Interactive Exploration of Developer Interaction Traces using a HMM

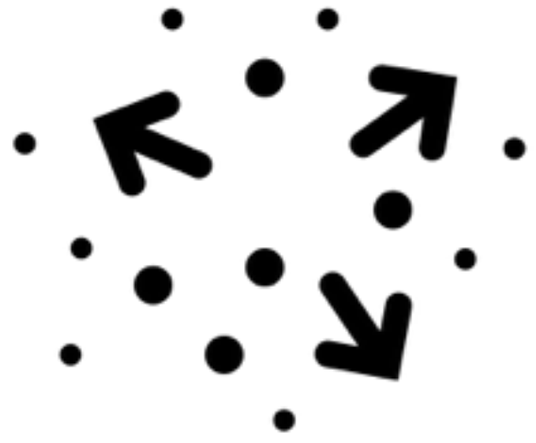
Kosta Damevski (Virginia Commonwealth University)
Hui Chen (Virginia State University)
David Shepherd (ABB Inc)
Lori Pollock (University of Delaware)

MSR 2016
Austin, TX

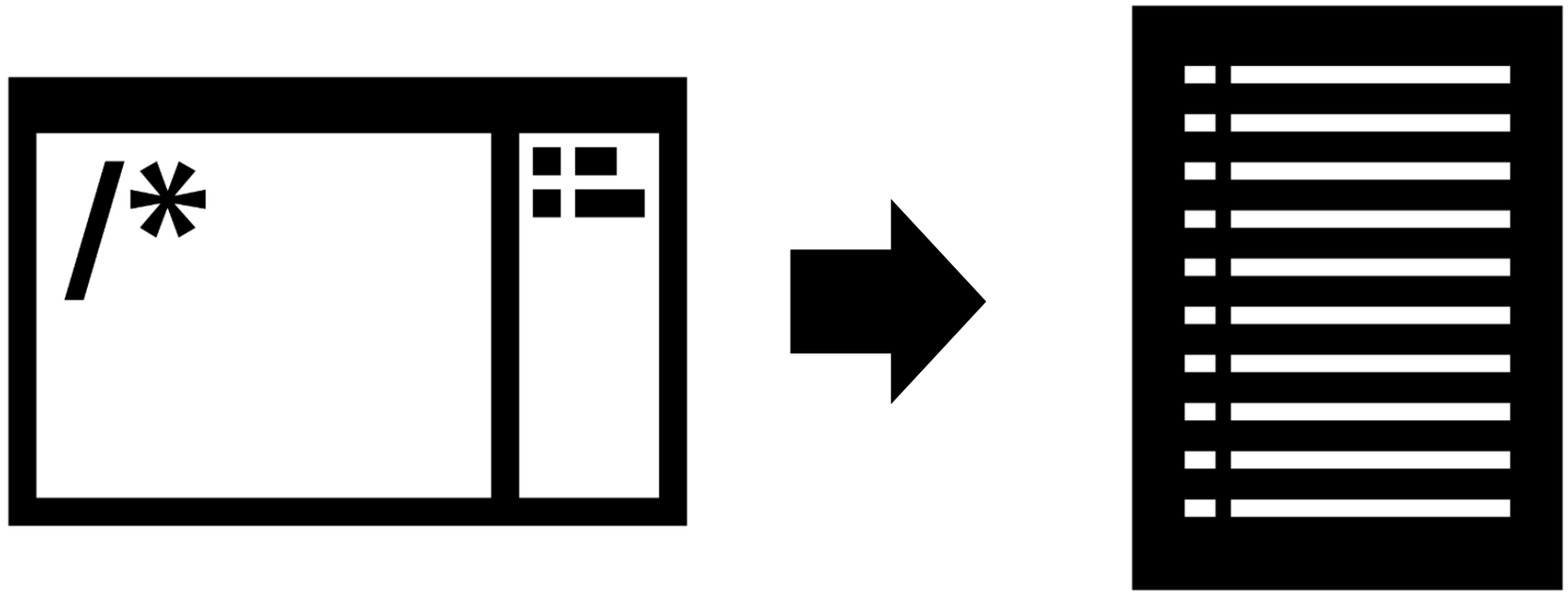
School of Engineering | Computer Science

Developer Interaction Traces

- Primarily collected from developer clicks and key presses in the IDE
    - Provide a perspective of ***development in the field***
- BUT
    - Noisy
    - Large
    - Low-level

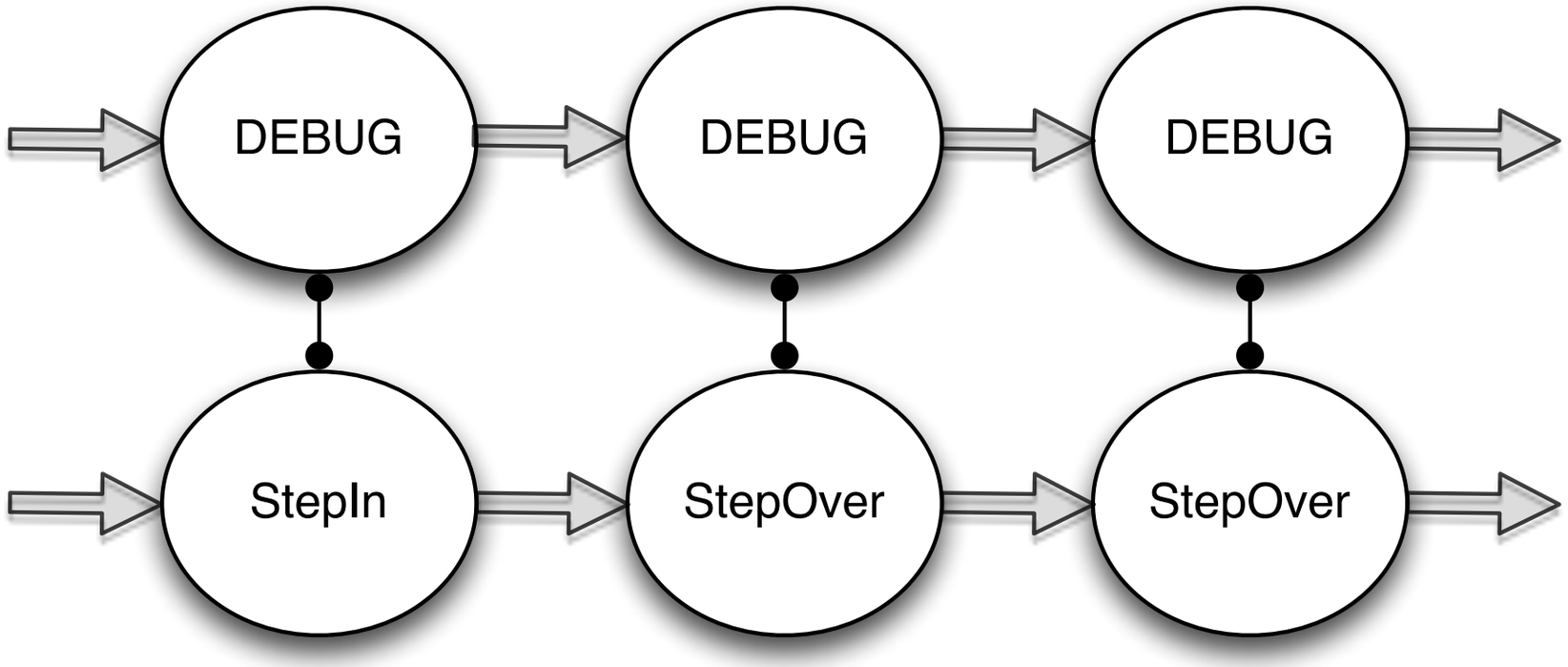


Typical Interaction Trace Analysis



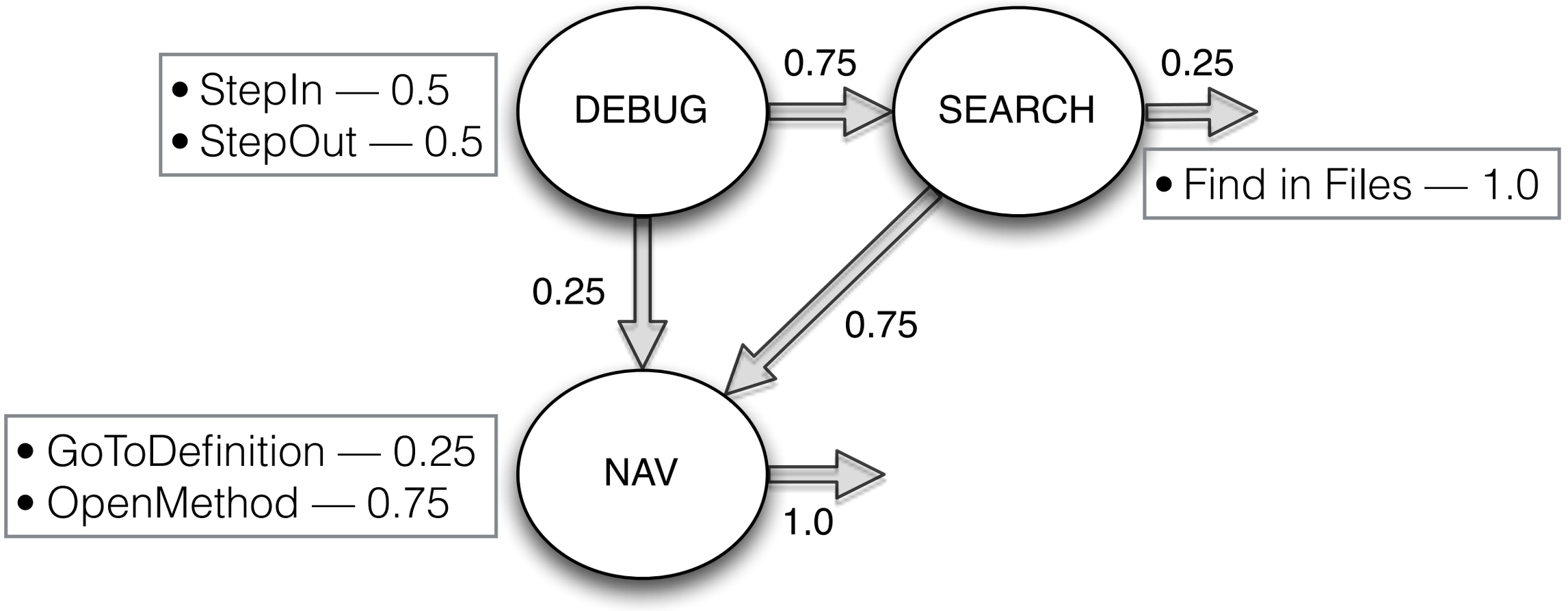
- A time ordered stream of a single developer's interactions and events
- Most analysis approaches extract low-level behaviors

Hidden Markov Models



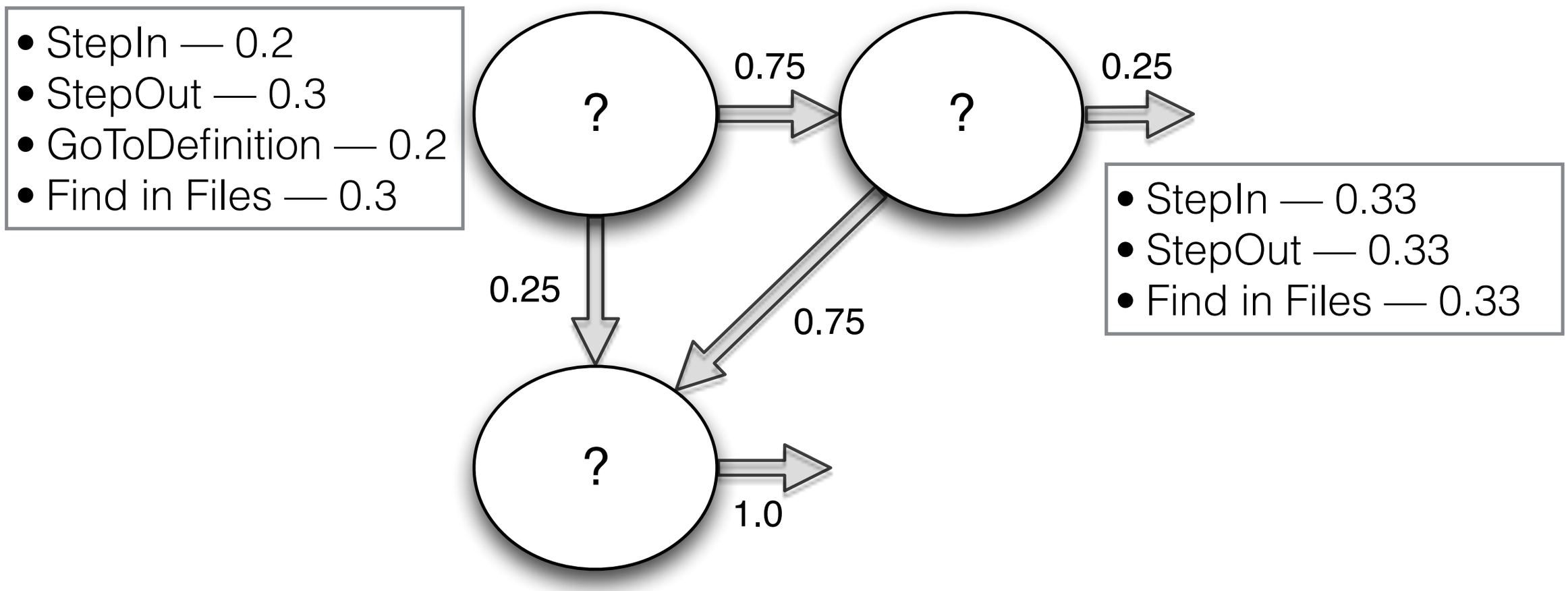
- dual stochastic processes
- lower process represents the individual interactions, higher represents behaviors

Baum-Welch Learning from Interaction Data



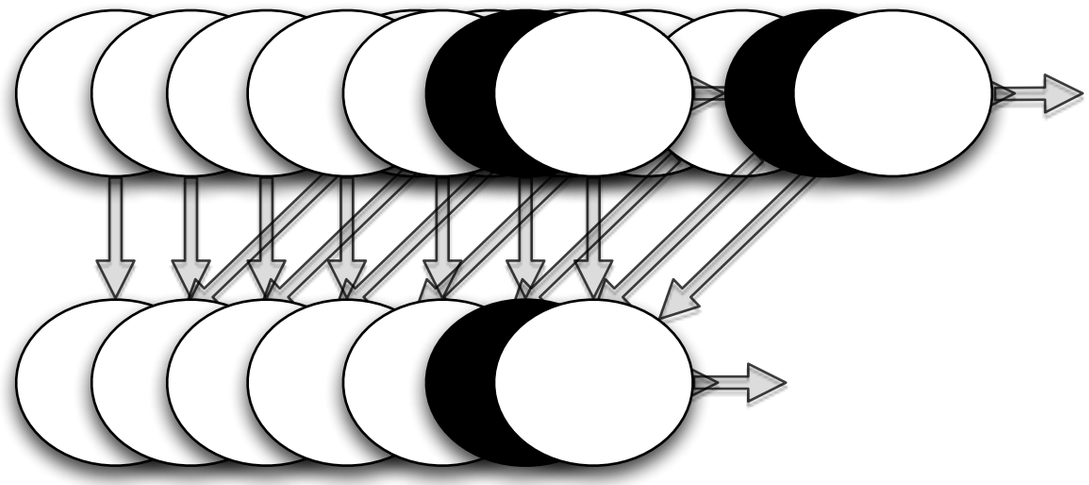
- Uses maximum likelihood estimation to infer a HMM, given some sequential data

Problem with Baum-W.



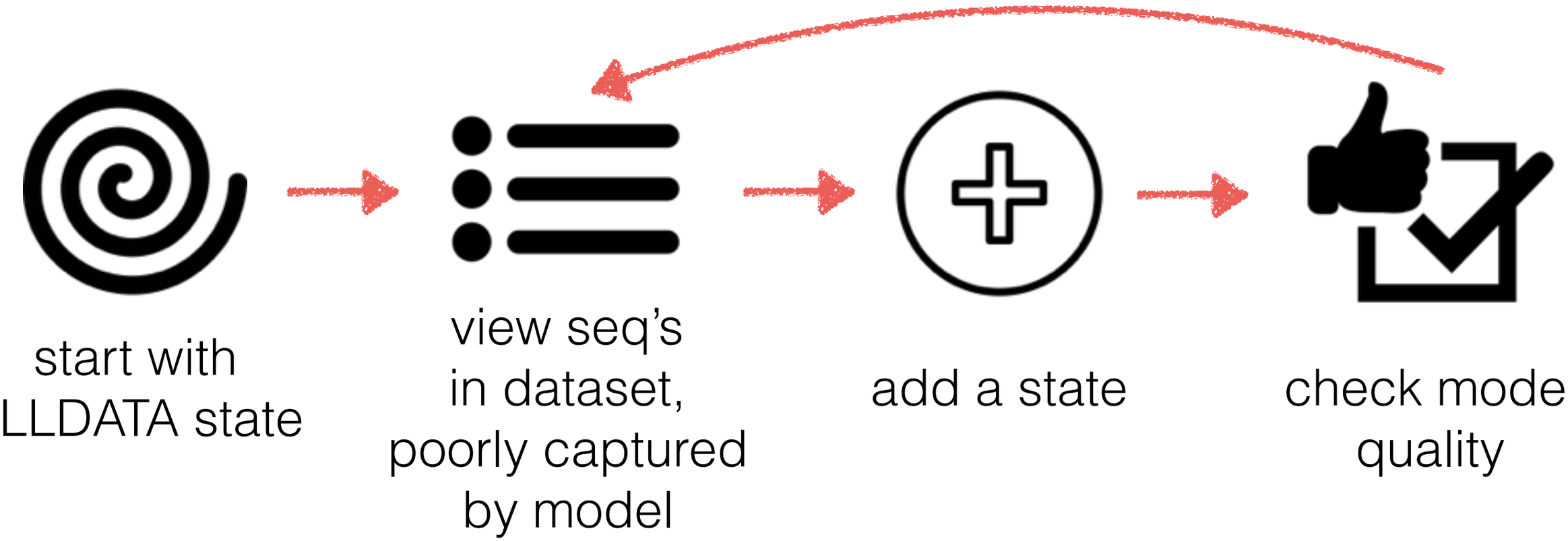
- In reality, most inferred models are ***not interpretable***

Key Idea: B.Y.O. HMM



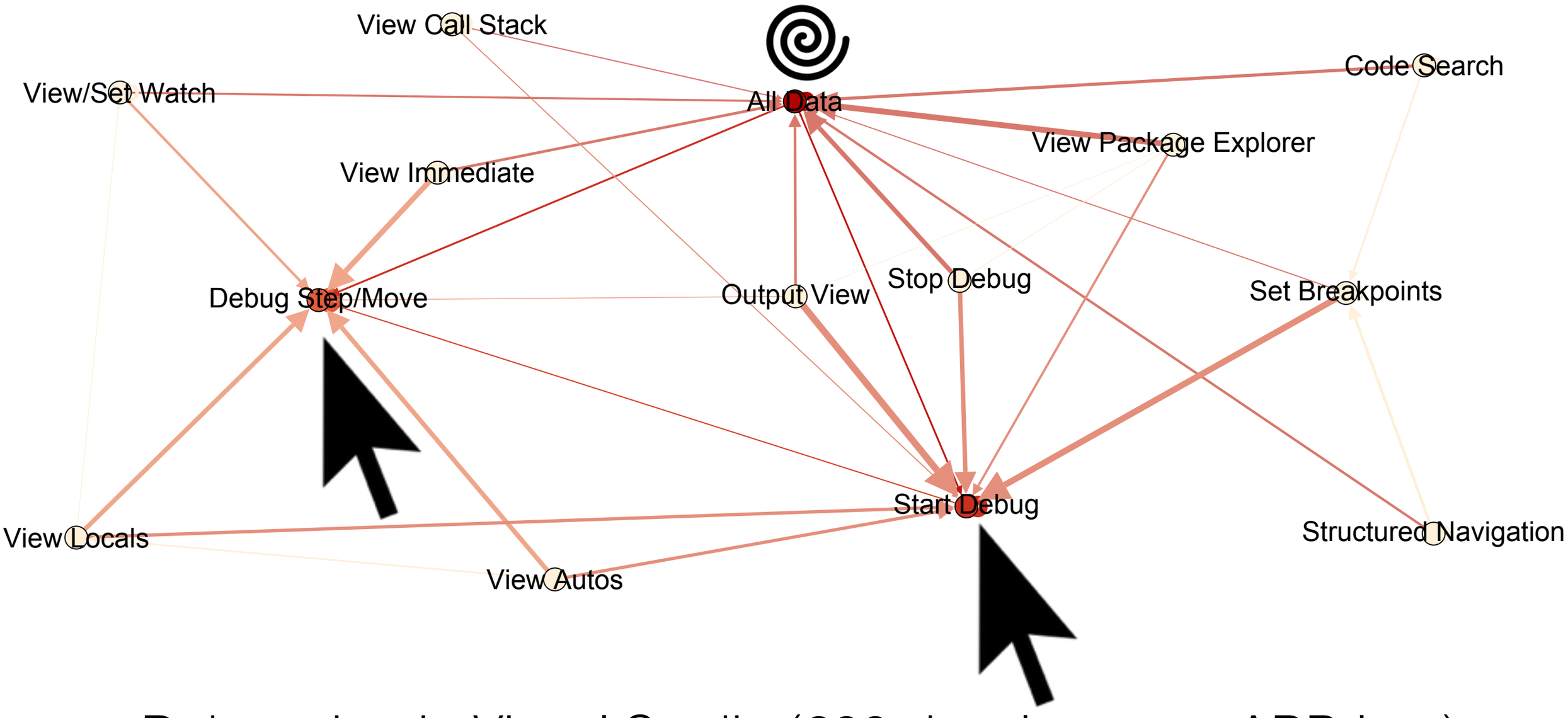
- There are many HMMs that fit the interaction data well
    - many similar local minima for the Baum-Welch algorithm to find
- Optimize understandability by building model ***interactively***, state by state

Interactive HMM Workflow



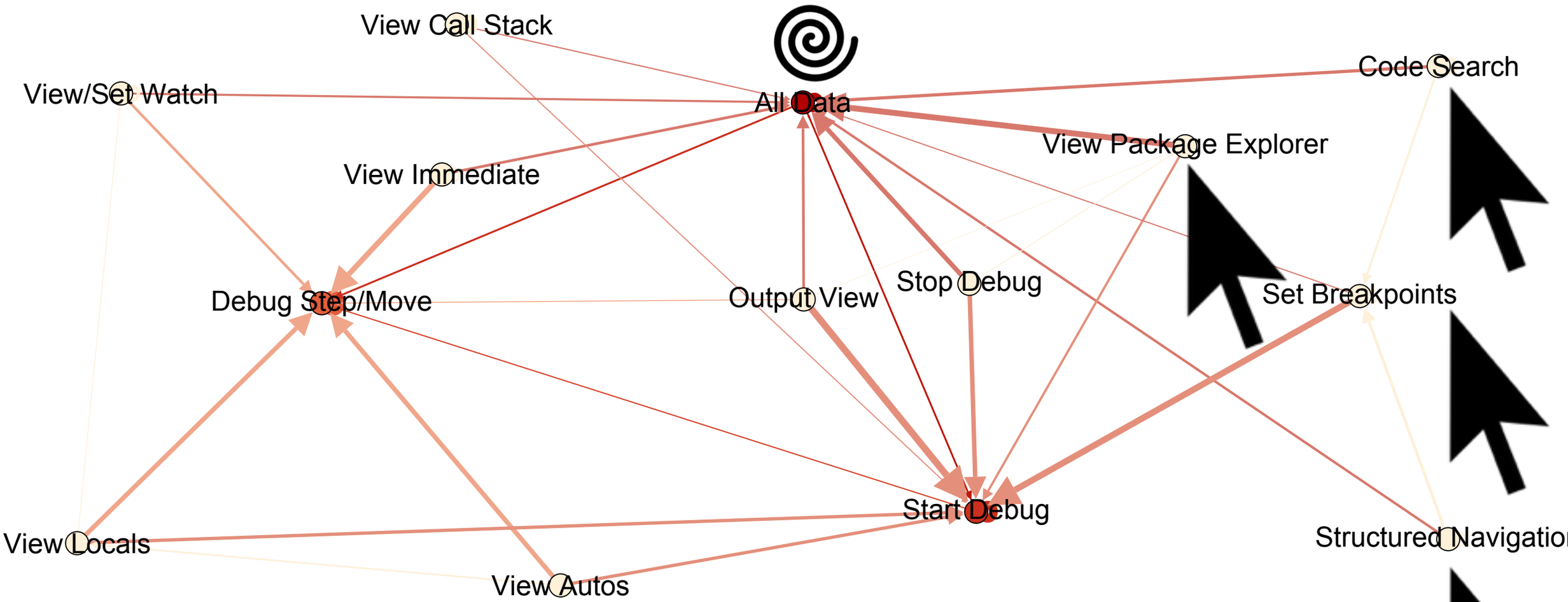
- Iteratively build up a quality, interpretable model, with appropriate number of states

HMM Workflow in Action



- Debugging in Visual Studio (200 developers at ABB Inc.)

Results on Debugging



- Code ***search and navigation lead to setting breakpoints***, while opening files does not

Contributions

- A way to ***interactively*** build ***interpretable*** HMMs from developer interaction data
    - Focusing on high-level behaviors of interest
- Feedback via model quality measurement and sequences in data poorly captured by mode

Thanks!

QUESTIONS?

damevski@acm.org