Qualitative Parameter Triangulation: A Formulated Approach to Parameterize Multimodal Models

Yeyu Wang[1], Andrew R. Ruis[1] and David Williamson Shaffer[1]

[1] *University of Wisconsin – Madison, 1025 W Johnson St, Madison, WI, USA, 53703.*

Abstract
Data fusion and parameterization based on qualitative insights are two key challenges in multimodal learning analytics. In this study, we propose Qualitative Parameter Triangulation (QPT) to address these two challenges. In particular, QPT generate optimized parameter values for the type of multimodal learning models that are event-based, process-oriented and connection-structured concerning with recent temporality.

Keywords
Quantitative Ethnography, Methodology, Model Elicitation.

1. Introduction

1.1. Multimodalities

Humans interact and communicate through variant modes. According to [1], mode refers to channels of representations which are socially and culturally shared. These socially shaped representations serve as different functionalities in the communication processes [2]. In Kress's book, he argues that "in communication several modes are always used together, in modal essembles, designed so that each mode has a specific task and function (p.28). " Thus, one of the key questions in multimodality studies is the modal affordance, which characterizes the "reach" of one mode influencing others.

However, multimodality is not a simple sum of various modes. Instead, multimodality studies the relationship between different modes [3]. According to [4], multimodality is defined as "an inter-disciplinary approach drawn from social semiotics that understands communication and representation as more than language and attends systematically to the social interpretation of a range of forms of making meaning. (p.250) " This definition creates link between communication and learning. Social semiotics, defined by [2], is a product of knowledge construction. [2] conceptualize knowledge as a problem-solving tool, which is created by multimodal representations based on their modal affordances. However, multimodality is not a simple sum of modal affordances. Instead, multimodality studies the relationship between different modes [4]. Thus, multimodal learning analytics study cross-modal interactions during learning processes.

To study interactional processes of learning, [5] emphasizes the importance of process-oriented approach. He conducted multimodal learning analytics in the context of an engineering challenge and compared non-process-oriented and process-oriented approaches. The process-oriented approach provides critical insights about the characteristics of learners, as "planner" and "thinker", which was not manifested in the non-process-oriented approach. Specifically, he claims that temporality and sequences in learning activities are essential in the process-oriented multimodal learning models. Similarly, argued by [2], multimodal communication and interactions are provisional and temporality-critical. That is, communication theories assumes that humans evaluate social scenarios and shapes their communicational encounters within a recent temporal frame. Thus, studying process of multimodal learning is to investigate relationships across modes and connections between events that are temporally organized.

1.2. Challenges in Multimodal Models

To represent an *event-based*, *process-oriented* and *connection-structured* multimodal learning process, there are two major challenges.

First, how to fuse multimodal data with varying time scales and frequencies. Depending on utilities and assumptions involved, there are three categories of data fusion: naive fusion, low-level fusion and high-level fusion [6]. Naive fusion is commonly used in exploratory studies, in which data is aggregated into features without specific assumptions. With prior knowledge and assumptions of data, researchers construct features in a small-time scale which describes the relationship across events. High-level fusion requires more assumptions and theoretical foundations about turning data into meanings. According to [7], data fusion face challenges exist in both collection and modeling. Noncommensurability and incompatible size of data are issues for data collected from different instruments and devices. Noncommensurability refers to the issue that the raw format of data does not commute. For example, data collected from electrodermal activity (EDA) is not directly related to eye movements in a study of mind wondering, which requires the first step of transformation [8]. Also, due to different observational mode, the size of data samples varies, which may result in large uncertainty and bias in modeling.

Second, how to elicit a quantitative parameter based on qualitative understanding. In the operationalization step, that is how to transform the qualitative information to quantitative parameters for further modeling. This challenge is not uncommon in a mixed-method study and solutions exist in unimodal analysis. For example, quantitative ethnography [9] defines the mechanical grip between observations and interpretations as *codes*, which transforms the qualitative records to binary numbers. Additionally, the operationalization of common ground in a discourse is defined as a window function [10], which assumes that codes within the recent temporal context are connected to each other. Here, the size of the window is an operationalization of recent temporal context. Within the window, the connections between codes is in presence; otherwise, there is no connections. However, these methods are easier to be applied to a unimodal dataset. With multimodality, the complexity of eliciting a parameter increases due to interactions and the irregularity of various modes. For example, it becomes a challenge for qualitative researchers or domain experts to elicit relationship between modes: how many times an eye-gaze event has a longer impact compared to a event of log data?

2

1.3. A Solution: Qualitative Parameter Triangulation (QPT)

We propose a formulated approach called *Qualitative Parameter Triangulation* (QPT) to address the two challenges above.

First, modeling involving QPT doesn't require data fusion; instead, QPT helps with determining parameter values in a pre-defined function. That is, the dataset can preserve its raw representation as long as meeting the requirement of *evidentiary completeness, ontological consistency* and *terminlogical consistency* [9]. For example, data collected from different streams can be organized into data spreadsheet: rows contains all kinds of information while each column contains one type of information. Instead of generate aggregated features as traditional data fusion, QPT facilitates parameterization by describing relationships between modes and events as mathematical functions. As mentioned above, the temporal impact between two events can be operationalized as a window, which is a step function in its mathematical form. Based on a theory in communication sciences that each mode serves a different functionality, we can vary the mathematical function to describe relationship between events for various modes. With the simplest example, we can vary the length of window to describe the survival of impact for different modes.

Second, QPT provides a formulated structure to help researchers to elicit their hypothesis on the qualtiative data. Instead of asking directly about the relationship between two modes, QPT automatedly constructs networks based on qualitative researcher's narratives and optimizes parameter values for the next-step modeling.
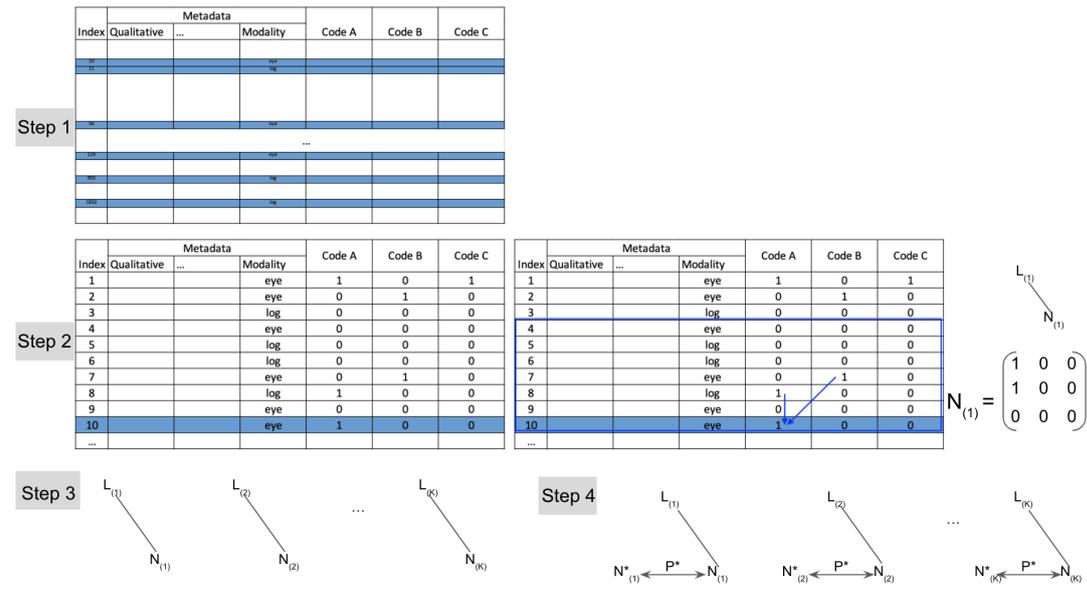
1.4. Usage and Combination with Other Models

QPT can be combined with any models that is event-based, process-oriented and connection-structured with consideration of temporalities, such as lag sequential pattern mining, process mining, etc. In this paper, we use Epistemic Network Analysis (ENA) as an example to demonstrate how QPT can be used to determine parameters: the window size of different modes. We select ENA due to its affordances of modeling interactivity and interdependence in problem-solving processes [11], which is event-based, process-oriented, and connection-structured temporally. which is aligned with the context of test dataset. The dataset of demonstration is collected from puzzle-solving game, called *baba is you*. See the next section for more details. However, as a conceptual and methodological framework, QPT can be combined with other methods, such as lag sequential mining, transition status analysis, etc. In this paper, we will use ENA as an example.

2. QPT Approach

2.1. Overview

QPT helps with eliciting assumptions based on qualitative understanding and optimizes the parameters using automated approach for further modeling. For example, if the goal is to determine the length of active-impact window for different modes, the inputs include: (1) human's qualitative interpretation of connections made, given randomly sampled time points, and (2) number of parameters. Then, QPT outputs the optimized window size for each mode that can

Step 1

Index	Metadata			Code A	Code B	Code C
	Qualitative	...	Modality			

Step 2

Index	Metadata			Code A	Code B	Code C
	Qualitative	...	Modality			
1			eye	1	0	1
2			eye	0	1	0
3			log	0	0	0
4			eye	0	0	0
5			log	0	0	0
6			log	0	0	0
7			eye	0	1	0
8			log	1	0	0
9			eye	0	0	0
10			eye	1	0	0
...						

Index	Metadata			Code A	Code B	Code C
	Qualitative	...	Modality			
1			eye	1	0	1
2			eye	0	1	0
3			log	0	0	0
4			eye	0	0	0
5			log	0	0	0
6			log	0	0	0
7			eye	0	1	0
8			log	1	0	0
9			eye	0	0	0
10			eye	1	0	0
...						

$$N_{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Step 3 $L_{(1)} \searrow N_{(1)}$ $\quad L_{(2)} \searrow N_{(2)}$ $\quad \cdots \quad L_{(K)} \searrow N_{(K)}$

Step 4 $L_{(1)} \searrow N^*_{(1)} \xleftarrow{P^*} N_{(1)}$ $\quad L_{(2)} \searrow N^*_{(2)} \xleftarrow{P^*} N_{(2)}$ $\quad \cdots \quad L_{(K)} \searrow N^*_{(K)} \xleftarrow{P^*} N_{(K)}$

be used as the window parameters in ENA. We refers to three key concepts in this triangulation: qualitative story, network representation (connections) and parameter determination. By optimizing the parameters, QPT minimize the differences between qualitative story and quantitative connections.

2.2. Worked Example: Determining Window Sizes for Two Modes Using QPT

Start with a multimodal dataset with evidentiary completeness, ontological consistency and terminlogical consistency. Define parameters needed to describe impacts of different modes. In this worked example, we defines the two parameters: eye-gazing and log data.

Step 1: Randomly select K lines from the whole dataset.

Step 2: For each line as a referring line, let the qualitative researcher tell a story about the learning event. For example, in the context of a digital learning game, line 10 is an eye-gazing event which captures the player looking at a specific object. The qualitative researcher elaborates on their understanding about why the player looked at such an object. Then, based on the content included in researcher's narrative, we can determine the connections between codes in a network representation. In the example, there is a connection from code A to itself and a connection from code B to code A. Thus, the entry (1,1) and (2,1) in the adjacency matrix are marked as 1; for the rest of entries, mark as 0.

Step 3: Repeat Step 2 for K time. The coded network structures serve as K ground-truth labels.

Step 4: Use an automated optimization algorithm to determine the parameters for eye-gazing and log-data, which result in the least differences between the ground-truth labels ($N_{(k)}$) and estimated network when plugging in parameters ($N^*_{(k)}$) into ENA. For example, let P^*_{eye} be the window of impact parameter for eye-gaze mode. If P^*_{eye} is 10, the model assumes that one eye-gaze event will have *approximately* active impact on all events happening in the next 10

seconds. Similarly, we need to optimize P_{log}^* to derive a complex model.

After QPT provides optimized parameters, researchers close the interpretive loop by checking whether the parameters are aligned with their original understanding. With validated parameters and interpretive alignment, use the parameter values to construct an multimodal ENA model.

2.3. Mathematical Notations

2.3.1. Human Labels of Connections

We randomly select K lines from the dataset regardless of modes. Let a randomly selected line be the referring line $L_{(k)}$, $k \in \{1, 2, 3, ..., K\}$. Each $L_{(k)}$ has a cooresponding adjacency matrix to represent what connections were made between any two codes, determined by *qualitative* understanding by researchers or domain experts. Let $N_{(k)}$ be the matrix, which represents the presence of connections between two codes. Let $N_{(k)}^{ij}$ be the binary value to indicate the presence of connections between code i and code j, given the referring line $L_{(k)}$. If $N_{(k)}^{ij} = 1$, there is a connection between code i and code j; otherwise, no connection between code i and code j.

2.3.2. Deriving Parameters for Different Modes

A multimodal dataset may include M modes. For each mode m, we will determine one parameter (P_m) as the length of a temporal window, which describes how long one event from such a mode have impacts on other events. Let $\mathbb{P}$ be a vector to record all parameters for multimodalities.

QPT optimizes the vector of parameters $\mathbb{P}^*$, which represents the active impact windows for different modes with least error. With any model $\mathbb{M}$, which describes interdependence between two events using connections, we can plug in $\mathbb{P}^*$ into function $f_{\mathbb{M}}$. That is, $N_{(k)}^* = f_{\mathbb{M}}(L_{(k)}^*, \mathbb{P}^*)$. To derive $\mathbb{P}^*$, let $N_{(k)}$ ($k \in \{1, 2, 3, ..., K\}$) be the ground truth. Define $\mathbb{L}(N_{(k)}, N_{(k)}^*)$ as the loss function which describes the sum of difference between $N_{(k)}$ and $N_{(k)}^*$. We optimize $\mathbb{P}^*$ by:

$$\operatorname*{argmin}_{\mathbb{P}^*} \sum_k \mathbb{L}(N_{(k)}, f_{\mathbb{M}}(L_{(k)}, \mathbb{P}^*))$$

Start with random value $\mathbb{P}_0$ and use gradient descent algorithm to converge to the local minimum of $\mathbb{L}$. Usually, we use gradient descent algorithm to derive the minimum local mean.

3. Discussion

In this paper, we propose a method called Qualitative Parameter Triangulation to determine the parameter values in a multimodal learning model. Specifically, this approach address the challenges of data fusion and parameterization based on qualitative insights. Additionally, the key concept of engaging qualitative researchers in the loop ensures the interpretive alignment, which provides potentials for closing feedback loop with other stakeholders in the multimodal study. Future work are as follows: (1) Use empirical data to test the efficacy of QPT. (2) Try

different models besides ENA. (3) Try different methods of optimization (such as using Gibbs sampling to estimate parameters of different modes iteratively). (4) Create multimodal interface to facilitate assumption elicitation for research efficiency and closing-loops in human-computer interactions.

References

[1] T. Van Leeuwen, G. Kress, Discourse semiotics, Discourse studies: A multidisciplinary introduction 2 (2011) 107–125.

[2] G. Kress, Multimodality: A social semiotic approach to contemporary communication, Routledge, 2009.

[3] C. Jewitt, Multimodal methods for researching digital technologies, The SAGE handbook of digital technology research (2013) 250–265.

[4] C. Jewitt, Multimodal analysis, in: The Routledge handbook of language and digital communication, Routledge, 2015, pp. 83–98.

[5] P. Blikstein, Multimodal learning analytics, in: Proceedings of the third international conference on learning analytics and knowledge, 2013, pp. 102–106.

[6] M. Worsley, Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors, in: Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge, 2014, pp. 1–4.

[7] D. Lahat, T. Adali, C. Jutten, Multimodal data fusion: an overview of methods, challenges, and prospects, Proceedings of the IEEE 103 (2015) 1449–1477.

[8] I. Brishtel, A. A. Khan, T. Schmidt, T. Dingler, S. Ishimaru, A. Dengel, Mind wandering in a multimodal reading setting: Behavior analysis & automatic detection using eye-tracking and an eda sensor, Sensors 20 (2020) 2546.

[9] D. W. Shaffer, Quantitative ethnography, Lulu. com, 2017.

[10] A. Ruis, A. Siebert-Evenstone, R. Pozen, B. Eagan, D. W. Shaffer, Finding common ground: A method for measuring recent temporal context in analyses of complex, collaborative thinking (2019).

[11] Z. Swiecki, Measuring the impact of interdependence on individuals during collaborative problem-solving., Journal of Learning Analytics 8 (2021) 75–94.