Bayesian Visual Analytics (BaVA): A Formal Visual Updating Procedure

Scotland C. Leman\textsuperscript{1}, Leanna House\textsuperscript{1}, Chris North\textsuperscript{2,3}

\textsuperscript{1}Department of Statistics
\textsuperscript{2}Department of Computer Science
\textsuperscript{3}Center for Human Computer Interaction

Virginia Tech
leman@vt.edu

FODAVA
Dec 2009
Foraging + Sensemaking

Extraction

Synthesis
Deep Interaction

- User guided modeling, via direct manipulation
- Analyst injects domain knowledge into the visualization
- The visualization responds meaningfully
Mapping Visual to Model

- Natural interaction within the visual metaphor
- Level of detail of interactive input control
- Visual feedback of underlying model updates
A New Framework...

- Beyond the Visualization Pipeline
Outline

1. The BaVA process of incorporating visualizable feedback.
2. A Cartoon Illustration of the BaVA process
3. Formalities
4. Insights
The BaVA Process

The BaVA paradigm is a hierarchical process, entailing 5 layers. Some of these layers will rely on stochastic models, while others might utilize deterministic transformations.

Figure 1: Schematic illustration of the BaVA process.
The BaVA Process

Preliminary: Obtain data $D$, which will in general be high dimensional, and massive in size. Like most analyses, our objective is to find hidden structure, and associations within the data.

Figure 2: Schematic illustration of the BaVA process.
Step 1: From a parameterized mathematical model, $\pi(D|\theta)$, which relates relevant unknowns ($\theta$) to the data $D$, obtain inferences about $\theta$ using Bayes’ Theorem: $\pi(\theta|D) = \pi(D|\theta)\pi(\theta)/\int \pi(D|\theta)\pi(\theta)d\theta$.

Figure 3: Schematic illustration of the BaVA process.
The BaVA Process

Step 2: Visualize both data ($\mathbf{D}$) and inferences ($\theta$) (perhaps only a summary: $\hat{\theta}$) in a meaningful way.

Figure 4: Schematic illustration of the BaVA process.
The BaVA Process

Step 3: Allow the user to reconfigure the visualization, in order to express relationships between a small set of objects. We refer to this as cognitive feedback.

Figure 5: Schematic illustration of the BaVA process.
The BaVA Process

Step 4: The information contained in the cognitive feedback is translated (parameterized) back into model parameters ($\mathbf{\theta}$). This is referred to as \textit{parametric feedback}, and usually invokes a black-box layer, which does not involve the user.

Figure 6: Schematic illustration of the BaVA process.
The BaVA Process

Step 5: The parameterized feedback is injected into the system, to provide updated parameter assessments, and subsequently updated visualizations. This relies on the a new iteration of Bayes’ Theorem:

\[
\pi(\theta|f, v, D) = \frac{\pi(f|v, \theta)\pi(\theta|D)}{\int \pi(f|v, \theta)\pi(\theta|D)d\theta}, \text{ where } f = \{f_p, f_c\}
\]

Figure 7: Schematic illustration of the BaVA process.
Step 1: Perform Bayesian analysis to obtain model inferences. For this example, we will focus on projection based clustering (PCA), but will rely on its model based variant Probabilistic-PCA (PPCA) (Tipping and Bishop, 1999).
Step 1: Perform Bayesian analysis.
Step 2: Visualize the inferences. (Project through principal subspace)

Figure 8: RAW PCA projection
Step 2: Visualize the inferences. (Project through principal subspace)

Figure 9: RAW PCA projection

This is really pretty uninformative, but is not unlikely to occur in real applications.
Step 2: Visualize the inferences.
Step 3: Identify anomalous structure in the visual display.

Figure 10: Identify odd structure
Step 2: Visualize the inferences.
Step 3: Identify anomalous structure in the visual display.

Perhaps these points should not appear so close together.

Figure 11: Identify odd structure
Step 3: Identify anomalous structure in the visual display.
Supply feedback by adjusting the display.

Figure 12: Supply Feedback (Cognitive)
Step 3: Supply feedback by adjusting the display

Given the visualization, the user is defining a random variable, governed by a cognitive distribution: $\pi(f_c|v)$. 

Figure 13: Supply Feedback (Cognitive)
Step 3: Supply feedback by adjusting the display (Cognitive Feedback)
Step 4: Parameterize the feedback

Express a mathematical relationship, between feedback, and parameters: $g(f_p|f_c, \theta)$. 

Figure 14: Supply Feedback (Cognitive)
Step 4: Parameterize the feedback
Step 5: Update the display.

Figure 15: Sequentially update
Formalities

Core BaVA steps:

1. Posterior inferences: \( \pi(\theta | D) = \pi(D | \theta)\pi(\theta) / \int \pi(D | \theta)\pi(\theta) d\theta \).
2. Visualize: \( g(v | \hat{\theta}, D) \), or \( g(v | D) = \int g(v | \theta, D) d\theta \).
3. Supply Cognitive Feedback: \( \pi(f_c | v) \).
4. Parameterize Feedback: \( g(f_p | f_c, v) \).
5. Sequentially update using Bayes’ Sequential updating formula:

\[
\pi(\theta | f, v, D) = \pi(f | v, \theta)\pi(\theta | D) / \int \pi(f | v, \theta)\pi(\theta | D) d\theta
\]

\( \propto \pi(f | v, \theta)\pi(\theta | D) \)

\( = g(f_p | f_c, \theta)\pi(f_c | v)\pi(\theta | D), \)

where \( f = \{f_p, f_c\} \)
Figure 16: a cartoon illustration of the BaVA process
The mathematics suggests an uninteresting projection.

Figure 17: An uninteresting projection
Insights

A user keys into interesting structure.

Figure 18: An uninteresting projection
A user keys into interesting structure and provides feedback.

Figure 19: User provides cognitive feedback
This feedback is parameterized, using information in both the low and high dimensional feature spaces.

\[ \Delta^t_R = [\Delta_{R_x}, \Delta_{R_y}, \Delta_{R_z}] \], is the vector of residues in the high dimensional feature space.
\[ f_p = \frac{1}{\sqrt{2\pi}} \int \exp \left( -\frac{1}{2} f_c \frac{1}{\Delta_R \Delta_t^f} \right) \text{Inv-Wish}(f_p | \Sigma) \]
The new principal directions corresponding to the parameterized feedback matrix $f_p = (\Delta_R \Delta^t_R)[-f_c]$ are displayed.
The new principal directions corresponding to the parameterized feedback matrix
\[ f_P = (\Delta_R \Delta_R^t)^{-f_c} \] are displayed.
Insights

The system is sequentially updated

Figure 21: Feedback is parameterized
The system is sequentially updated

Figure 22: Feedback is parameterized

$$\pi(\theta|f, v, D) = \frac{\pi(f|v, \theta)\pi(D)}{\int \pi(f|v, \theta)\pi(D)d\theta} \propto \pi(f|v, \theta)\pi(D)$$

$$= g(f_p|f_c, \theta)\pi(f_c|v)\pi(D),$$

where $f = \{f_p, f_c\}$. 

---

28
Demonstration

Enough already. Does this actually work??
Conclusion

Summary:
- Motivation (Chris)
- BaVA process (Scotland)
- Proof of concept by demo (Leanna)

Future
- Goal: Generalize BaVA to update any visualization
- Fine tune PPCA to assess complex, large datasets
- Develop BaVA methods for different types of models, e.g., tree diagrams, graphical models
- Develop meaningful visualizations that relate to update-able parameters
- Test our tool at PNNL with real analysts
Conclusion

» Summary:
  » Motivation (Chris)
  » BaVA process (Scotland)
  » Proof of concept by demo (Leanna)

» Future
  » Goal: Generalize BaVA to update any visualization
Conclusion

▶ Summary:
  ▶ Motivation (Chris)
  ▶ BaVA process (Scotland)
  ▶ Proof of concept by demo (Leanna)

▶ Future
  ▶ Goal: Generalize BaVA to update any visualization
  ▶ Fine tune PPCA to assess complex, large datasets
Conclusion

Summary:
- Motivation (Chris)
- BaVA process (Scotland)
- Proof of concept by demo (Leanna)

Future
- Goal: Generalize BaVA to update any visualization
- Fine tune PPCA to assess complex, large datasets
- Develop BaVA methods for different types of models e.g., tree diagrams, graphical models
Conclusion

Summary:
▶ Motivation (Chris)
▶ BaVA process (Scotland)
▶ Proof of concept by demo (Leanna)

Future
▶ Goal: Generalize BaVA to update any visualization
▶ Fine tune PPCA to assess complex, large datasets
▶ Develop BaVA methods for different types of models e.g., tree diagrams, graphical models
▶ Develop meaningful visualizations that relate to update-able parameters
Conclusion

▶ Summary:
   ▶ Motivation (Chris)
   ▶ BaVA process (Scotland)
   ▶ Proof of concept by demo (Leanna)

▶ Future
   ▶ Goal: Generalize BaVA to update any visualization
   ▶ Fine tune PPCA to assess complex, large datasets
   ▶ Develop BaVA methods for different types of models
e.g., tree diagrams, graphical models
   ▶ Develop meaningful visualizations that relate to update-able
   parameters
   ▶ Test our tool at PNNL with real analysts
Thank you very much