ASKE Milestone 2 for AMIDOL

Eric Davis[1], Alec Theriault[1], Max Orhai[1], Eddy Westbrook[1], and Ryan Wright[1]

[1]Galois, Inc

1 Introduction

Complex system analysis currently requires teams of domain experts, data scientists, mathematicians, and software engineers to support the entire life cycle of model-based inference. The models that result are often bespoke, lack generalizability, are not performable, and make it difficult to synthesize actionable knowledge and policies from their raw outputs. In this report we describe the current prototype system for AMIDOL: the Agile Metamodel Interface using Domain-specific Ontological Languages, a project that aims to reduce the overhead associated with the model life cycle and enables domain experts and scientists to more easily build, maintain, and reason over models in robust and highly performable ways, and to respond rapidly to emerging crises in an agile and impactful way. We discuss the current design principles of the AMIDOL prototype, its capabilities, plans for development, and formal aspects of the system.

AMIDOL is designed to support models in a number of scientific, physical, social, and hybrid domains by allowing domain experts to construct meta-models in a novel way, using visual domain specific ontological languages (VDSOLs). These VDSOLs utilize an underlying intermediate abstract representation to give formal meaning to the intuitive process diagrams scientists and domain experts normally create. AMIDOL's abstract representations are executable, allowing AMIDOL's inference engine to execute prognostic queries on reward models and communicate results to domain experts. AMIDOL binds results to the original ontologies providing more explainability when compared to conventional methods.

AMIDOL addresses the problem of machine-assisted inference with two high-level goals:

1. improving the ability of domain experts to build and maintain models and
2. improving the explainability and agility of the results of machine-inference.

Our techniques for achieving these goals incorporate abstract functional representations, intermediate languages, and semantic knowledge representation and binding in graph structures into traditional machine learning and model solution techniques.

2 VDSOL Definition

2.1 Basic Language Properties

Nouns :

Verbs :

5 Reward Variables and Reward Models

5.1 Rate Reward Variables

5.2 Impulse Reward Variables

5.3 Temporal Characteristics of Reward Variables

5.4 Translation of Reward Variables to IR

5.5 Expressions on Reward Variables

6 Design of Experiments and Results Database

6.1 Results Database

6.2 Prognostic Queries

6.3 Model Comparison

6.4 Design of Experiments

6.5 Conterfactural Exploration, Planning, Crisis Response

6.6 Correctness and Uncertainty

6.7 Communication of Results

7 Domain Models

We are currently testing AMIDOL using several domain models whose primary domain is epidemiology. We have selected a range of models to test different scenarios, use cases, and assumptions to aid in the prototype design of AMIDOL.

7.1 SIS/SIRS

The SIS/SIRS model is one of the simplest models we have deployed for testing with AMIDOL, with the advantage that the model itself is relatively simple, but utlizes real data, and can be used to answer important epidemiological questions. The primary objective of the SIS/SIRS model is to identify the *basic reproduction number* associated with an infection, also known as R_0, or *r nought*. R_0 was first used in 1952 when studying malaria and is a measure of the potential for an infection to spread through a population. If $R_0 < 1$, then the infection will die out in the long run. If $R_0 > 1$, then the infection will spread. The higher the value of R_0, the more difficult it is to control an epidemic.

Given a 100% effective vaccine, the proportion of the population that needs to be vaccinated is $1 - 1/R_0$, meaning that R_0 can be used to plan disease response. This assumes a homogenous population, and contains many other simplifying assumptions and does not generalize to more complex numbers. We have several main goals for SIS/SIRS models:

1. Fitting the models for the data in hindsight to perform goodness of fit estimates.
2. Finding the *retrospective* R_0 estimate over the entire epidemic curve.
3. Finding the *real-time* R_0 estimate while the epidemic is ongoing.

Data : For these models we will be working with the WHO/NREVSS (World Health Organization/National Respiratory and Enteric Virus Surveillance System) data sets at the resolution of Department of Human and Health Services designated regions.

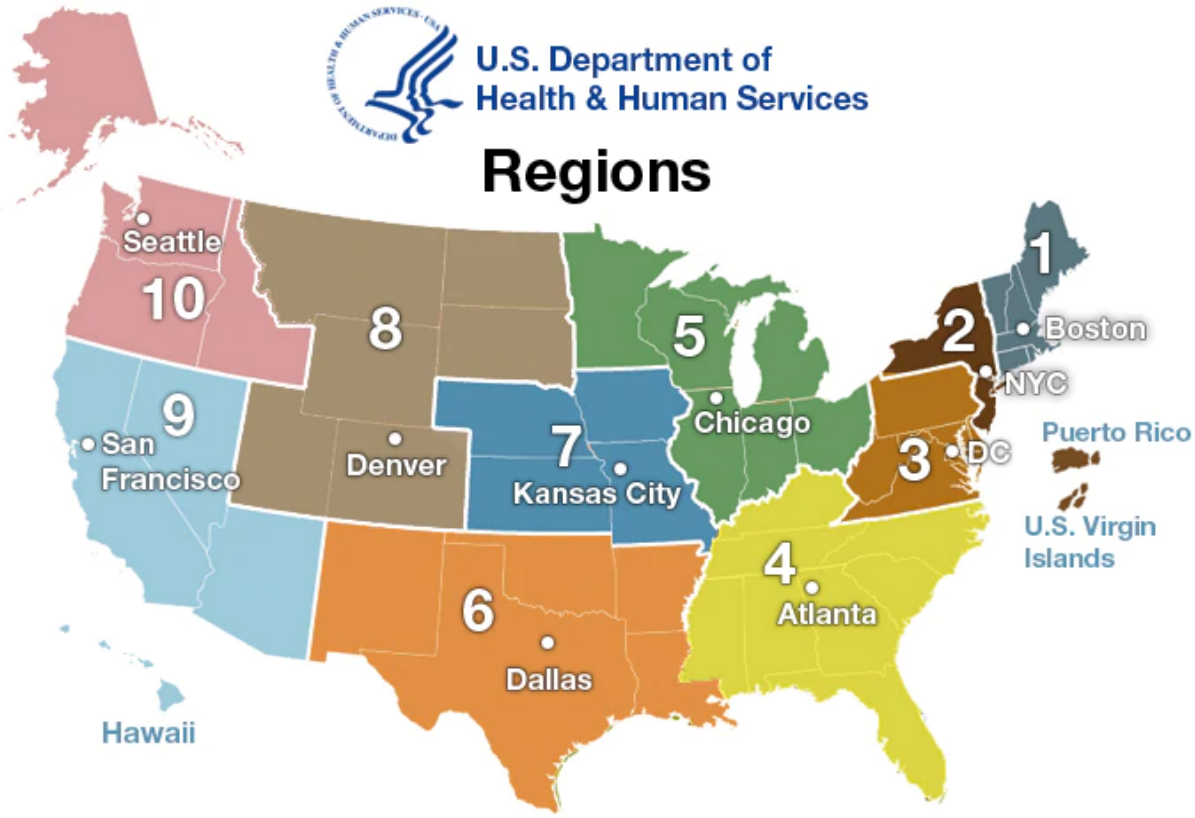


Figure 1: Department of Human and Health Services designated regions.

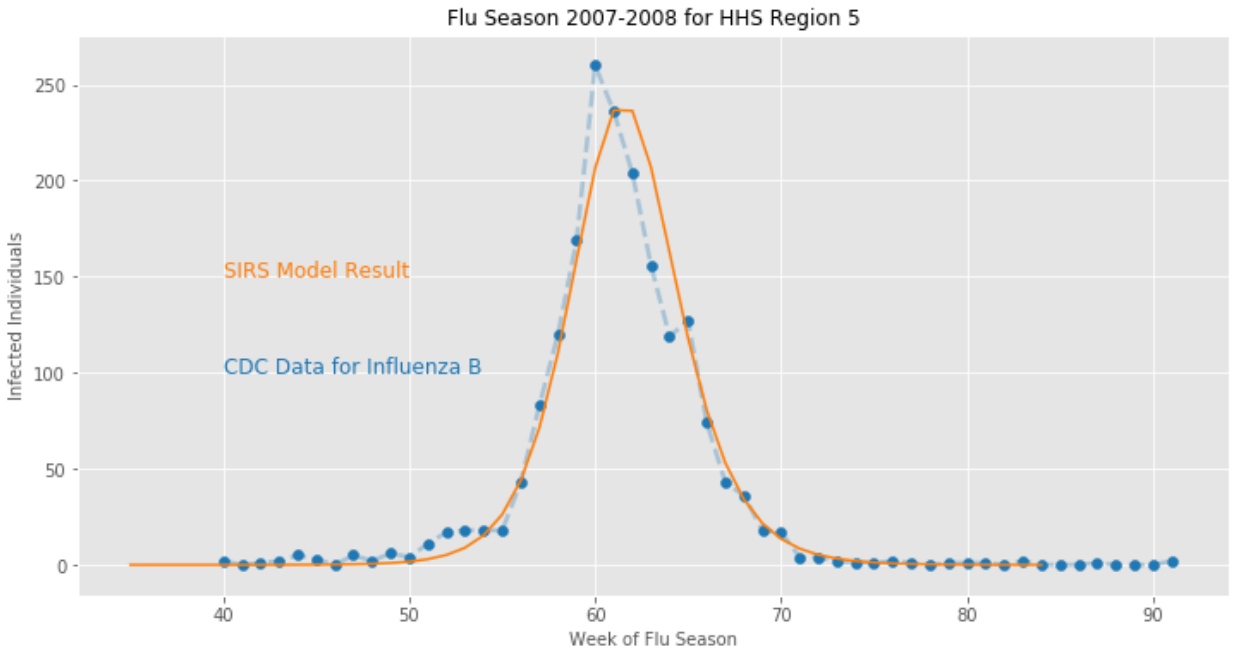


Figure 2: 2007 - 2008 Flu Season

Using data from a given region, and a given strain, we will estimate R0 for the epidemic curve as shown in Figure 2

7.2 Artificial Chemistry

7.3 Viral Infection Model

Note the use of multiple "Tat" symbols in Figure 3a. Sometimes scientists draw the same symbol multiple places as an "alias" for the same underlying state variable.

$$LTR \overset{k_{basal}}{\rightarrow} LTR + nRNA \tag{1}$$

$$nRNA \overset{k_{export}}{\rightarrow} cRNA \tag{2}$$

$$cRNA \overset{k1_{translate}}{\rightarrow} GFP + cRNA \tag{3}$$

$$cRNA \overset{k2_{translate}}{\rightarrow} Tat + cRNA \tag{4}$$

$$Tat \overset{k_{bind}/k_{unbind}}{\leftrightarrow} pTEFb_d \tag{5}$$

$$LTR + pTEFb_d \overset{k_{acetyl}/k_{deacetly}}{\leftrightarrow} pTEFb_a \tag{6}$$

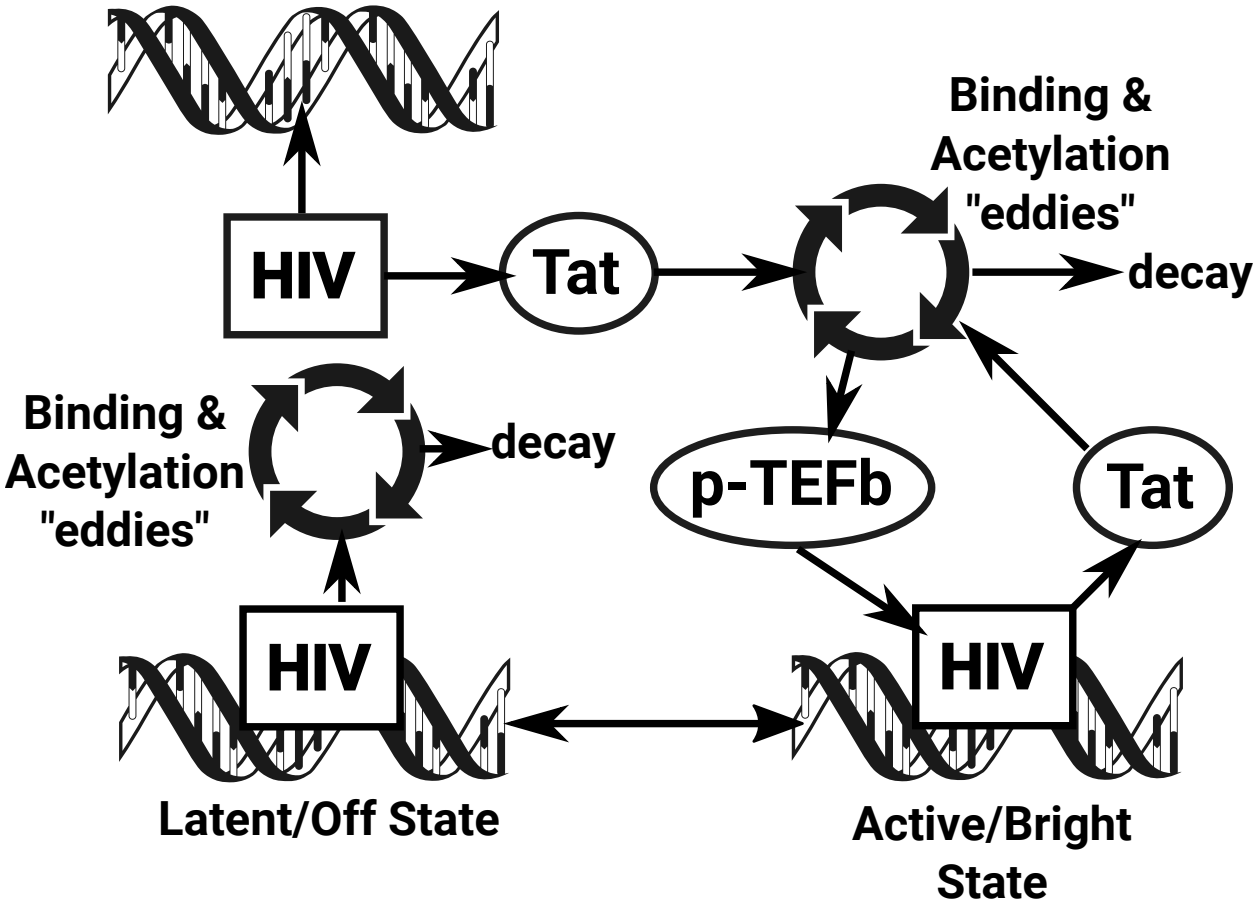
$$pTEFb_a \overset{k_{transact}}{\leftrightarrow} LTR + nRNA + Tat \tag{7}$$

$$GFP \overset{d_{GFP}}{\rightarrow} \emptyset \tag{8}$$

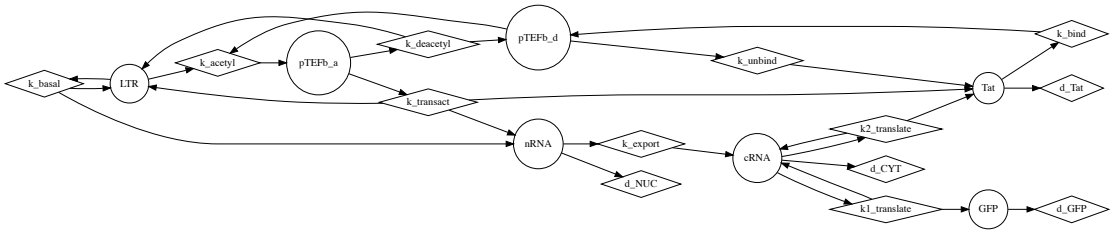
$$Tat \overset{d_{Tat}}{\rightarrow} \emptyset \tag{9}$$

$$cRNA \overset{d_{CYT}}{\rightarrow} \emptyset \tag{10}$$

$$nRNA \overset{d_{NUC}}{\rightarrow} \emptyset \tag{11}$$



(a) Semi-formal diagram of the molecular model of the Tat transactivation circuit.



(b) Simple noun (circle) and verb (square) representation of Tat model without ambiguity and aliasing.