A Logical Characterization of Individual-Based Models

James F. Lynch

Department of Computer Science
Clarkson University

June 26, 2008

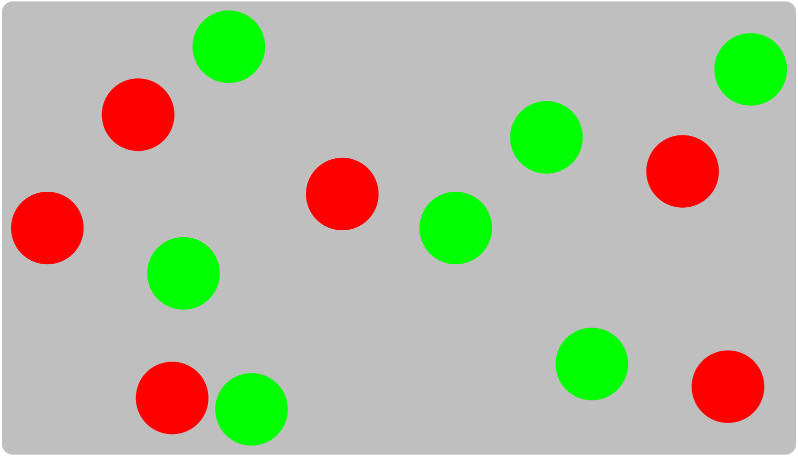
What Is an Individual-Based Model? (v. 1)

An IBM consists of populations of individuals.

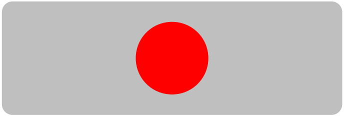
It evolves via interactions among the individuals.

An Example

Two species: Predator and Prey.
Individuals move freely and rapidly in an enclosed space.

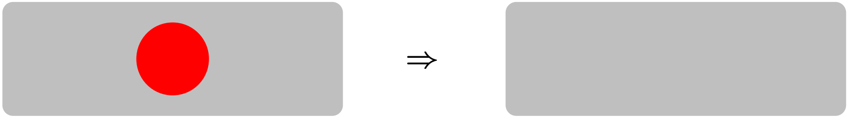


Death of Predator

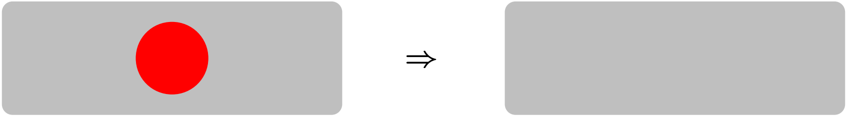


Interactions

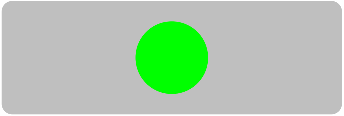
Death of Predator



Death of Predator

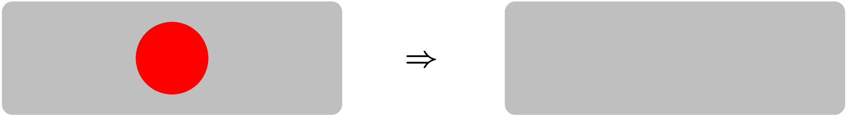


Birth of Prey

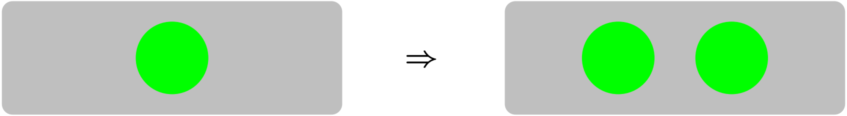


Interactions

Death of Predator

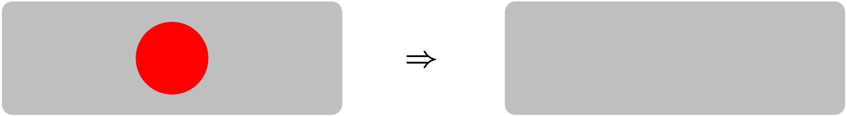


Birth of Prey

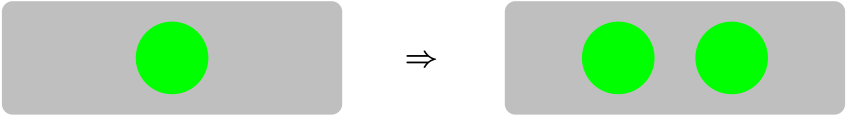


Interactions

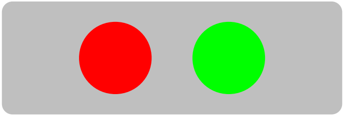
Death of Predator



Birth of Prey

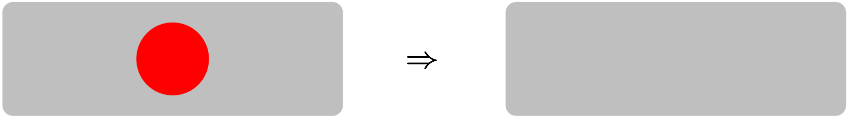


Predation

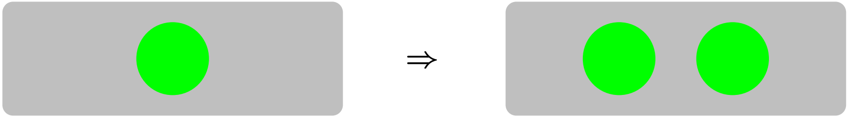


Interactions

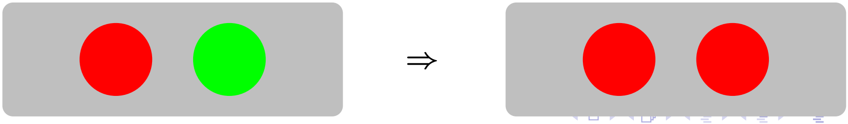
Death of Predator



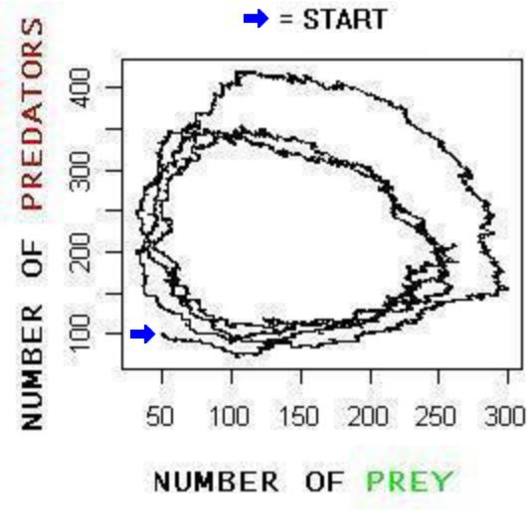
Birth of Prey



Predation

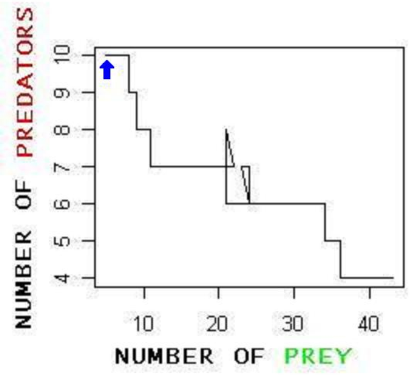


Typical Behavior

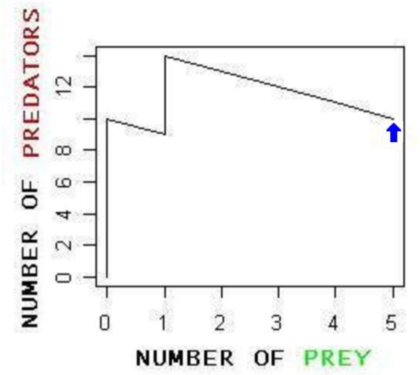


Extinction Is Inevitable

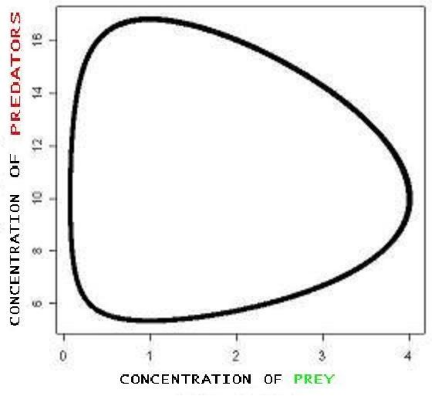
Extinction of Predator



Extinction of Both



Continuous Approximation of Large Populations (A State-Variable Model)



Lotka-Volterra Model

$$\frac{dx}{dt} = ax - bxy$$
$$\frac{dy}{dt} = -cy + bxy$$

Some Areas That Use Individual-Based Models

- Biology
  - Molecular Biology
  - Ecology
- Chemistry
- Computer Science
  - Internet Graphs
- Economics
- Physics
  - Statistical Mechanics
  - Galaxy Formation

The Need for a Formal Approach

Model complexity

- Many species of individuals
- Many types of interactions
- High cost of simulation and analysis

Design issues

- Discrete vs. continuous
- Probabilistic vs. deterministic
- Individuals vs. aggregations
  Reasons for aggregation:
  - More efficient
  - More “realistic" model for large populations
- Rigorous justifications for using SVMs

Goals

- A unifying conceptual framework for IBMs
- Classification of IBMs
  - Relationship between IBMs and SVMs
- Methodologies for dealing with complexity
  - Determining appropriate level of abstraction
  - Stepwise refinement
  - Efficient algorithms for simulation and analysis

Our Results

- A formal language for IBMs
- Classification of IBMs, including SVMs
- Definition of abstraction
- Characterization of IBMs that can be abstracted to SVMs
- Examples of IBMs that can not be abstracted to SVMs

What Is an Individual-Based Model? (v. 2)

- An IBM is a dynamical system whose states are metafinite models
- metafinite models = finite models
  +
  weight functions
  +
  numeric functions
  +
  multiset operations
- State transitions are probabilistic rules defined on metafinite models

Metafinite Models

Definition

1. A weight function of arity k on a set A is a partial function
$$w\colon A^k \to \mathbb{R}$$
2. A numeric function of arity k is a function
$$f\colon \mathbb{R}^k \to \mathbb{R}$$
3. A multiset over a set S is an unordered collection of elements from S with repetitions allowed. Ex: $\{\!|2, 5, 3, 2|\!\}$.
4. A multiset operation on S is a function
$$\Gamma\colon \{\text{finite multisets over } S\} \to \mathbb{R}$$

Example

$S = \mathbb{R}$, $\Gamma(M) = \sum_{r \in M} r$.

Metafinite Models Continued

Definition

1. A vocabulary is a triple $(\mathcal{W}, \mathcal{F}, \mathcal{G})$ where
   - $\mathcal{W}$ is a set of weight function symbols
   - $\mathcal{F}$ is a set of numeric function symbols
   - $\mathcal{G}$ is a set of multiset operation symbols
2. A metafinite model $\mathfrak{A}$ over $(\mathcal{W}, \mathcal{F}, \mathcal{G})$ is a structure $(A, \mathcal{W}^{\mathfrak{A}}, \mathcal{F}^{\mathfrak{A}}, \mathcal{G}^{\mathfrak{A}})$ where
   - A is the universe—a finite set (of individuals)
   - $\mathcal{W}^{\mathfrak{A}}$ is a set of interpretations on A of the weight function symbols in $\mathcal{W}$
   - $\mathcal{F}^{\mathfrak{A}}$ is a set of interpretations of the numeric function symbols in $\mathcal{F}$
   - $\mathcal{G}^{\mathfrak{A}}$ is a set of interpretations on $\mathbb{R}$ of the multiset operation symbols in $\mathcal{G}$

A Logic of Metafinite Models

The logic is a pure term calculus over the vocabulary.
Two types of variables:

1. Individual variables: values range over the universe A
2. Numeric variables: values range over $\mathbb{R}$

Two kinds of atomic terms:

1. Numeric variables.
2. $w(x_1, \ldots, x_k)$ where w is a k-ary weight function symbol and $x_1, \ldots, x_k$ are free individual variables.

A Logic of Metafinite Models Continued

Recursively,

1. If f is a k-ary numeric function symbol
   and $\tau_1, \ldots, \tau_k$ are terms,
   then $f(\tau_1, \ldots, \tau_k)$ is a term.
2. If Γ is a multiset operation symbol
   and τ is a term with free individual variables $x_1, \ldots, x_k, y$,
   then $(\Gamma y \tau)$ is a term with free variables $x_1, \ldots, x_k$.
   For $a_1, \ldots, a_k \in A$,

$$(\Gamma y \tau)^{\mathfrak{A}}(a_1, \ldots, a_k) = \Gamma(\{\!|\tau(a_1, \ldots, a_k, b) | b \in A|\!\})$$

Example

A weighted graph $\mathfrak{G} = (V, \{w_1^{\mathfrak{G}}, w_2^{\mathfrak{G}}\}, \{+, \times, -, /\}, \{|\ldots|, \sum\})$ where

V = vertices

for $a \in V$, $w_1^{\mathfrak{G}}(a) = 1$.

for $a, b \in V$, $w_2^{\mathfrak{G}}(a, b) = \begin{cases} \text{weight of edge } (a, b) \text{ if it exists} \\ \text{undef otherwise} \end{cases}$

$|\ldots|$ is the cardinality operator on multisets: $|\{\!|2, 5, 3, 2|\!\}| = 4$.

Expressing the number of vertices:

$$|\{\!| w_1(v) \mid v \in V |\!\}|$$

Expressing the outdegree of vertex v:

$$|\{\!| w_2(v, u) \mid u \in V |\!\}|$$

The average outdegree of $\mathfrak{G}$:

$$\sum \{\!| \, |\{\!| w_2(v, u) \mid u \in V |\!\}| \mid v \in V |\!\} / |\{\!| w_1(v) \mid v \in V |\!\}|$$

Transition Rules

$\mathfrak{A} = (A, \mathcal{W}^{\mathfrak{A}}, \mathcal{F}, \mathcal{G})$ and $\mathfrak{A}' = (A', \mathcal{W}^{\mathfrak{A}'}, \mathcal{F}, \mathcal{G})$ denote the states of the IBM before and after a transition.

Probability of a transition from $\mathfrak{A}$ to $\mathfrak{A}'$ is defined by a term in the vocabulary $(\{A, A'\} \cup \mathcal{W} \cup \mathcal{W}', \mathcal{F}, \mathcal{G})$.

Example

Graph growth model with preferential attachment. Probability of transition $\mathfrak{A} \to \mathfrak{A}'$ is

$$\sum_v \Big[v \notin A \wedge A' = A \cup \{v\} \wedge \mathrm{outdeg}(v) = 1 \\ \times \Big(\sum_u E'(v,u) \times \mathrm{indeg}(u) \Big) \Big] \Big/ \Big(\sum_u u \in A \times \mathrm{indeg}(u) \Big)$$

What Is an Individual-Based Model? (v. 3)

Definition

An IBM over vocabulary $(\mathcal{W}, \mathcal{F}, \mathcal{G})$ is a pair $(\mathcal{S}, \delta)$ where

- $\mathcal{S}$ is a set of metafinite models over $(\mathcal{W}, \mathcal{F}, \mathcal{G})$
- δ is a term over $(\{A, A'\} \cup \mathcal{W} \cup \mathcal{W}', \mathcal{F}, \mathcal{G})$ that defines a Markov process on $\mathcal{S}$.

Abstractions

Definition

Let $(\mathcal{S}, \delta)$ be an IBM over vocabulary $(\mathcal{W}, \mathcal{F}, \mathcal{G})$,
$(\mathcal{S}^\alpha, \delta^\alpha)$ be an IBM over vocabulary $(\mathcal{W}^\alpha, \mathcal{F}, \mathcal{G})$.
$(\mathcal{S}^\alpha, \delta^\alpha)$ is an abstraction of $(\mathcal{S}, \delta)$ if

- For every $w \in \mathcal{W}^\alpha$ there is a term τ_w in the logic of $(\mathcal{W}, \mathcal{F}, \mathcal{G})$ of the same arity as w.
- There is a map $\alpha \colon \mathcal{S} \to \mathcal{S}^\alpha$ such that for every $\mathfrak{A} \in \mathcal{S}$, if $\mathfrak{A}^\alpha = \alpha(\mathfrak{A})$ then
  - $A^\alpha \subseteq A$
  - for all $a_1, \ldots, a_i \in A^\alpha$ and $w \in \mathcal{W}^\alpha$ of arity i,

$$w^{\mathfrak{A}^\alpha}(a_1, \ldots, a_i) = \tau_w^{\mathfrak{A}}(a_1, \ldots, a_i)$$

Predator-Prey IBM Abstracted to Lotka-Volterra SVM

Example

States in $\mathcal{S}$ are metafinite models of the form
$(A, \{P^{\mathfrak{A}}, X^{\mathfrak{A}}, Y^{\mathfrak{A}}, Z^{\mathfrak{A}}\}, \{+, \times, -, /\}, \{|\ldots|, \sum\})$ where

- A is the set of all predators and prey
- $P^{\mathfrak{A}}(a) = 1$ if a is a predator; 0 otherwise
- $X^{\mathfrak{A}}(a) = x$-coordinate of a
- similarly for $Y^{\mathfrak{A}}$ and $Z^{\mathfrak{A}}$

States in $\mathcal{S}^{\alpha}$ are of the form
$(\emptyset, \{w_0^{\mathfrak{A}^{\alpha}}(), w_1^{\mathfrak{A}^{\alpha}}()\}, \{+, \times, -, /\}, \{|\ldots|, \sum\})$ where

$$w_0() \equiv |A| - \sum_{a \in A} P(a) \quad \text{(number of prey)}$$

$$w_1() \equiv \sum_{a \in A} P(a) \quad \text{(number of predators)}$$

Accuracy of Abstraction

Definitions

- For any time t, let $\mathfrak{A}_t$ (resp. $\mathfrak{A}^\alpha_t$) be the state of $(\mathcal{S}, \delta)$ (resp. $(\mathcal{S}^\alpha, \delta^\alpha)$) at time t.
- Let $[t, t + \Delta t]$ be a time interval. For $\mathfrak{A} \in \mathcal{S}$ and $r \in \mathbb{R}$, let

$$Q(\mathfrak{A}, r) = \Pr(\tau_w^{\mathfrak{A}_{t+\Delta t}} \leq r \mid \mathfrak{A}_t = \mathfrak{A})$$

and for $\mathfrak{A}^\alpha \in \mathcal{S}^\alpha$, let

$$Q^\alpha(\mathfrak{A}^\alpha, r) = \Pr(w^{\mathfrak{A}^\alpha_{t+\Delta t}} \leq r \mid \mathfrak{A}^\alpha_t = \mathfrak{A}^\alpha)$$

(The conditional cumulative distribution function of $\tau_w^{\mathfrak{A}_{t+\Delta t}}$ (resp. $w^{\mathfrak{A}^\alpha_{t+\Delta t}}$).)

Accuracy of Abstraction Continued

Definitions

- Let $\gamma \geq 0$ and $\epsilon \in [0, 1]$. w approximates τ_w with accuracy γ and confidence ϵ over $[t, t + \Delta t]$ if for all $r \in \mathbb{R}$, there is $s \in \mathbb{R}$ such that $|r - s| \leq \gamma$ and

$$\mathfrak{A}_t^\alpha = \alpha(\mathfrak{A}_t) \Rightarrow |Q(\mathfrak{A}, r) - Q^\alpha(\mathfrak{A}^\alpha, s)| \leq 1 - \epsilon \qquad (1)$$

- τ_w converges to w over $[t, t + \Delta t]$ if for all γ and ϵ, (1) holds for sufficiently large $|A_t|$.
- $(\mathcal{S}, \delta)$ converges to $(\mathcal{S}^\alpha, \delta^\alpha)$ over $[t, t + \Delta t]$ if τ_w converges to w for all $w \in \mathcal{W}^\alpha$.

$$\begin{array}{lccc}
\delta: & \mathfrak{A}_t & \xrightarrow{\Delta t} & \mathfrak{A}_{t+\Delta t} \\
& \alpha \big\downarrow & & \big\downarrow \alpha \\
\delta^\alpha: & \alpha(\mathfrak{A}_t) = \mathfrak{A}_t^\alpha & \xrightarrow{\Delta t} & \mathfrak{A}_{t+\Delta t}^\alpha \approx \alpha(\mathfrak{A}_{t+\Delta t})
\end{array}$$

Conditions That Imply Convergence

Let $(\mathcal{S}, \delta)$ be a discrete state and time IBM, τ a term. For $\mathfrak{A} \in \mathcal{S}$,

$$q(\mathfrak{A}, r) = Pr(\tau^{\mathfrak{A}_{t+1}} - \tau^{\mathfrak{A}_t} = r \mid \mathfrak{A}_t = \mathfrak{A})$$

Definition

Lipschitz: $|\tau^{\mathfrak{A}_{t+1}} - \tau^{\mathfrak{A}_t}| < c$ for some constant c.

Smoothness: $|\tau^{\mathfrak{B}} - \tau^{\mathfrak{A}}| < c \Rightarrow q(\mathfrak{B}, r) = q(\mathfrak{A}, r)(1 + o(1))$

Convergence to Mean-Field Approximation

Theorem

Let $(\mathcal{S}, \delta)$ and τ satisfy the Lipschitz and smoothness conditions. For any $\gamma > 0$ and $\epsilon \in [0, 1]$, for sufficiently large $\mathfrak{A}$,

$$\Pr(|\tau^{\mathfrak{A}_{t+\Delta t}} - \mathbf{E}(\tau^{\mathfrak{A}_{t+\Delta t}})| < \gamma \Delta t \mid \mathfrak{A}_t = \mathfrak{A}) > \epsilon$$

Corollary

Assuming $(\mathcal{S}, \delta)$ and $\tau_1, \ldots, \tau_k$ satisfy the Lipschitz and smoothness conditions for arbitrarily small intervals Δt, $(\mathcal{S}, \delta)$ converges to a k-dimensional deterministic SVM whose transitions are defined by a system of k ODEs.

Examples of IBMs That Converge to SVMs

Definition

A set of states $\mathcal{S}$ has bounded degree if there are finitely many isomorphism types among the Gaifman neighborhoods of radius 1 for $a \in \mathfrak{A} \in \mathcal{S}$.

Fact

If $(\mathcal{S}, \delta)$ is an IBM where $\mathcal{S}$ has bounded degree, then for any term τ, there is an integer k, an abstraction $\alpha : \mathcal{S} \to \mathbb{N}^k$, and a function $\Gamma : \mathbb{N}^k \to \mathbb{R}$ such that $\tau^{\mathfrak{A}} = \Gamma(\alpha(\mathfrak{A}))$.

Theorem

Let $(\mathcal{S}, \delta)$ be an IBM such that

- *$\mathcal{S}$ is of bounded degree*
- *All state transitions change only finitely many values of the weight functions*
- *δ and Γ satisfy certain Lipschitz and smoothness conditions*

Then $(\mathcal{S}, \delta)$ converges to a deterministic SVM.

Examples

- Models of chemical kinetics
- Lattice models of coupled chemical reactions
- Trophic webs
- Patch-occupancy models
- Graph growth models with bounded degree

Examples Not Satisfying Convergence

It is well-known that counting logics such as SQL cannot define topological properties like connectedness.

Using similar proof techniques, we show that certain processes of cellular metabolism cannot be modeled by SVMs:

Examples

- Chemical reaction systems controlled by membranes
- Transcription of long polymers