Primer on Grounded Theory

Jeffrey Young

June 28, 2017

Executive Summary

What is Grounded Theory

Grounded Theory is a research methodology that was created by sociologists in the late 60s. It is qualitative, but still tries to be as systematic, rigorous and sound as quantitative methods.

How will it be used for the project?

We do not need to use the whole of grounded theory for our project because we are not in need of a theory per se. Rather we need to use grounded theory's coding ideas to generate a systematic, orthogonal database of codes that describe the content in algorithm explanations

Why is it of use to XOP project?

It is roots the project in a well used, well studied, research method and gives us a sound, systematic way to create the codes that are required to analyze algorithm explanations.

Other considerations

You may want to look over the "Reference that will be useful" section as this paper is well sources, and is a good branching point for other papers to check out. I've highlighted some interesting ones.

Context

This document is for the XOP Encoding project, participants include Eric Walkingshaw and Jeffrey Young. This is the fourth such primer in the project, the first one is a primer on DN-Theory, the second a primer on typologies of explanation, specifically Bellack et al's typology, the third on Dan Hillman's PhD thesis. This document is meant to provide a decent summary for understanding grounded theory. We are interested in grounded theory because it offers a rigorous, scientific method for determining sets of tags or "codes" from varied documents. In terms of our work, we believe that a comprehensive set of codes, or rather, an ontology of algorithm-explanation descriptors will serve as the basis for an XOP-based DSL.

Orientation

This document is meant to explicate three things:

1. What is Grounded Theory?
2. Where did Grounded Theory come from?
3. How does one *do* grounded theory?
4. What will grounded theory applied to the XOP project look like?

Grounded Theory Terminology

memos are defined as a specialized type of written record, that contains the products of analyses. These are basically thoughts a researcher wants to record during an analysis phase. Memos are used for:

1. Open data exploration
2. Identifying/developing the properties and dimensions concepts/categories
3. Elaborating the paradigm: the relationships between conditions, actions/interactions, and consequences
4. Developing a story line

Theoretical Sampling is a type of sampling that disregards typical statistical considerations and instead focuses on filling in gaps in the data based on a theory. For example, one might have data of apple growth that has only been sampled during the day. A grounded theorist might say "well we have no data on growth during the nighttime, so let's go collect that". This is different than normal statistical sampling, which is just focused on getting a *representative* sample of some population.

Open Coding is the first stage of coding, in this stage a researcher writes down terms, or tags that describe the data. The point is to develop a sense of the salient categories that occur in the data.

Axial Coding is the second stage of coding, in this stage the researcher is trying to identify relationships between the tags that were developed in the Open Coding phase. The goal of this phase is to develop a *coding paradigm* which is a theoretical model that visually displays the inter relationships between the codes

Selective Coding is the last stage of coding, in this stage the researcher tries to identify one, or two central categories upon which forms the basic and central phenomena for their theory. Then the researcher tries to systematically relate the core categories to the other categories or tags.

Saturation refers to the end point in coding. This occurs when a grounded theorist adds new data (per Theoretic sampling) and no new codes are identified in it. Some grounded theorists refer to this as being as much a feeling as an identifiable end point.

What is Grounded Theory

Grounded theory is a methodology for research that was created by sociological researches in the mid 60s. The main idea is to generate, or discover theory *based on* data. In this sense, the theory is *grounded* in the data. In the original authors own words: "A grounded theory is one that is inductively derived from the study of the phenomena it represents."[1]

Where did Grounded Theory Come From?

Grounded theory was developed by Strauss and Glaser working on sociological health research in the 60s. It is qualitative, and is an attempt to show that qualitative research can have a rigorous, sound, and useful methodology just as quantitative research has.

What exactly is the method of Grounded Theory?

The grounded theory method is robust and nuanced. From the sources I've read I believe it is roughly as follows:

1. Pick a problem to research
2. Identify sources of data related to 1

3. Collect that data, whether it be interviews, public records, quantitative data, personal letters
4. Open Code that data to develop tags, and some over-arching categories
5. Axially Code the tags and categories to identify and develop relationships in the tags and categories
6. Selectively code the tags and categories to identify the central phenomena in the data.
7. From Selective coding, the researcher should have developed a *theory* that is *grounded* in the data.

Central to grounded theory is that these steps are not linear, rather they happen simultaneously. The theorist is always moving back and forth between these steps in order to develop a robust theory.

Central Tenets of Coding

Open Coding

See the definition above. In open coding the researcher should be asking several questions as these code. These include "what is the core point of this", "What would happen if x became y?", "how is this instance of this code similar to other data that is coded similarly?". The latter is called constant comparison.

Constant Comparison

Constant Comparison is a technique that grounded researches use to make sure that they are coding validly throughout the data. The central idea is that when you code some data you think back to other times you've coded that data and ask yourself if you are being consistent. If so, then good, if not then you should revise the code or category in some way.

Saturation

One is done coding when introducing more data does not generate any new tags. This is called *saturation* by grounded theorists.

How will we apply grounded theory to our project?

The product of grounded theory is a theory that is grounded in the data. For our use case we don't need a theory per se, rather we need a robust set of identifiable, orthogonal codes that we can use to describe the information in algorithm explanations. Using these codes we can construct our own Typology of Explanation, or ontology of explanation (see previous primer's on Typologies). Once we have a rigorous, system for describing the content of algorithm explanations, and the way an explanation matures over the course of a document, then we have the bones or foundation for a XOP DSL.

Sources

This primer was drawn from the following sources:

1. A youtube lecture series by Graham R. Gibbs: `https://www.youtube.com/watch?v=4SZDTp3_New`
2. See references below: [2]

References

[1] J. Corbin, A. Strauss, and A.L. Strauss. *Basics of Qualitative Research*. SAGE Publications, 2014.

[2] K. Charmaz. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Introducing Qualitative Methods series. SAGE Publications, 2006.