Evolutionary Computing, Neuroevolution and Genetic Algorithms: Toward Tumor-Growth Simulation

a.holzinger@hci-kdd.org

http://hci-kdd.org/machine-learning-for-health-informatics-course
ML needs a concerted effort fostering integrated research

Interactive

<table>
<thead>
<tr>
<th></th>
<th>Data Mining</th>
<th>Knowledge Discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data Mapping</td>
<td>Preprocessing</td>
</tr>
<tr>
<td>2</td>
<td>Learning Algorithms</td>
<td>Data Fusion</td>
</tr>
<tr>
<td>3</td>
<td>GDM</td>
<td>Graph-based Data Mining</td>
</tr>
<tr>
<td>4</td>
<td>TDM</td>
<td>Topological Data Mining</td>
</tr>
<tr>
<td>5</td>
<td>EDM</td>
<td>Entropy-based Data Mining</td>
</tr>
</tbody>
</table>

Privacy, Data Protection, Safety and Security

Red thread through this lecture

- 00 Reflection
- 01 Evolution
- 02 Neuroevolution
- 03 Genetic Algorithms
- 04 Medical Example: Tumor-Growth Simulation
00 Reflection
Let us start with a warm-up Quiz
Who watched this video?

Stephen Hawking's The Meaning of Life (John Conway's Game of Life segment) has 216,009 views.
01 Evolutionary Principles

“Evolution is the natural way to program”
Thomas S. Ray, University of Oklahoma,
http://life.ou.edu/

Who is Who in the Theory of Evolution

- **Jean Baptiste de Lamarck**, 1801. Theory of Inheritance of Acquired Characteristics, Paris


The goal of aML is to build systems that learn and make decisions without the human.

Early aML efforts, e.g. the perceptron [1], had been truly inspired by human intelligence.

Today, probabilistic modelling has become the cornerstone of aML [2], with applications in neural processing [3] and human learning [4].

Based on the evolutionary theories of Darwin, Lamarck, Baldwin, Mendel.

Since the 1980s, EAs have been used for optimization problems.

Exploring the possibility of optimizing machine learning algorithms rather recently [1]

Terminology of EC [1]

Evolutionary algorithms

Genetic algorithms
Evolutionary programming
Genetic programming, Neuroevolution,

Particle swarm optimization,
Artificial Bee algorithm,
Invasive Weed Optimization
Intelligent Water Drops
Ant colony optimization,
...

### Biological Universe vs. Computational Universe

<table>
<thead>
<tr>
<th>NOTION</th>
<th>BIOLOGICAL UNIVERSE</th>
<th>COMPUTATIONAL UNIVERSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome</td>
<td>DNA, protein, and RNA sequence in cells</td>
<td>Sequence of information objects</td>
</tr>
<tr>
<td>Fitness</td>
<td>Determines chances of survival and reproduction</td>
<td>Determines chances of survival and reproduction</td>
</tr>
<tr>
<td>Gene</td>
<td>Part of a Chromosome, determines a (partial) characteristic of an individual</td>
<td>Information object, e.g. a bit, a character, number etc.</td>
</tr>
<tr>
<td>Generation</td>
<td>Population at a point in time</td>
<td>Population at a point in time</td>
</tr>
<tr>
<td>Individual</td>
<td>Living organism</td>
<td>Solution candidate</td>
</tr>
<tr>
<td>Population</td>
<td>Set of living organisms</td>
<td>Bag or multi-set of Chromosomes</td>
</tr>
</tbody>
</table>

The General Evolutionary Computation Framework [1]

Lamarckian/Baldwin Adaptation [1]

- Modify chromosomes to adapt to the environment
  - can be used additionally or instead of mutation process
- A local search optimization is applied (e.g. Hill Climbing)
- **Baldwin** uses only pseudo adaptation

- **Naive Bayes** is a very effective classifier
- EAs need parameters that can be modified
- A **Weighted Naive Bayesian (wnb)** [1] classifier offers the possibility of easy optimization:

\[
p(a_1, a_2, \cdots, a_n|c) = \prod_{i=1}^{n} p(a_i|c). \quad V_{nb}(E) = \arg \max_c p(c) \prod_{i=1}^{n} p(a_i|c)
\]

\[
V_{wnb}(E) = \arg \max_c p(c) \prod_{i=1}^{n} p(a_i|c)^{w_i}
\]

- **Dataset**: Pima Indians Diabetes dataset [8]
  - 768 instances (patients)
  - 8 attributes
  - 2 classes

- **Fitness** of an chromosome determined by: number of correctly classified instances in training set

- **Performance** was compared to algorithms in Weka

Algorithm 1 Fitness function

1: procedure FITNESS FUNCTION(weightings[], List trainingSet)
2:     for all instances of trainingset do
3:         for i = 1 to NumberOfClass do
4:             for all attribute to MaxNumberOfAttributes do
5:                 probability[i] *= NORMDISTRIBUTION(attribute + weightings[attribute])
6: 
7:     index ← INDEX OF MAX(probability[])
8: 
9:     if index == CLASS OF(instance) then
10:        INCREMENT(fitness)
11:    else
12:        DECREMENT(fitness)
13:    
14:     RETURN fitness
Results and Encountered Problems


- **Advantages:**
  - Fast to train and fast to classify
  - Not sensitive to irrelevant features
  - Handles real and discrete data

- **Disadvantages:**
  - Assumes independence of features
 Conclusion

- Offers many possibilities to improve machine learning algorithms, but finding the right parameters is a difficult task
- Not many machine learning algorithms are suitable for direct function optimization
- Implementation of EA:
  - straightforward
  - simple
- EAs are suitable for many tasks in health informatics beyond function optimization

Future Research

- Improvement of function optimization strategy
- Use EAs in different **fields**
  - Graph Optimization
  - Text Mining [1]
  - Feature selection
- Usage of novel **evolutionary strategies**
  - Intelligent Water Drops
  - Invasive Weed
  - Ant Colony with humans-in-the-loop (Super-Ants)

Evolutionary Algorithms for Text Mining

- Text mining with EAs on unstructured information:
  - Doctors/Nurse reports
  - Different Medical Records
  - ...

- Sample applications:
  - Categorizing Texts into subject groups [1]
  - Mining “interesting” details [2] like:
    - Gender
    - Age
    - Addresses
    - Occupation

Evolutionary Computing
The evolution of evolutionary computing

- 1948 Alan Turing: “genetical or evolutionary search”
- 1962 Hans-Joachim Bremermann: optimization through evolution and recombination
- 1964 Ingo Rechenberg: introduces evolution strategies
- 1965 Lawrence J. Fogel, Owens and Walsh: introduce evolutionary programming
- 1975 John Holland: introduces genetic algorithms
- 1992 John Koza: introduces genetic programming
Evolution of evolutionary computing conferences

http://dblp2.uni-trier.de/db/conf/evoW/

http://dblp.uni-trier.de/db/conf/cec/

http://dblp.uni-trier.de/db/conf/gecco/
Macroscopic View on Natural Evolution

The text and diagram discuss macroscopic views on natural evolution, possibly including branches or taxa in a evolutionary tree.

Elements of the diagram include:
- A tree-like structure
- Branches labeled with letters (A, B, C, D, etc.)
- Handwritten notes around the diagram

The handwritten notes may provide additional context or explanations related to the evolutionary relationships or classifications depicted in the diagram.
Microscopic View on Natural Evolution

Genes contain instructions for making proteins.

Proteins act alone or in complexes to perform many cellular functions.
An evolving population is conceptualized as moving on a surface whose points represent the set of possible solutions = search space

BEGIN
INITIALISE population with random candidate solutions;
EVALUATE each candidate;
REPEAT UNTIL (TERMINATION CONDITION is satisfied) DO
1 SELECT parents;
2 RECOMBINE pairs of parents;
3 MUTATE the resulting offspring;
4 EVALUATE new candidates;
5 SELECT individuals for the next generation;
END

General Scheme of an Evolutionary Algorithm

- Initialisation
- Population
- Parent selection
- Parents
- Recombination (crossover)
- Mutation
- Survivor selection
- Offspring
- Termination
Basic Model of Evolutionary Process

- Population of individuals
- Each individual has a fitness function
- Variation operators: crossover, mutation, ...
- Selection towards higher fitness by “survival of the fittest” and “mating of the fittest”

**Neo Darwinism:**
Evolutionary progress towards higher life forms = Optimization according to some fitness-criterion (optimization on a fitness landscape)
Two competing forces

1) Increasing population diversity by genetic operators (e.g. mutation, recombination, ...)
Push towards creating novelty

2) Decreasing population diversity by selection of parents and survivors
Push towards quality
Role: provides code for candidate solutions that can be manipulated by variation operators, and leads to two levels of existence:

- **phenotype**: object in original problem context (outside)
- **genotype**: code to denote that object, the inside (chromosome, “digital DNA”)

Implies two mappings:

- **Encoding**: phenotype → genotype (not necess. 1:1)
- **Decoding**: genotype → phenotype (must be 1:1)

Chromosomes contain genes, which are in (usually fixed) positions called loci and have a value (allele)
In order to find the global optimum, every feasible solution must be represented in the genotype space.

Evaluation function = Fitness function

- **Role:**
  - Represents the task to solve, the requirements to adapt to (can be seen as “the environment”)
  - Enables selection (provides basis for comparison)
  - e.g., some phenotypic traits are advantageous, desirable, e.g. big ears cool better, these traits are rewarded by more offspring that will expectedly carry the same trait

- **A.k.a. quality function or objective function**
  - Assigns a single real-valued fitness to each phenotype which forms the basis for selection
    - So the more discrimination (different values) the better
  - Typically we talk about fitness being maximised
    - Some problems may be best posed as minimisation problems, but conversion is trivial
Population

- **Role:** holds the candidate solutions of the problem as individuals (genotypes)
- Formally, a population is a multiset of individuals, i.e. repetitions are possible
- Population is the basic unit of evolution, i.e., the population is evolving, not the individuals
- Selection operators act on population level
- Variation operators act on individual level
- Some sophisticated EAs also assert a spatial structure on the population e.g., a grid
- Selection operators usually take whole population into account i.e., reproductive probabilities are relative to current generation
- **Diversity** of a population refers to the number of different fitness / phenotypes / genotypes present (note: not the same thing)
Selection mechanism

Role:
- Identifies individuals
  - to become parents
  - to survive
- Pushes population towards higher fitness
  - Usually probabilistic
    - high quality solutions more likely to be selected than low quality
    - but not guaranteed
    - even worst in current population usually has non-zero probability of being selected
  - This *stochastic* nature can aid escape from local optima
Most EAs use fixed population size so need a way of going from (parents + offspring) to next generation

Often deterministic (while parent selection is usually stochastic)

Fitness based: e.g., rank parents + offspring and take best

Age based: make as many offspring as parents and delete all parents

Sometimes a combination of stochastic and deterministic (elitism)
Variation operators

- Role: to generate new candidate solutions
- Usually divided into two types according to their arity (number of inputs):
  - Arity 1: mutation operators
  - Arity >1: recombination operators
  - Arity = 2 typically called crossover
  - Arity > 2 is formally possible, seldom used in EC
- There has been much debate about relative importance of recombination and mutation
- Nowadays most EAs use both
- Variation operators must match the given representation
Mutation

- Role: causes small, random variance
- Acts on one genotype and delivers another
- Element of randomness is essential and differentiates it from other unary heuristic operators
- Importance ascribed depends on representation and historical dialect:
  - Binary GAs – background operator responsible for preserving and introducing diversity
  - EP for FSM’s / continuous variables – only search operator
  - GP – hardly used
- May guarantee connectedness of search space and hence convergence proofs
Recombination

- Role: merges information from parents into offspring
- Choice of what information to merge is stochastic
- Most offspring may be worse, or the same as the parents
- Hope is that some are better by combining elements of genotypes that lead to good traits
- Principle has been used for millennia by breeders of plants and livestock
Initialization/ Termination

- Initialisation usually done at random,
- Need to ensure even spread and mixture of possible allele values
- Can include existing solutions, or use problem-specific heuristics, to “seed” the population

- Termination condition checked every generation
- Reaching some (known/hoped for) fitness
- Reaching some maximum allowed number of generations
- Reaching some minimum level of diversity
- Reaching some specified number of generations without fitness improvement
Types of EAs

- Historically different EAs have been associated with different data types to represent solutions
- Binary strings: Genetic Algorithms
- Real-valued vectors: Evolution Strategies
- Finite state Machines: Evolutionary Programming
- LISP trees: Genetic Programming
- These differences are largely irrelevant, best strategy
- choose representation to suit problem
- choose variation operators to suit representation
- Selection operators only use fitness and so are independent of representation
Typical EA behaviour

Goldberg, D. E. 1989. Genetic algorithms in search, optimization, and machine learning, Reading (MA), Addison-Wesley
- **Individuals**: hypothesis $x$ from a hypothesis space $X$
- **Population**: collection $P$ of $\mu$ hypotheses $P = \{x_i \mid i = 1, \ldots, \mu\}$
- **Evaluation**: $f : X \rightarrow R$ (fitness function) to all individuals
- **Selection mechanism**: selects individuals $x \in P_i$ for reproduction (mating); selects individuals from off-springs and $P_i$ to form the new population $P_i + 1$
- **Reproduction**: combination of two or more individuals (Crossover) and random alteration (Mutation).
Algorithm 1 Fitness function

1: procedure FITNESS FUNCTION(weightings[], List trainingSet)
2:   for all instances of trainingset do
3:     for $i = 1$ to NumberOfClass do
4:       for all attribute to MaxNumberOfAttributes do
5:         probability[$i$] *= NORMDISTRIBUTION(attribute + weightings[attribute])
6:     index ← INDEX OF MAX(probability[])
7:     if index == CLASS OF(instance) then
8:       INCREMENT(fitness)
9:     else
10:    DECREMENT(fitness)
11:   RETURN fitness
02 Neuro Evolution *)

Note: this is ML - not to confuse with neural evolution!
Def.: Risto Miikkulainen tutorial at GECCO 2005
Define: Neuroevolution

- is a form of machine learning that uses evolutionary algorithms to train deep learning networks.
- method for optimizing neural network weights and topologies using evolutionary computation. It is particularly useful in sequential decision tasks that are partially observable (i.e. POMDP) and where the state and action spaces are large (or continuous).
- Application: Games, robotics, artificial life
- neuroevolution can be applied more widely than supervised learning algorithms

http://nn.cs.utexas.edu/keyword?neuroevolution
An example

Neuroevolution - Car learns to drive

[YouTube video thumbnail]

Published on Apr 19, 2014

test of neuroevolution with cars that learn to drive a track

73,917 views
Future Challenge: to utilize domain knowledge for problem solving
Sometimes we have knowledge and sometimes random initial behavior is not acceptable
Grand question: How can domain knowledge be utilized?
by incorporating rules
by learning from examples
Neuroevolution is closely related to Genetic Algorithms

http://rednuht.org/genetic_cars_2/

http://rednuht.org/genetic_walkers/
Video Sample of Genetic Algorithms

https://www.youtube.com/watch?v=uxourrlPlf8
03 Genetic Algorithms
- Similar to stochastic optimization
- Iteratively trying to improve a possibly large set of candidate solutions
- Few or no assumptions about the problem (need to know what is a good solution)
- Usually finds good rather than optimal solutions
- Adaptable by a number of adjustable parameters
The landscape of Natural Computing

Image Credit to Johann Dréo, Caner Candan - Metaheuristics classification CC BY-SA 3.0 https://commons.wikimedia.org/w/index.php?curid=16252087
K= 2 → **Two-armed bandit problem:**
Arm 1: award $\mu_1$ with variance $\sigma_1^2$
Arm 2: award $\mu_2$ with variance $\sigma_2^2$
$\mu_1 > \mu_2$

Question: Which arm (left/right) is which index 1, 2?

Can be used for motivation of the Schema Theorem by John Holland (1975): is widely taken to be the foundation for explanations of the power of genetic algorithms: low-order schemata with above-average fitness increase exponentially in successive generations.

Holland, J. H. 1975. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, U Michigan Press (as of 01.06.2016 49,320 citations !)
The famous theorem by John Holland

- \( N = \) total number of trials

- \( b = \frac{\sigma_1}{\mu_1 - \mu_2} \)

- **Conclusion:**
  Expected loss is minimal if approximately:

\[
n^* \approx b^2 \cdot \ln \left( \frac{N^2}{8\pi \cdot b^4 \cdot \ln(N^2)} \right)
\]

- Consequently, trials are allocated to the observed worst arm
The “expected loss” is minimal if (approx.)

\[ N - n^* \approx \sqrt{8\pi \cdot b^4 \cdot \ln(N^2)} \cdot \exp\left(\frac{n^*}{2b^2}\right) \]

- The trials are allocated to the **observed** best arm
- This 2-arm bandit can be generalized to a k-armed bandit, resulting in:
  - A) Generalized corollary: The optimal strategy is to allocate an exponentially increasing \( n \) of trials to the **observed** best arm
  - B) This links-up to Genetic Algorithms because: Minimizing expected losses from k-armed bandits \( \approx \) Minimizing expected losses while sampling from order \( \log_2(k) \) schemata (=GA’s allocate trials opt.)
A point of criticism came from David P. Fogel (1995)

- Why would this be optimal for global optimization?
- Minimizing expected losses does not always correspond to maximizing potential gains.
04 Tumor Growth Simulation
1) What is Cancer: A biological introduction
2) The multistep process of cancer
3) Key Problems for cancer research
4) Overview of Machine Learning for cancer
4) Tumor Growth Modeling
5) Cellular Potts Model > Tumor Growth Simulation
6) Implementation of Tumor Growth Visualization
7) Summary and Open Problems
Keywords

- In Silico
- Differentiation
- Benign & Malignant Tumor Cells
- Proliferation
- Migration
- Tissue
- Adhesion
- Tumor Growth Modeling
- Agent Based Modeling
- Cellular Automaton (CA)
- Extra-Cellular Matrix (ECM)
- Cellular Potts Model (CPM)
Part 1: Biology
Cancer Statistics

OGB – German Health News, apr 5th 2016 - Annually mortality causes in EU28 and other industrial countries

**Occupational Deaths**

- Infectious diseases
- Cancer diseases
- Pulmonary diseases
- Cardiac diseases
- Psychological disorders
- Stomach/bowel diseases
- Urogenital diseases
- Accidents/Traumata
Cancer Sites of All Cases Worldwide

- Lung: 13%
- Breast: 12%
- Bowel: 10%
- Prostate: 8%
- Stomach: 6%
- Liver: 6%
- Cervix: 8%
- Oesophagus: 8%
- Bladder: 6%
- Non-Hodgkin Lymphoma: 6%
- Other sites: 8%

Cancer Statistics

World Health Organization - WHO - Age standardized rate of Cancer Site Incidence & Mortality Worldwide 2012

Holzinger Group, hci-kdd.org
Cancer Definitions

- **Tumor / Neoplasm**
  ... *Abnormal mass of tissue - cells divide more than they should*

- **Cancer**
  ... *group of diseases in which abnormal cells divide without control, can invade nearby tissues*

- **Malignant**
  ... *Cancerous: invasive, destroy nearby tissue, spread to other parts of the body*

- **Benign / Non-malignant**
  ... *normal (not cancerous): grow larger but do not spread*

- **Hyperplasia / Dysplasia**
  ... *increased number / abnormal form*

Most common Cancer types, based on their origin (primary manifestation):

- Skin
- Lung
- Breast
- Prostate
- Colon & rectum
- Uterus
Cell Types originate from 3 different germ layers

- **Endoderm** (inner layer) → digestive tract, liver, lung, pharynx ...
- **Ectoderm** (outer layer) → skin, nails, hair, skin glands, salivary glands, nerve tissue ...
- **Mesoderm** (middle) → muscles, fibrous tissue, bone, cartilage, adipose tissue, blood cells, blood&lymph vessels

The-three-germ-layers - http://madmnemonics.blogspot.co.at - 06/2015
- Tissue
- Extracellular Matrix (ECM)
- Cell → Organelles
- Microfibril → Fiber → Protein

Loose connective tissue - by Adrignola, Sunshineconelly, Lawson R. 2011
Histological Types: Hundreds of different cancers, summed up to 6 major categories:

- **Carcinoma** (epithelial tissue)
- **Sarcoma** (supportive/connective tissue)
- **Myeloma** (plasma/bone marrow cells)
- **Leukemia** (bone marrow → blood production)
- **Lymphoma** (lymphatic system)*
- **Mixed** Types (eg. Carcinosarcoma)

* Hodgkin/Non-Hodgkin lymphoma depending on presence of Reed-Sternberg cells

ICD-0-3 - The International Standard for the classification and nomenclature of histologies is the International Classification of Diseases for Oncology, 3rd Ed.
NIH SEER Training Modules - US Department of Health & Human Services, National Institutes of Health, National Cancer Institute.
Cancer Classification

Nomenclature based on tissue type and malignancy/benignancy:

eg.  *Adenoma* (benign) & *Adenocarcinoma* (malignant),
     *Fibroma* & *Fibrosarcoma*
     *Neuroma* & *Neuroblastoma*

ICD-O-3 - The International Standard for the classification and nomenclature of histologies is the International Classification of Diseases for Oncology, 3rd Ed.
NIH SEER Training Modules - US Department of Health & Human Services, National Institutes of Health, National Cancer Institute.
Tumor Grading

- Grades
  - G1 (undetermined)
  - G2 (well differentiated)
  - G3 (poorly differentiated)
  - G4 (undifferentiated)

- Cancer type-specific grading
  - Gleason Scoring - prostate cancer - calculated from pattern 1-5
    - X, 1-6 (well diff.), 7 (moderately), 8-10 (poorly diff.)
  - Nottingham system - breast cancer
    (based on tubule formation, nuclear grade, mitotic rate)
- Tumor location
- Cell type
- Tumor Size
- Spread to lymph nodes
- Spread to different parts of body
- Tumor grade = cell abnormality
  (proliferation rate, nuclear hyperchromasia, mitoses)
**Cancer Staging**

- **TNM** system (extent/number/metastasis)
  
  X,0,T1-4, N1-3, M1
  
  eg. T3N0M0 (large tumor, no cancer in nearby lymph nodes/tissue, not spread to distant body parts)

- **Stage**
  
  0 (carcinoma in situ)
  
  I-III (size and spread to nearby tissue)
  
  IV (metastasis to distant parts)

NCI – National Cancer Institute, at the National Institutes of Health - About Cancer – Diagnosis and Staging, March 9th 2015
- **In situ** - abnormal cells present but not spread to nearby tissue
- **Localized** - cancer limited to origin, not spread
- **Regional** - cancer spread to nearby lymph nodes, tissues, organs
- **Distant** - cancer spread to distant parts of the body
- **Unknown**

NCI – National Cancer Institute, at the National Institutes of Health - About Cancer - Diagnosis and Staging, March 9th 2015
Development of cancer

Cancerogen, Promoting Stimuli

Mutation

Programmed cell death
Proliferation & Apoptosis Inhibition

Uncontrolled cell growth

DNA structure - Krebsinformationsdienst, Deutsches Krebsforschungszentrum Jan 2016.
S.Jorhaa’ir / Garak76, 2010 cell division – normal vs. cancer.
Cancer cycle and cancerous cells - © 2015 OncoSera.

Holzinger Group, hci-kdd.org
Development of cancer

DNA mutation

Onset of cancer - modified from Lü et al. Spandidos Publications 2012
What is Cancer: A biological background

- Differentiation
  ... *cell changing to a more specialized cell type*

- Proliferation
  ... *growth: increase in cell number via cell division*

- Mitosis
  ... *cell division*

- Apoptosis
  ... *programmed cell death, blocked in cancer cells*

- Necrosis
  ... *unprogrammed / general cell death*

*NCI Dictionary of Cancer Terms - US Department of Health & Human Services, National Institutes of Health, National Cancer Institute.*
*Intro to Apoptosis - Gewies A. Nov. 6th 2014, http://de.slideshare.net/richardhastings589/kumc-measuring-apoptosis-using-flow-cytometry*
Signal transduction pathways – Commons, by Bohog2 sept. 6th 2008
Carcinogenicity

- Increased frequency in spontaneous formation of tumors
- Reduced latency time
- Tumor occurrence in additional tissues
- Increased number of tumors

- Genotoxicity: direct DNA damage
- Non-genotoxic: indirect damage on external genetic influence factors

BG BAU – Berufsgenossenschaft der Bauwirtschaft, Berlin 2016, GISBAU – Gefahrstoff-Informationssystem
Toxicity / Cancerogenity studies

- Daily administration of test substance to animal (oral, dermal, inhalative) –18-30 months (live-long for rodents)
- Chron. toxicity by repeated dosing > 12 months
- Histopathological changes (hyperplasia, atypia), rates of cell division

---

Humanized mice in translational biomedical research, Leonard D. Shultz, Fumihiko Ishikawa and Dale L. Greiner, Nature Reviews Immunology 7, 2007
Humanized mice in translational biomedical research, Leonard D. Shultz, Fumihiko Ishikawa and Dale L. Greiner, Nature Reviews Immunology 7, 2007
Mouse xenograft surgery - www.youtube.com/watch?v=R2Wka7YhhAo
mouse xenograft surgery

https://www.youtube.com/watch?v=R2Wka7YhhAo
Cancer studies

Humanized mice in translational biomedical research, Leonard D. Shultz, Fumihiko Ishikawa and Dale L. Greiner, Nature Reviews Immunology 7, 2007
Jeanquartier et al. 2016
Dank Bioinformatik haben wir nun ein klares Bild von der Krankheit.
“Somehow your medical records got faxed to a complete stranger. He has no idea what’s wrong with you either.”
Part 2
Computational Modelling
Are Computers better doctors?
Key Problems

- Inter- and intracellular **dynamics**
- avoiding **hard-to-measure** variables
- **Inflexible** models
- *in silico* complements *in vivo*
- **executable (cell) biology**
- **reduce** animal experiments (resources)
- **boost** *in silico* for awareness & breakthrough
- **patient-personalized** prediction


Cancer Research

- Images of tissue

Overview of Machine Learning for cancer research

- ML in Genomics
  - such as DNA micro array analysis for cancer classification etc.
  - => for identification & treatment

- ML in image analysis
  - such as for classifying and/or differentiating benign from malignant samples etc.
  - => for diagnosis & prognosis

- ML in cancer research is growing rapidly
  - combination of molecular patterns and clinical data
  - deep text-mining offers new possibilities
  - etc.

Example: Glioma Classification

- Using gene expression data
- Unsupervised ML approach on genome-wide gene expression profiles of 159 gliomas
- Model predicts
  - two major groups,
  - separated into six subtypes,
  - previously unrecognized prognostic groups within TCGA published data could be found

Aiguo Li, Jennifer Walling, Susie Ahn, Yuri Kotliarov, Qin Su, M. Quezado, J. C. Oberholtzer, J. Park, J. C. Zenklusen, H. A. Fine: Unsupervised Analysis of Transcriptomic Profiles Reveals Six Glioma Subtypes, DOI: 10.1158/0008-5472.CAN-08-2100 Published 1 March 2009
Example: Modeling glioma tumor growth

- Using image data (MRI scans)
- Learn the parameters of a diffusion model
  - Using patient data
  - Preprocessing images
    - noise reduction, linear register and warp to standard coordinate system, reducing inhomogeneity, Intensity standardization, segmentation between grey and white matter...
  - Feature extraction
- => Prediction through classification & diffusion

There are different kinds of models in biology, such as spatial ones, space free ones but also cell descriptive models based on density, or cell-based, or sub-cellular or molecular, (relating to their scale of phenomenon)

Visualization Applications At Different Biological Scales

- Atoms
- DNA/RNA (nucleic acids and bigger biomolecules...)
- Proteins
- Viruses
- Bacteria
- Cells
- Tissues

Log. scale:
- 0.1nm
- 1nm
- 10nm
- 100nm
- 1micron
- 1mm

Nanoscale | Microscale | Mesoscale
- **Tumor growth**
- **complex disease**: simplification & approximation
- **differentiation** of normal cells
- excessive **proliferation**
- either **dormant** or **growing**
- **critical** mass
  - growth stops
  - migration (metastasis)
- **underlying** network structure
- environmental heterogeneities

A tumor can be seen as *spatio-temporal* pattern formation

Spatial & temporal data exist and can be used for improving existing simulation & analysis tools

Several attempts have been made to *model* and *predict* malignant tumor


- Tumor growth **kinetics follow** simple laws
- Mathematical models **exist** f.i. Gompertz or power law
- No **universal** law
- Prediction rate low and/or distinct

Tumor Growth Modeling

**Tumor model categories**

- Biological Scale
  - Tissue Behavior
  - Cell-cell Interactions (communities)
  - Cellular Function
  - Functioning and Regulation of Pathways
  - Molecular Interactions
  - Gene Expression
  - Genetics, Genomics

- Statistical Models
  - Pathway Enrichment Models
  - Gene Expression Models

- Network-based Models
  - Stoichiometric Models
  - Kinetic Models
  - Biochemical Reaction Networks

- Tissue-level Models
  - Continuum Models
  - Agent-Based Models

---


Holzinger Group, hci-kdd.org
## Continuum vs. Discrete/Agent-based Modeling

<table>
<thead>
<tr>
<th>Continuum</th>
<th>Discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuously distributed variables</td>
<td>discrete entities in discrete time intervals</td>
</tr>
<tr>
<td>interactions between factors representing several effects of physiological/biochemical events</td>
<td>interactions in a single space representation</td>
</tr>
<tr>
<td>f.i.: simulating population dynamics, combinatoric effects of several nutrient availability and other parameters etc.</td>
<td>f.i.: simulating agent dynamics, probabilistics of each time step, a small number of individuals</td>
</tr>
</tbody>
</table>

- Cellular Automaton (CA) approach to modeling biological cells

- CAs are:
  - Discrete
  - Abstraction of a system
  - Computational
  - At each time, each cell instantiates one state of a finite set of states

- CA for tumor growth with rules:
  - Cell division, movement, change or not change state...

- “On-lattice” modelling (see next slide)

In abstract algebra it is a fundamental algebraic structure, consisting of a partially ordered set in which every two elements have a unique supremum (join) and a unique infimum (meet). An example is given by the natural numbers, partially ordered by divisibility, for which the unique supremum is the least common multiple and the unique infimum is the greatest common divisor.

In geometry a lattice in $\mathbb{R}^n$ is a subgroup of $\mathbb{R}^n$, which is isomorphic to $\mathbb{Z}^n$, and which spans the real vector space $\mathbb{R}^n$, i.e. for any basis of $\mathbb{R}^n$ the subgroup of all linear combinations with integer coefficients of the basis vectors forms a lattice. A lattice may be viewed as a regular tiling of a space by a primitive cell.
Simulated Growth of Solid Tumors in Confined Heterogeneous Environment

Tumor Growth Modeling

Tumor model categories

Biological Scale

- Tissue Behavior
- Cell-cell Interactions (communities)
- Cellular Function
- Functioning and Regulation of Pathways
- Molecular Interactions
- Gene Expression
- Genetics, Genomics

Statistical Models
- Pathway Enrichment Models
- Gene Expression Models

Network-based Models
- Stoichiometric Models
- Kinetic Models
- Biochemical Reaction Networks

Tissue-level Models
- Statistical Inference Networks
- Cellular Potts Model (CPM)
- Agent-Based Models
- Continuum Models

CPM for Tumor Growth Simulation

CPM

- Model for cell sorting
- **Describes** cell-cell interaction, motion, rearrangement, pressure inside tissue
- **Suitable** for pathol. developmental mechanisms in cancer
- **Cell-based** method on the **lattice**
- 2D lattice represents tissue
- Collection of particles to represent the cell
- Each cell is represented as an object with a possible **adhesive** state, spatially extended
- cells are composed of adjacent lattice sites with similar id nr.
- system tends to minimize overall surface energy (energy per unit of area)

CPM originally developed by Graner & Glazier 1992

\[
p(\sigma_{i,j} \rightarrow \sigma_{i',j'}) = \begin{cases} 
  e^{-\frac{\Delta H}{T}} & \text{if } \Delta H > 0; \\
  1 & \text{if } \Delta H \leq 0;
\end{cases}
\]

Probability of accepting/rejecting a spin copy


Note that this is an approach with Reals – not discrete – but we are working on discrete multi-agent approaches in the future!
CPM originally developed by Graner & Glazier 1992

\[
H = J \sum_{i,j} (1 - \delta_{\sigma_i, \sigma_j} \sigma_{i', j'}) + \lambda \sum_{\sigma} (v(\sigma) - V_t(\sigma))^2
\]

Kronecker delta

\[
\delta_{ij} = \begin{cases} 
0 & \text{if } i \neq j, \\
1 & \text{if } i = j.
\end{cases}
\]

\(\sigma\) spin of a cell

\(J\) surface energies between spins (adhesion)

\(\lambda\) cellular constraint, function of elasticity

\(v(\sigma)\) area/volume of a cell

\(V_t(\sigma)\) target area for cells of type

Image of a cell sorting time series

- initial: random assigned cell types
- each step represents a growing number of Monte Carlo Step (MCS)
- figure shows pattern

Our idea:

• **Reducing** animal experiments
• **Visualizing** tumor dynamics towards better understanding
• **Easy-to-use**
• **Easy-to-extend**
• **Implementation** of Cellular Potts Model
  visualized with cytoscape.js (web application)
  other client rendering frameworks
  ... based on network visualization in biology
• **Support** biologists and clinical scientists
  ultimate goal

Jeanquartier, F., Jean-Quartier, C., Cemernek, D. & Holzinger, A. In Silico Modeling For Tumor Growth Visualization. BMC In revision.
for Number of MCS do
  for Appropriate number of samples (substeps) do
    Calculate the Hamiltonian in current state, $H_0$;
    Select a lattice site, $i$, from the domain at random;
    Select a neighbour, $j$, of this site at random;
    Change config so that site $i$ refers to same cell as site $j$ (if not ecm)
    Calculate the Hamiltonian in new configuration, $H_1$;
    if $\Delta H = H_1 - H_0 \leq 0$, then
      Accept change;
    else
      Evaluate $p = \exp((-\Delta H)/(T))$;
      Sample a number $u$ from $U(0, 1)$;
      if $p < u$, then
        Accept change;
      end
    end
  end
  If change is rejected, then restore original configuration.
end

CPM implementations already exist:

- CompuCell3d
- Tissue Simulation Toolkit

However

- though „community-driven“, not maintained
- context-specific
- static
- lack of re-usability
- hard to be combined with visualization libraries
- no web implementation
- not useful for interactive visualization

Jeanquartier, F., Jean-Quartier, C., Cemernek, D. & Holzinger, A. In Silico Modeling For Tumor Growth Visualization. BMC In revision.
Implementation Example: Tumor Growth Visualization

**Cross-browser User Interface**

- html5 & css,
- **visualization components** built with cytoscape.js, jquery.js
- ... can be customized and extended with other 3rd party plugins like flotcharts, d3.js, jit.js etc.

**BACKEND**

- **JSONCPMServlet**: for communication of frontend & backend using JSON
- **Java CPM Implementation**, consisting of **CPM Lattice Computation, Graph Converter** and **Cytoscape Converter** ...

**FRONTEND**

- Configurable parameters updated via AJAX
Implementation Example: Tumor Growth Visualization

```java
energyArea = 0;
newEnergyArea = 0;

// if cell is of type ECM then the Area calculation is suppressed
if (cell > 0) {
    energyArea = params.getLambdaArea() * Math.pow((area[cell] - getTargetAreaForCell(cell)), 2);
    //logger.debug("energyArea=" + energyArea + " (0.05 * (" + area[cell] + "+ " + getTargetAreaForCell(cell) + "+")^2 )");
}

// if cell neighbor is of type ECM then the Area calculation is suppressed
if (cellNeighbour > 0) {
    newEnergyArea = params.getLambdaArea() * Math.pow((area[cellNeighbour] - getTargetAreaForCell(cellNeighbour)), 2);
    //logger.debug("newEnergyArea=" + newEnergyArea + " (0.05 * (" + area[cellNeighbour] + "+ " + getTargetAreaForCell(cellNeighbour) + "+")^2 )");
}

/* The overall deltaH = AFTER-BEFORE
deltaH = (newEnergyAdhesion + newEnergyArea) - (energyAdhesion + energyArea);
//logger.debug("deltaH = " + deltaH + "+ energyAdhesion + " + newEnergyArea + ") - (" + energyAdhesion + "+ energyArea + ")");

// spin-copy for a temperature > 0 is accepted
// with prob = 1.0 if it would decrease the value of globally Hamiltonian
// or with Boltzmann probability if it would increase the value of Hamiltonian
if (params.getTemperature() > 0) {
    if (deltaH > 0) {
        prob = Math.exp(-deltaH / params.getTemperature()); // Boltzmann
        //logger.debug("prop=" + prob);
    } else if (deltaH <= 0) {
        prob = 1.0;
    } else if (params.getTemperature() == 0) {
        if (deltaH > 0) {
            prob = 0;
        }
```
Implementation Example: Tumor Growth Visualization

Implementation of Tumor growth visualization

- "cpm-cytoscape" already on GitHub:
  https://github.com/davcem/cpm-cytoscape

- and available as online DEMO:
  http://styx.cgv.tugraz.at:8080/cpm-cytoscape/

Jeanquartier, F., Jean-Quartier, C., Cemernek, D. & Holzinger, A. In Silico Modeling For Tumor Growth Visualization. BMC, Manuscript in rev
Nodes as Cellular bricks

→ Compartmental states:

Localized phenomena

- Intra & inter-cellular interactions
- PPI
- GO
- Disease signaling
- Absorption, Excretion, Distribution
- Modulators, Inhibitors, Promoters

Summary and Conclusion
Hot topics:

- Learning from image data for Initialization
  - image preprocessing & feature extraction
  - comparison, refinement & optimization
- ML for tumor growth profiles and model validation
- On using open tumor growth data for ML
  - Histologic data
  - Drug targeting data etc
- On multi-scale trends in cancer modelling
  - Compare results of different models
  - Link between different scales
  - Combining microscopic characteristics
  - with macroscopic parameters
- Sensitivity plots for tumor modelling
- Trajectory visualization of tumor dynamics
Thank you!
Sample Questions (1)

- What is the difference between tumor and cancer?
- What does the term differentiation in biological context stand for? Give an example.
- What means in vivo, in vitro and in silico?
- What types of computational tumor growth models exist?
- What is a cellular automaton?
In Silico Modeling for Tumor Growth Visualization

http://styx.cgv.tugraz.at:8080/cpm-cytoscape/
Appendix
Remember: Many problems in health informatics are hard

- **P**: algorithm can solve the problem in polynomial time (worst-case running-time for problem size n is less than F(n))
- **NP**: problem can be solved and any solution can be verified within polynomial time (P ⊆ NP)
- **NP-complete**: problem belongs to class NP and any other problem in NP can be reduced to this problem
- **NP-hard**: problem is at least as hard as any other problem in NP-complete but solution cannot necessarily be verified within polynomial time