like *dog* and *tree* to refer to objects in the world. Many words, however, do not have clear category boundaries and are used flexibly to describe subsets of objects in the world. For example, *tall* buildings and *fat* pigs. This class of words, known in the linguistic community as VAGUENESS, is an important facet of grounded language use. People leverage vague words to communicate ideas—broadly ranging from prototypical examples (such as *blue* describing a pure blue) to borderline cases (such as *blue* describing a blue-green color). Thus, vague language is a referential shortcut that remains underdetermined until it is made more specific in a context.

Parameterizing vague language in models of grounded language use is vital for grounded communciation.

Chapter 2

Related Work

2.1 Meaning as Classifiers

In this section, I discuss the intuitions underlying grounded semantics. Specifically, I introduce the idea that the meaning of words can be thought of as clasification functions over perceptual stimuli.

There is a wide range of work that has taken this perspective.

There is also a running thought: words have meaning insofar as they refer to the thing the speaker had meant. This is a meaning-by-denotation or meaning-as-extension. The meaning of a phrase is just the set of things that it picks out.

I don't think that this means meaning is only loaded into the extension—for example, a person may have meant a different thing and say "I mean something else", but that person is able to negotiate what they meant. So, meaning has an intention, a purpose, and a reception.

The earliest grounded semantics work is SHRDLU (Winograd). In this, words are thought to map to the world deterministically and as a boolean value.

There are several other lines which saught to investigate different aspects of linguistic phenomena. Such cases include the semantic parsing community. The more recent semantic parsing people (Krishnamurthy,Zettlemoyer) have relaxed this view and taking the classification view.

The classification view captures the duality of meaning I described above. That is, the meaning of a word is a classifier. A classifier is an intention to categorize objects in the world to some discrete labeling scheme. For example, we may classify moving beings as dogs, cats, humans, etc.

At the same time, a classifier relates to meaning only while it correctly classifies. A misclassification is a mistake.

The meaning as classifiers paradigm has strong routes in Deb Roy's dissertation. In this, he related language to the sensory data of a robot.

It also has deep routes in Stephan Harnad's "Symbol grounding" paper. In this paper, Harnad petitions that the only way to give content to meaningless symbols (Searle) is to use a sub-symbolic system that relates domain knowledge to the symbol.

2.2 Grounded Language Models

I need to find when grounded language models first started becoming a thing.

In essence, language models see sentences as sequences of symbols and use probabilistic functions to relate some current state to the production the next word in the sequence.

Grounded language models follow in this line of thinking, but add in the capability to base the production on some multimodal signal.

Work on this paragraph more. GLMs have grown in popularity in recent years for several key reasons. First, there has been a significant increased in the amount of labeled data which relates language to vision. Second, the increase in data propelled techniques which can be best described as energy based models or factored representations to become substantially better.

Some of the early work on GLMs were from the Berg and Mitchell communities. In this, they used a CRF and other considerations to generate phrases and sentences for images.

2.3 Vagueness: Flexibility in Meaning

The previous computational methods treat variation in people's grounded language behavior primarily as noise. However, there is a class of words which can be used flexibly to denote different things in different contexts. These words—named vagueness by linguists—are used in clear, prototypcial cases and in borderline cases.

In borderline cases, the word is used in the context, but it may not extend outside of that context.

All vagueness is not created equal. Scalar vagueness is a measurable property whose use depends on the context and the prevailing common usage. For example, “tall” can refer to a skyscraper or a toddler.

Go over the different kinds of vagueness and how they can be interpretted in a computationally grounding framework.

Chapter 3

Grounded Color Semantics

3.1 Vagueness and Color

In this section, I discuss the properties of color. Specifically, I relate how it is measured and why it might be tricky to categorize it. This includes both cultural differences in color naming as well as perceptual variability.

3.2 A Bayesian Approach

In this section, I discuss most of the details that were in the TACL paper on this topic. Specifically, I talk about the Bayesian interpretation of the rational observer and the membership functions which lead to our formulation.

3.2.1 XKCD Dataset

3.2.2 Parameterized Boundaries

3.2.3 Markov-Chain-Monte-Carlo (MCMC) Training

3.2.4 Evaluations

3.3 A Neural Approach

In this section, I outline our neural approach. Specifically, how we extend the boundary interpretations of our original LUX model to be included in neural architecture. This begins from an encoding layer to map onto a latent layer. This latent layer is where the boundary distributions take place. The boundary distributions are then integrated into a final decision.

I also discuss how we extended this to include real images. This presented several new challenges. First, it required that we align the posterior of the image label distributions to those of our model. This required confronting a long-standing problem in natural language generation: optimizing a statistical model to produce unique descriptions rather than just reproducing the observed distribution in the data.

3.3.1 C3I Dataset

3.3.2 Parameterized Hypersurfaces

3.3.3 Training Objectives

3.3.4 Evaluations

Chapter 4

Grounded Language Models

In this chapter, I discuss the task of constructing sentences from semantic input.

4.1 Neural Tree Grammar

4.2 TBD

Chapter 5

Distinguishing Referents

5.1 Color Patches

5.2 Image Patches

References

Quine, M. (1960). Word and object.