

The AnyScale Learning for All (ALFA) Group



November 2016

ALFA
ANYSCALE LEARNING FOR ALL

CSAIL

STEALTH

Coevolving Tax Evasion And Audits

Fraud Anticipation and Detection in Partnership Taxation

Una-May O'Reilly
ALFA Group
CSAIL, MIT

ALFA
ANYSCALE LEARNING FOR ALL

CSAIL

Agenda

- Introduction
- Big Data era projects
 - Scalable Evolutionary computation for machine learning
 - Enabling data science
- STEALTH

ALFA
ANYSCALE LEARNING FOR ALL

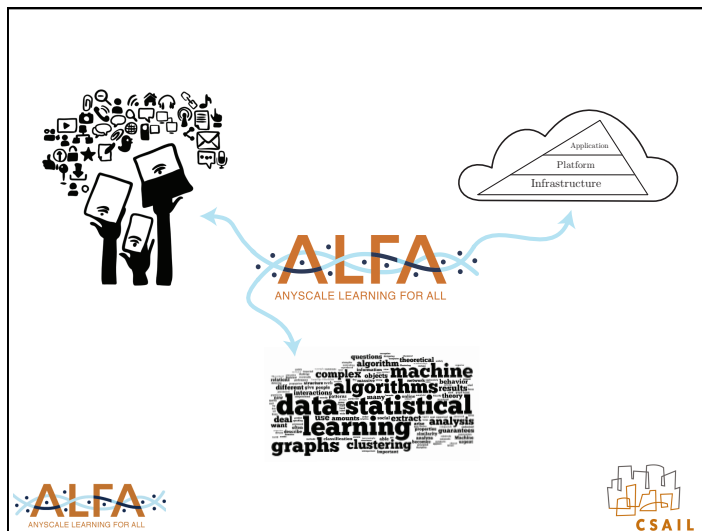
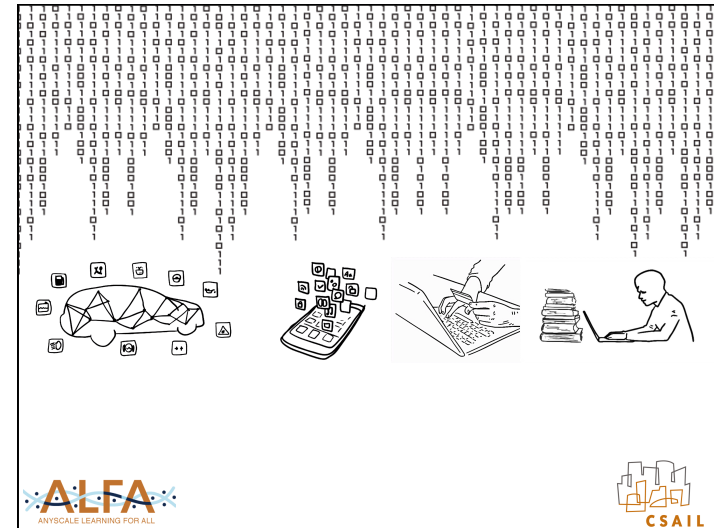
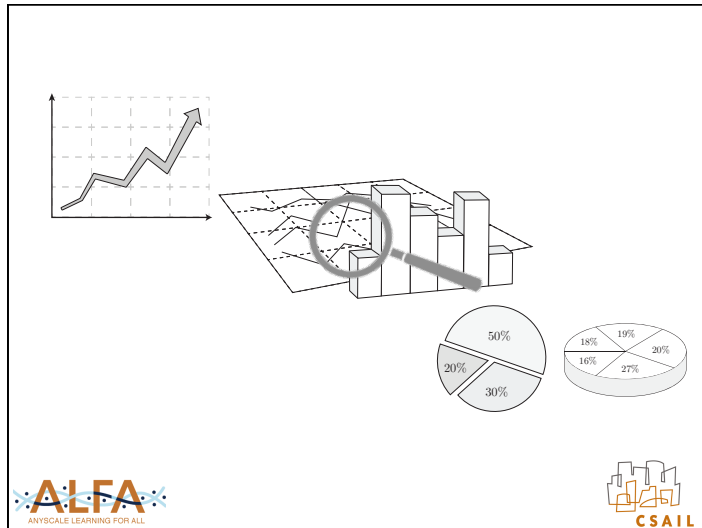
CSAIL

Agenda

- Introduction

ALFA
ANYSCALE LEARNING FOR ALL

CSAIL

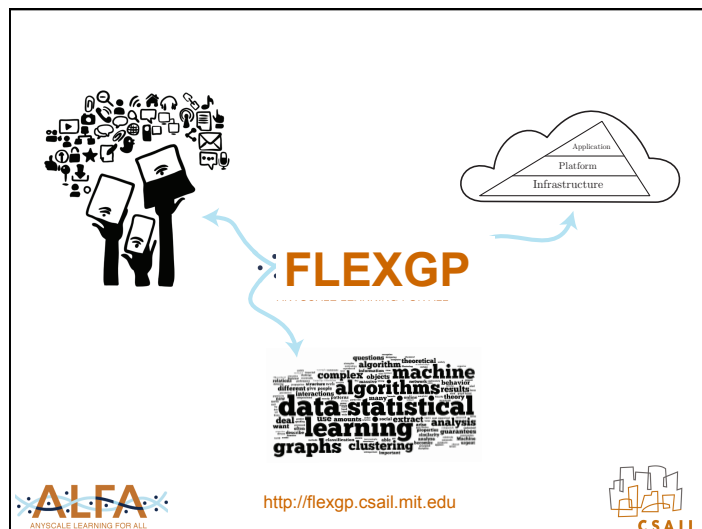
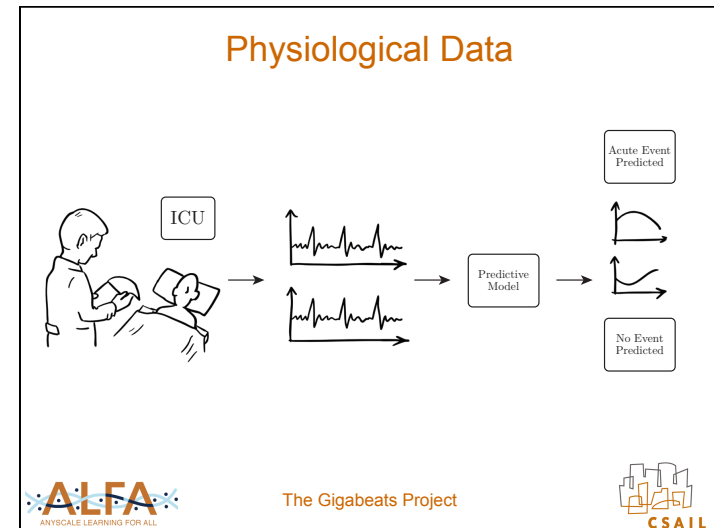
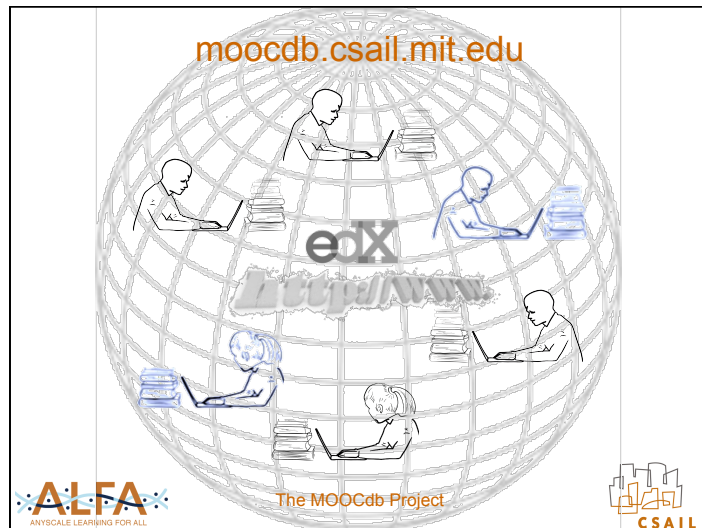


Agenda

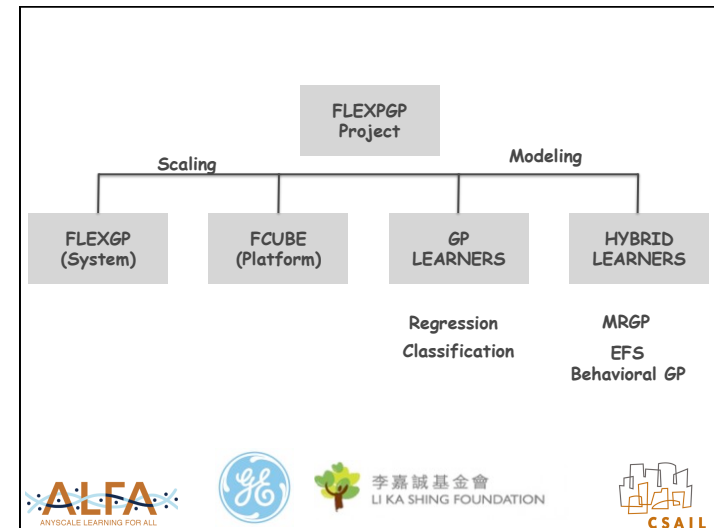
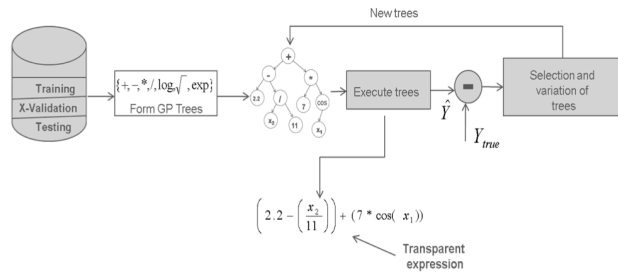
- Introduction
- Big Data era projects
 - Scalable Evolutionary computation for machine learning
 - Enabling data science

ALFA
ANYSCALE LEARNING FOR ALL

CSAIL



Genetic Programming



The FlexGP Project

In a nutshell, the FlexGP project goal is scalable machine learning using genetic programming (GP).

Genetic programming is a mature, robust multi-point search technique (inspired by evolution) which supports readable, and flexibly specified learning representations which can readily express linear or non-linear data relationships. It is well suited to parallelization and machine learning. It has a strong record in real world domains.

- Evolutionary learners:** this layer provides access to the learners so that one could run them on their desktop. See description of the learners [here](#) and a tutorial to running them on multiple examples [here](#)
- FlexGP:** a cloud based platform for generating transparent non-linear large scale regression problems
- FCUBE:** A data parallel approach to building ensemble of classifiers
- Feature learning:** Evolutionary Feature Synthesis (EFS) generates accurate, readable, nonlinear features for tabular data.

ALFA ANYSIZE LEARNING FOR ALL, CSAIL

The FlexGP Project

GP-based Learners

We provide a tutorial for the core learners developed within the frame of the FlexGP project.

Regression

- Multiple Regression Genetic Programming
- SR learner: Symbolic Regression

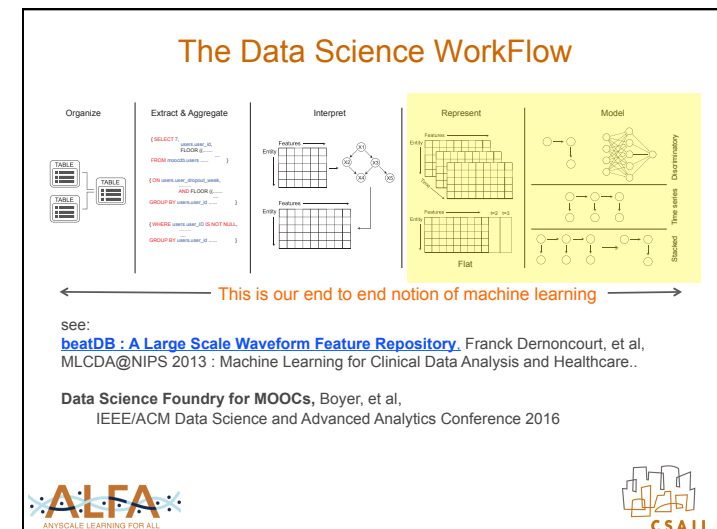
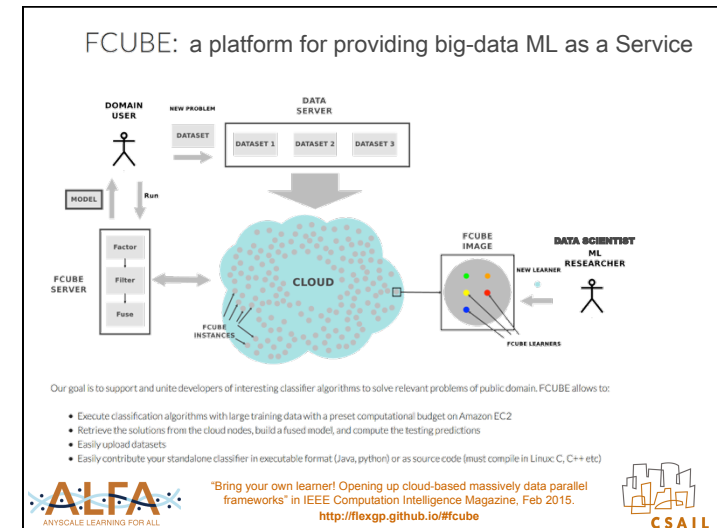
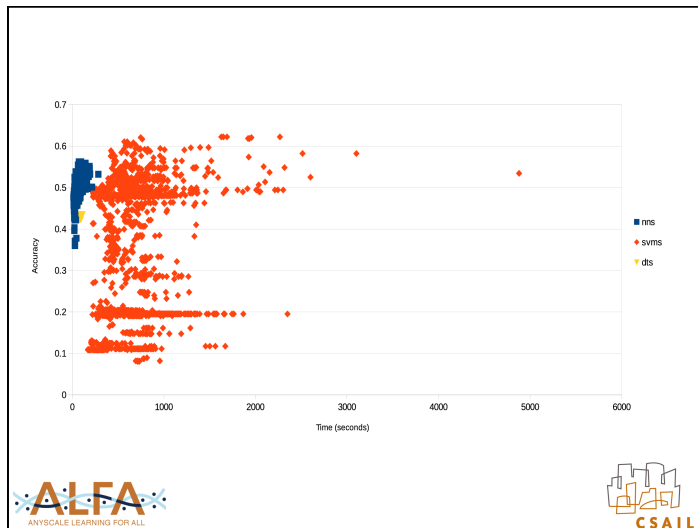
Classification

- GP function classification
- Rule Tree classification

Examples

To check examples and reports visit our blog: [FlexGP Blog](#)

ALFA ANYSIZE LEARNING FOR ALL, CSAIL



Building predictive models via feature synthesis

Ignacio Araldo
Una-May O'Reilly
Kalyan Veeramachaneni

ALFA Group, CSAIL, MIT



Project website:
<http://flexgp.github.io/efs/>



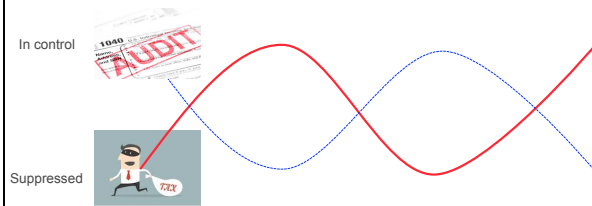
STEALTH Coevolving Tax Evasion And Audits

Fraud Anticipation and Detection in Partnership Taxation

Jacob Rosen, Erik Hemberg, Una-May O'Reilly,
Geoff Warner, Sanith Wijesinghe (MITRE Corp)



IRS Oscillatory Dynamics



- Audit dept shifts audit resources to some subset of observables
- Evaders shift schemes to sneak by where there's no attention
- Adversarial co-evolution oscillation
- \$91B tax gap from PARTNERSHIP activities
 - Some deliberate, intentional fraud



IRS = *Direcção-Geral dos Impostos in Portugal, HMRC in UK, etc*



iBOB

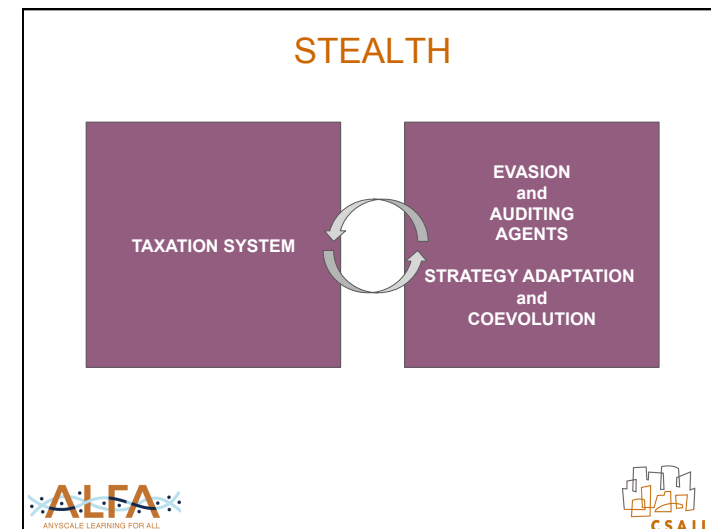
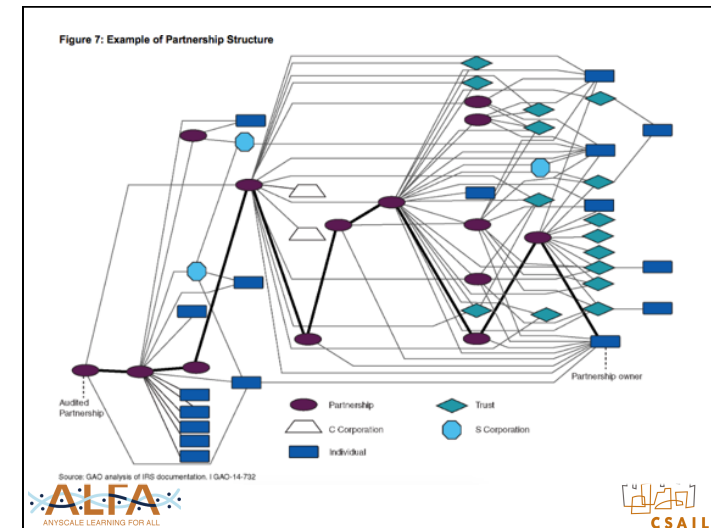
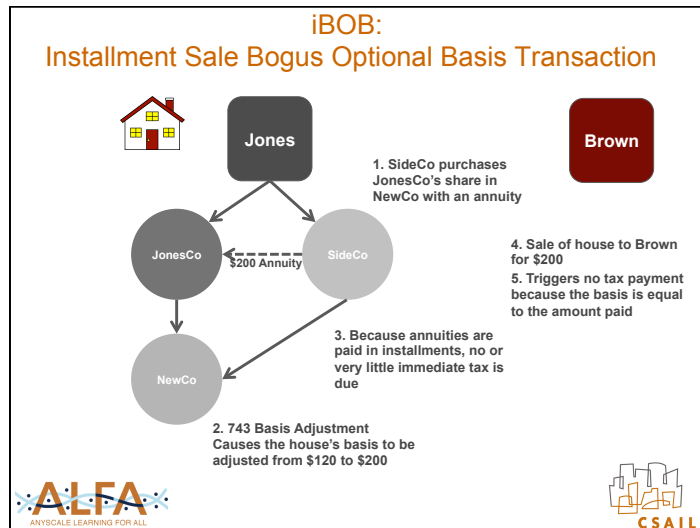
Jones wants to sell a
house he bought for
\$120 for \$200 to Brown



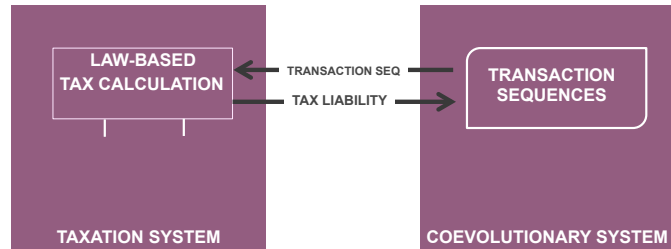
This would result in Jones
being taxed on \$80 in gain

i.e. the house has a *basis* of \$120 and
a *fair market value (FMV)* of \$200





Stealth



Representing the Tax Code

U.S. Code § 754 - Manner of electing optional adjustment to basis of partnership property

If a partnership files an election, in accordance with regulations prescribed by the Secretary, the basis of partnership property shall be adjusted, in the case of a distribution of property, in the manner provided in section 734 and, in the case of a transfer of a partnership interest, in the manner provided in section 743. Such an election shall apply with respect to all distributions of property by the partnership and to all transfers of interests in the partnership during the taxable year with respect to which such election was filed and all subsequent taxable years. Such election may be revoked by the partnership, subject to such limitations as may be provided by regulations prescribed by the Secretary.

Formalization of IRC SC K

- **Entities:** tax payer, not a flow through
 - Properties: Ordinary Income, Capital Income, Outside Basis, partnership share
- **Assets**
 - Properties: adjusted basis and Type (\$, ordinary, capital)
- **Functions**
 - $FMV(A)$,
 - $Built-in\ Income(entity, asset)$
 - $AccruedIncome(e,a) \rightarrow$ gain to e from a 's change in FMV
- **Conditional logic** that culminates in tax **position update** (basis adjustment) organized by type of **event** within a transaction:
 - Sell a partnership interest/share
 - Sell property held by a partnership
 - Distribution: liquidating or not
 - Contribution of an asset to gain entry to a partnership

Formalism to Tax Calculation

An asset is a tuple (b, τ) consisting of

- Adjusted Basis: A scalar $b \in \mathbb{R}^+$
- Type: A positive integer τ that whether the asset is category 0 (cash), category 1 (ordinary) and category 2 (capital).

An entity is a tuple $(\theta, \kappa, s, \sigma)$ consisting of

- Ordinary Income: A scalar $\theta \in \mathbb{R}$ that records ordinary taxable income for the entity
- Capital Income: A scalar $\kappa \in \mathbb{R}$ that records capital taxable income for the entity
- Share: A scalar $s \in (0, 1)$ that represents the entity's share of partnership income and liabilities.³
- Outside Basis: A scalar $\sigma \in \mathbb{R}^+$

Grammar of Partnership Transactions

```

<transactions> ::= <transactions> <transaction> | <transaction>
<transaction> ::= Transaction(<entity>, <entity>, <Asset>, <Asset>)
<entity> ::= Brown | NewCo | Jones | JonesCo | FamilyTrust
<Asset> ::= <Cash> | <Material> | <Annuity> | <PartnershipAsset>
<Cash> ::= Cash(<Cvalue>)
<Material> ::= Material(200, Hotel, 1)
<Annuity> ::= Annuity(<Avalue>, 30)
<PartnershipAsset> ::= PartnershipAsset(99, <Pname>)
<Share> ::= Share(<Sshare>)
<Cvalue> ::= 200 | 300 | 100
<Avalue> ::= 200 | 300 | 100
<Pname> ::= NewCo | JonesCo | FamilyTrust
<Sshare> ::= 30 | 50 | 20

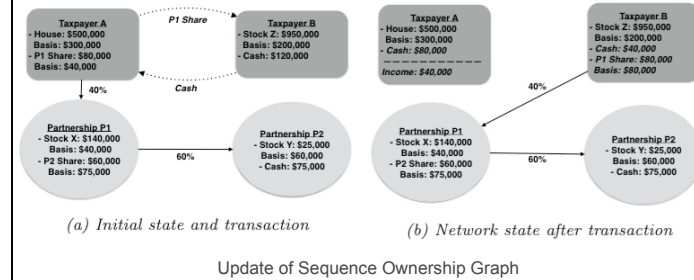
```



Allowable Sequences: Coevolutionary Module



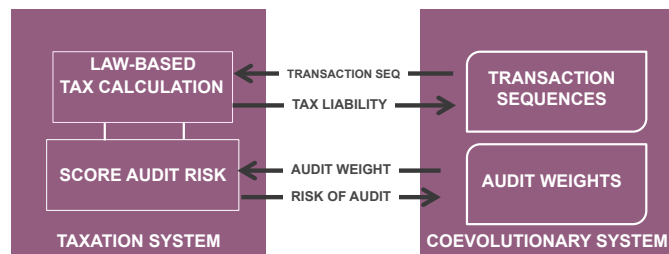
Tax Calculation of a Transaction



Transaction Examples



Stealth



Auditor Observables Examples

The basis of partnership property shall not be adjusted as the result of (1) a transfer of an interest in a partnership by sale or exchange or on the death of a partner unless (2) the election provided by §754 (relating to optional adjustment to basis of partnership property) is in effect with respect to such partnership or (3) unless the partnership has a substantial built-in loss immediately after such transfer.

743 Alteration (2004)

Observables

1. The sale of a partnership interest in exchange for a taxable asset.
2. The partnership whose shares are being transferred has not made a §754 election.
3. The seller's basis in respect to the non-cash assets owned by the partnership exceeds their FMV by more than \$250,000



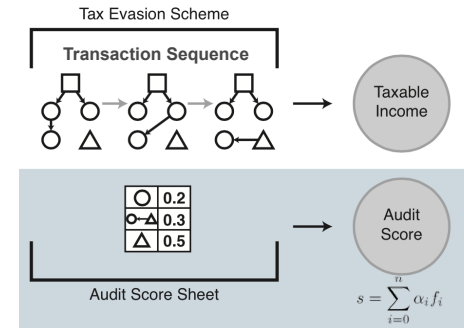
3 Auditors' Score Sheet

Observable	Independent	Random	Joint
<i>Pship Interest Sale (1)</i>	0.33	0	0
<i>No §754 Election (2)</i>	0.33	0.15	0
<i>Substantial Built-in Loss (3)</i>	0.33	0	0
(1) and (2)	0	0	0
(1) and (3)	0	0.5	0
(2) and (3)	0	0.35	0
(1) and (2) and (3)	0	0	1.0

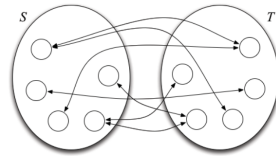
$$s = \sum_{i=0}^n \alpha_i f_i \text{ where } \sum_{i=0}^n \alpha_i = 1$$



Coupled Tax and Auditing Calculations



Two-population competitive CoEA



- ❖ Typical interaction scheme: all-to-all
- ❖ S and T co-evolve in parallel
- ❖ No transfer of individuals between S and T

July 9th, 2015

Solving Complex Problems with
Coevolutionary Algorithms

Competitive Coevolutionary Algorithm

	Evader/Adversary	Auditor/Defender
Population	Transaction sequence of entity (entity is partnership or tax payer)	Audit score sheets Set of weights for every observation
Fitness	\$ owed * average risk of audit	-\$ owed * average risk of audit
Representation	Grammar based	Real valued numbers
Interaction Scheme (with other population)	One to multiple	One to multiple
	Evolves in parallel with auditor population	Evolves in parallel with evader population



Grammatical Evolution

Map:

Grammar:

```
<transactions> ::= <transactions> <transaction> | <transaction>
<transaction> ::= Transaction(<entity>, <entity>, <Asset>, <Asset>)
<entity> ::= Brown | NewCo | Jones | JonesColFamilyTrust
<Asset> ::= <Cash> | <Material> | <Annuity> | <PartnershipAsset>
<Cash> ::= Cash(<Cvalue>)
<Material> ::= Material(200, Hotel, 1)
<Annuity> ::= Annuity(<Avalue>, 30)
<PartnershipAsset> ::= PartnershipAsset(99, <Pname>)
<Share> ::= Share(<Sshare>)
<Cvalue> ::= 200 | 300 | 100
<Avalue> ::= 200 | 300 | 100
<Pname> ::= NewCo | JonesColFamilyTrust
<Sshare> ::= 30 | 50 | 20
```

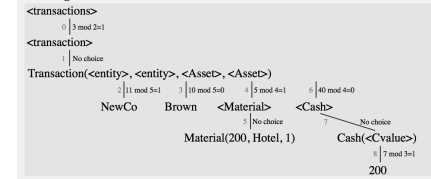
Genotype:

3, 11, 10, 5, 40, 7



Grammatical Evolution

Rewriting:



Phenotype:

Transaction(NewCo, Brown, Material(200, Hotel, 1), Cash(200))

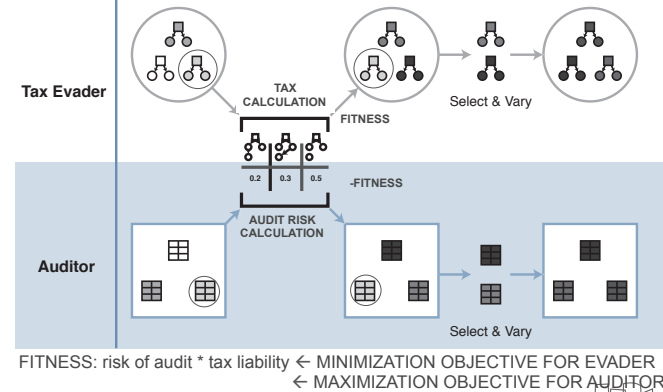


Auditor Weights

- Representation is simple now
 - Binary translation of integers
 - Normalization

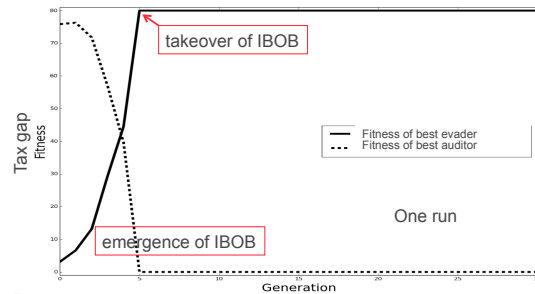


Coevolutionary Module



Demo 1: audit sheet does not have access to observables capable of detecting IBOB

Expectation: IBOB should emerge in the evader population and take hold permanently because it never gets audited

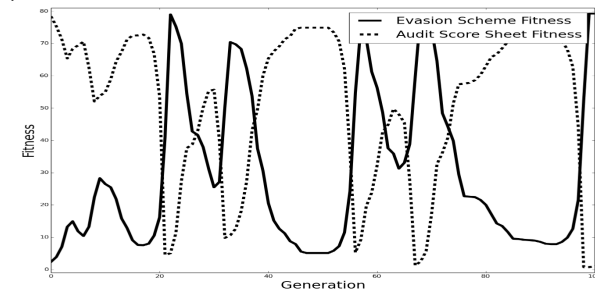


ALFA
ANYSCALE LEARNING FOR ALL

CSAIL

Demonstration #2

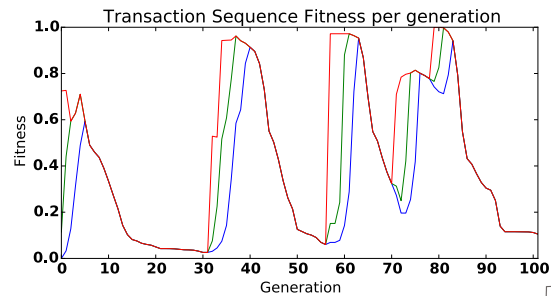
Model case of limited audit resources
Observables exist for IBOB and other sequences, but audit can only afford to weight/examine three of four observables



ALFA
ANYSCALE LEARNING FOR ALL

CSAIL

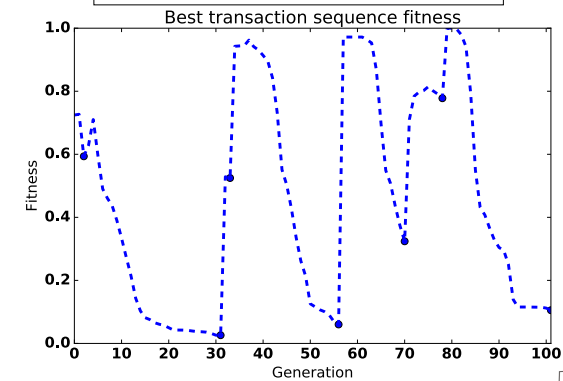
— Mean Transaction Sequence Fitness
— Top Layer Transaction Sequence Fitness
— Best Transaction Sequence Fitness



ALFA
ANYSCALE LEARNING FOR ALL

CSAIL

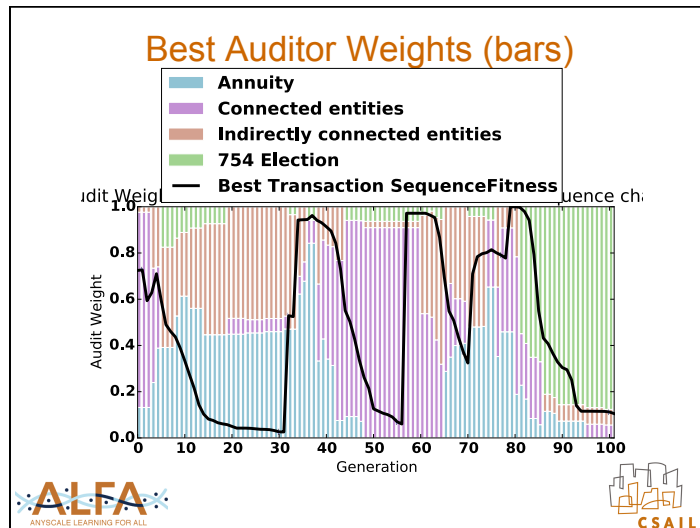
--- Best Transaction Sequence Fitness



ALFA
ANYSCALE LEARNING FOR ALL

UM_ndist_best_transaction_sequence_fitness_line.pdf

CSAIL



Publications, More Details

- **Website**
 - <http://groups.csail.mit.edu/ALFA/STEALTH/>
- **Pubs:**
 - [Computer Aided Tax Avoidance Policy Analysis](#) , Jacob Rosen, S.M Thesis MIT Dept of EDS, *Awarded Best Thesis in the Technology and Policy Program, 2015*
 - Computer Aided Tax Evasion Policy Analysis: Directed Search using Autonomous Agents, to appear in Shadow 2015 , Munster
 - Tax non-Compliance Detection Using Co-Evolution of Tax Evasion Risk and Audit Likelihood, International Conference on AI and Law, 2015, San Diego