Critical Abstraction: Generating Simplest Models for Causal Explanation

Brian C. Williams
Xerox PARC

Goal: Transition from being users to being producers of qualitative abstractions of how devices work.

Claims:

1. What is interesting is captured in part through **Critical Abstraction**

2. What is qualitatively and temporally interesting is inextricably tied to concepts of causality and interaction.

Focus: Model generation in support of causal explanation.

Issues

1. What is Interesting?
   $\Rightarrow$ **Critical Abstraction**

2. What Representation is adequate?
   - Concise histories
   - SR1 hybrid algebra

3. How do we generate what's interesting?
   - TCP
   - **Critical Abstraction Engine**

Features of Interest

1. What is temporally interesting?

2. What features of quantities are interesting?

3. What features of interactions are interesting?

Features of Interest

1. What is temporally interesting?

2. **What features of quantities are interesting?**
   ⇒ What are the interesting landmarks?

3. **What features of interactions are interesting?**
   ⇒ What are the simplest models?

Background

1. What is temporally interesting?
   - Events — changes in values and interactions
   - Orderings that affect sequence of events.

2. What representation is adequate?
   (concise histories)

3. How do we generate it?
   (TCP)

Lesson: temporal features fall out of local, causal interactions