The AnyScale Learning for All (ALFA) Group

November 2016

STEALTH
Coevolving Tax Evasion And Audits
Fraud Anticipation and Detection in Partnership Taxation

Una-May O'Reilly
ALFA Group
CSAIL, MIT

Agenda

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

Agenda

- Introduction
Adversarial Dynamics
Agenda

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

- username: Jane
- ip: 128.21.221.13
- agent: Mozilla/5.0 ...
- page: https://mitx.mit.edu/...
- time: 11/25/2015:5:40PM

- username: John
- ip: 164.28.233.15
- agent: Mozilla/5.0 ...
- page: https://mitx.mit.edu/...
- time: 09/12/2015:9:32AM

- username: Smith
- ip: 168.61.221.13
- agent: Mozilla/5.0 ...
- page: https://mitx.mit.edu/...
- time: 12/11/2015:10:03AM

------

Physiological Data

ICU

Predicted Model

No Event Predicted

FLEXGP

http://flexgp.csail.mit.edu

The Gigabeats Project

Ignacio Arnaldo

Kalyan Veeramachaneni

Una-May O'Reilly

Owen Derby

Krzysztof Krawiec
Genetic Programming

The FlexGP Project

GP-based Learners

- **Evolutionary learners**: This layer provides access to the learners so that one could run them on their devices. See description of the learners here and a tutorial to running them on cloud here.
- **FlexGP**: A tool-based platform for generating scalable non-linear models for regression problems.
- **EFS**: A data-parallel approach to building ensembles of classifiers.
- **Hybrid Learning**: Evolutionary Fuzzy Synthesis (EFS) generates accurate, readable, nonlinear features for tabular data.
FlexGP Blog

In this blog we will provide examples of use of the learners developed within the "FlexGP" project. We will analyze the performance of the released learners for different datasets.

1. RR learner
2. RF Tree classification
3. GP Function classification
4. Multiple Regression Genetic Programming

Symbolic Regression Learner: Predicting the quality of wine

The Wine Quality dataset is available at the UC Machine Learning repository website. This problem consists in modeling the quality (a grade from 1 to 10) of a given red or white wine given 11 features such as acidity, alcohol degree etc. Note that the first line of both datasets contains the labels of the different features and needs to be deleted. Additionally, the separation employed in the original dataset

FlexGP: Cloud-Based Ensemble Learning with Genetic Programming for Large Regression Problems

Arnaldo, Veeramachaneni, O’Reilly, J. Grid Computing, 2014

Multiple Regression GP

Multiple Regression Genetic Programming, Arnaldo, Krawiec, O’Reilly (GECCO 2014)

http://flexgp.github.io

http://flexgp.github.io
flexgp.csail.mit.edu
Machine Learning often tacitly starts from the training matrix

The Data Science Workflow

This is our end to end notion of machine learning

See:
- beatDB: A Large Scale Waveform Feature Repository, Franck Dernoncourt, et al, MLCUDA@NIPS 2013 - Machine Learning for Clinical Data Analysis and Healthcare...
- Data Science Foundry for MOOCs, Boyer, et al, IEEE/ACM Data Science and Advanced Analytics Conference 2016

The Data Science WorkFlow
- Organize
- Extract & Aggregate
- Interpret
- Model

This is our end to end notion of machine learning

The Data Science Machine: Emulating Human Intelligence in Data Science Endeavors, James Max Kanter, MIT Dept of EECS, 2015.
Building predictive models via feature synthesis

Ignacio Arnaldo
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

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

This would result in Jones being taxed on $80 in gain
iBOB: Installment Sale Bogus Optional Basis Transaction

1. SideCo purchases JonesCo's share in NewCo with an annuity
2. 743 Basis Adjustment Causes the house’s basis to be adjusted from $120 to $200
3. Because annuities are paid in installments, no or very little immediate tax is due
4. Sale of house to Brown for $200
5. Triggers no tax payment because the basis is equal to the amount paid

Project Goals

• One project goal is to abstractly replicate the fundamental oscillatory dynamics of evaders and auditors
• Understand how to model adversarial relationships as continuously adapting with some ultimate goal of anticipation and insight for government and legal stakeholders
• Technical research investigations
  – How we simulate co-evolutionary dynamics?
  – Transfer policy/law/accounting into a software system
  – Express rich behavioral strategy of evader
    » Infinite repertoire available
  – How to represent the IRS detection side?
Stealth

LAW-BASED TAX CALCULATION

TRANSACTION SEQ

TAX LIABILITY

TRANSACTION SEQUENCES

COEVOLUTIONARY SYSTEM

12/12/16

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 (S, ordinary, capital)
• Functions
  – FMV(A),
  – Built-in Income(entity, asset)
  – AccruedIncome(e,a) -> 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

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.

Formalism to Tax Calculation

An asset is a tuple \((s, \tau)\) consisting of

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

An entity is a tuple \((b, c, s, \tau)\) consisting of

• Ordinary Income: A scalar \(b \in \mathbb{R}\) that records ordinary taxable income for the entity
• Capital Income: A scalar \(c \in \mathbb{R}\) 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.
• Outside Basis: A scalar \(a \in \mathbb{R}^+\)
Grammar of Partnership Transactions

<transactions>::=<transactions><transaction> | <transaction>
<transaction>::=Transaction(<entity>,<entity>,<Asset>,<Asset>)
<Asset>::=Cash|Material|Amortity|PartnershipAsset
<Cash>::=Cash(<value>)
<Material>::=Material(200,Hotel,1)
<Amortity>::=Amortity(<value>,30)
<PartnershipAsset>::=PartnershipAsset(Wo,<Pname>)
<Share>::=Share(<Share>)
<Value>::=200|300|100
<Pvalue>::=200|300|100
<Pname>::=NewCo|JonesCo|FamilyTrust
<Share>::=30150|20

Allowable Sequences: Coevolutionary Module

Tax Calculation of a Transaction

(a) Initial state and transaction
Update of Sequence Ownership Graph

(b) Network state after transaction

Transaction Examples

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

<table>
<thead>
<tr>
<th>Observable</th>
<th>Independent</th>
<th>Random</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partnership Sale (1)</td>
<td>0.33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No 1754 Election (2)</td>
<td>0.33</td>
<td>0.15</td>
<td>0</td>
</tr>
<tr>
<td>Substantial Built-in Loss (3)</td>
<td>0.33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(1) and (2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(1) and (3)</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>(2) and (3)</td>
<td>0</td>
<td>0.35</td>
<td>0</td>
</tr>
<tr>
<td>(1) and (2) and (3)</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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

Coupled Tax and Auditing Calculations

**Tax Evasion Scheme**

Transaction Sequence

- **Taxable Income**

- **Audit Score**

- **Audit Score Sheet**

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

Competitive Coevolutionary Algorithm

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

Two-population competitive CoEA

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

Map
Grammar:

<transaction> ::= <transactions> <transaction> | <transaction>
<transactions> ::= <Transaction> <entity> <entity> <Asset> <Asset>
<entity> ::= <BusinessName> <Transaction> <entity> <entity> <Asset> <Asset>
<Asset> ::= <Cash> <Material> <Amorty> <PartnershipAsset>
<Cash> ::= <Cash> (<Value>)
<Material> ::= <Material> (200, Hotel, 1)
<Amorty> ::= <Amorty> (<Value>, 30)
<PartnershipAsset> ::= <PartnershipAsset> (99, <Phones>)
<Shares> ::= <Shares> (<Shares>)
<Value> ::= 2003000100
<Phones> ::= 2003000100
<Phones> ::= NewCoName, CoFamilyTrust
<Shares> ::= 305020

Auditor Weights

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

Coevolutionary Module

FINDING:

<transaction> ::= <transactions> | <transaction>
<transactions> ::= <Transaction> <entity> <entity> <Asset> <Asset>
<entity> ::= <BusinessName> <Transaction> <entity> <entity> <Asset> <Asset>
<Asset> ::= <Cash> <Material> <Amorty> <PartnershipAsset>
<Cash> ::= <Cash> (<Value>)
<Material> ::= <Material> (200, Hotel, 1)
<Amorty> ::= <Amorty> (<Value>, 30)
<PartnershipAsset> ::= <PartnershipAsset> (99, <Phones>)
<Shares> ::= <Shares> (<Shares>)
<Value> ::= 2003000100
<Phones> ::= 2003000100
<Phones> ::= NewCoName, CoFamilyTrust
<Shares> ::= 305020

Tax Evader

FITNESS: risk of audit * tax liability

MINIMIZATION OBJECTIVE FOR EVADER

Maximization Objective for Auditor

Coevolutionary Module

Tax Evader

Select & Vary

FITNESS: risk of audit * tax liability

MINIMIZATION OBJECTIVE FOR EVADER

Maximization Objective for Auditor

Audit Risk Calculation

Select & Vary
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.

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.
Best Auditor Weights (bars)

Publications, More Details

- Website
  - http://groups.csail.mit.edu/ALFA/STEALTH/

- Pubs: