STAR CANADA

A TECHWELL EVENT

W14

AI and Data Analytics
Wednesday, October 17th, 2018 3:00 PM

Improve Testing of AI Systems with "Grey-Box" Testing Technique

Presented by:

Yury Makedonov

Brought to you by:

350 Corporate Way, Suite 400, Orange Park, FL 32073
888-268-8770 · 904-278-0524 - info@techwell.com - http://www.starwest.techwell.com/

Yury Makedonov

Yury Makedonov was trained as a researcher and worked in a research and development institution dealing with composite materials. He has a Ph.D. degree in physics and math, though he is not a rocket scientist anymore; now he is using his skills and knowledge to improve software quality. Yury has more than twenty years of testing experience, from small startups to large companies and government organizations, and recently has been working as a QA manager, test manager, and consultant.

Improving Testing of AI Systems with Grey-box Testing Technique

Yury Makedonov
Accenture

Phone: (647) 342-6675
E-mail: yury.makedonov@gmail.com

Outline

1. Introduction
2. Business case for Artificial Intelligence
3. What is Artificial Intelligence
4. How a machine learns and how to test it
5. What's common and what's unique to AI testing
6. Demo of simple AI/Machine learning system
7. Conclusion
8. Q & A

Target audience

- This workshop is for everybody who is involved in testing of AI systems:
  - Test Managers
  - Testers
  - Developers
  - System support

When not to use AI

- Financial reporting in an ERP system
- Actuarial system to calculate pension contributions
- Time of use electricity billing system
- Government mandated employment equity reporting
- Any application based on:
  - government regulation
  - contractual mandate
  - laws of physics
  - etc.

When to use AI – theory

- When exact calculations are theoretically possible but practically not feasible
- When exact calculations are possible but are unimportant
- When exact calculations are not possible due to the nature of the task
- When you do not have any government etc. mandated algorithms

---

When to use AI – some examples

- Personalized offers from a telecom company
- Online ads network >$250B in 2018
- Social media feeds
- Other marketing voodoo stuff
- Video games
- Fraud prevention at banks
- Smart assistants with voice recognition
  Siri, Alexa, Cortana etc.

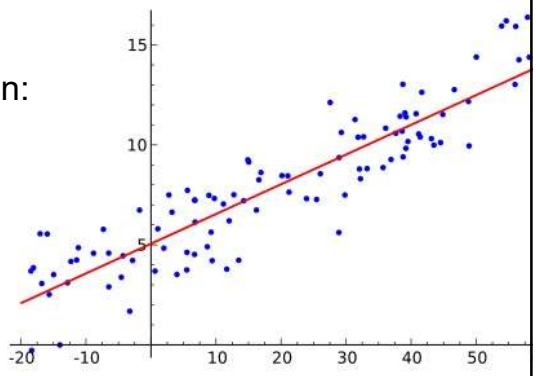
What is AI

- Artificial Intelligence doesn't exist!
- Artificial Intelligence = Machine Learning
- Lifetime Cycle of a Machine Learning system:
  - Build
  - Learn
  - Use
- Machine Learning:
  - a set of formulas with fitting coefficients
  - exact values of fitting coefficients are not known during the development phase
  - These values are adjusted during the "learning" phase

Machine Learning Algorithm Example #1

Simple linear regression:

$$y = a + bx$$



here:

y – predicted value

a, b – fitting coefficients

x – input variable a.k.a. predictor

Machine Learning Algorithm Example #2

More complex machine learning algorithm:

$$y = \sum_{i=1}^{m} a_i f(x_1 \dots x_n)$$

here:

y – predicted value

a_i – fitting coefficients

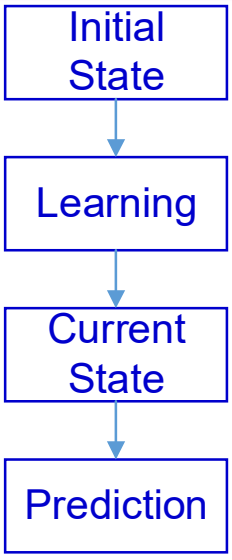
$x_1 \dots x_n$ – input variables a.k.a. predictors

How to test Machine Learning - theory

$$y = \sum_{i=1}^{m} a_i f(x_1 \dots x_n)$$

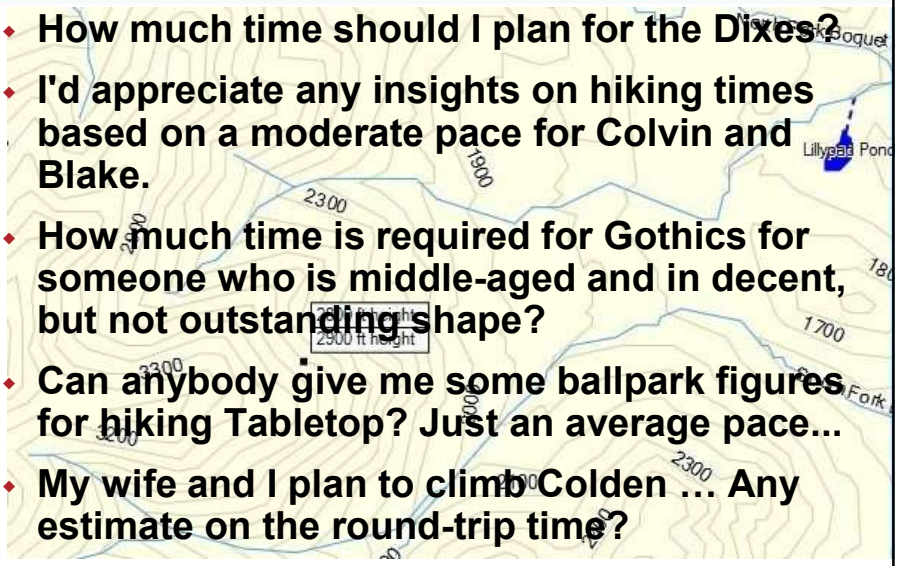
1. Classic GUI testing
2. Algorithm testing – most important
3. How to make AI testing predictable and reproducible:

   Control of the state of a system:
   - backup
   - restore
   - know all values
   - force a predefines state



Demo – Predicting Adirondack hiking times

- **How much time should I plan for the Dixes?**
- **I'd appreciate any insights on hiking times based on a moderate pace for Colvin and Blake.**
- **How much time is required for Gothics for someone who is middle-aged and in decent, but not outstanding shape?**
- **Can anybody give me some ballpark figures for hiking Tabletop? Just an average pace...**
- **My wife and I plan to climb Colden … Any estimate on the round-trip time?**



Demo – Predicting Adirondack hiking times

- Hiking time formula:

$$t = L / v$$

here: t – predicted time

L – distance

v – speed

For flat or very gradual terrain, the calculation of 30 minutes for every mile is used.

Demo – Predicting Adirondack hiking times

- Naismith's Rule (1892):
  - Allow one hour for every 3 miles (5 km) forward,
  - plus an additional hour for every 2,000 feet (600 m) of ascent.

$$t = L / v_{hor} + H / v_{vert}$$

here: t – predicted time

L – distance

v_{hor} – horizontal speed

H – total elevation gain

v_{hor} – vertical speed

Demo – Predicting Adirondack hiking times

- People have different physical capabilities:

Hike/ hiker:	Street & Nye	Whiteface & Esther	Allen
Joe	3:35	4:29	6:57
Yury	8:20	8:37	12:26
Gerard	13:35	12:49	n/a

- Calculation algorithm should be unique:
  - for each person
  - for each geographic area
- Let's use AI to predict individual hiking times!

Demo – Linear Regression 1

$$t = L\,p$$

here: t – predicted time

L – distance

p – pace (min/km of distance)

Starting values:

p_{hor} = 0.3 hours/km of distance

Demo – Least Squares Method

MS EXCEL >> File >> Options >> Add-ins >> Solver

Least Squares Method:

- trying to find the minimum value of a
- sum of squared differences between actual and predicted values
- by changing the values of fitting coefficients

$$S = \sum_{i=1}^{n} (Predicted_i - Actual_i)^2$$

Demo – Linear Regression 2

$$t = L\,p_{hor} + H\,p_{vert}$$

here: t – predicted time

L – distance

p_{hor} – horizontal pace (min/km of distance)

H – total elevation gain

p_{vert} – vertical pace (min/km of altitude gain)

Starting values:

p_{hor} = 0.3 hours/km of distance

p_{vert} = 2 hours/km of altitude gain

Demo – Linear Regression 3

$$t = L\,p_{hor} + H\,p_{vert} + L^2\,p_3 + H^2\,p_4$$

here: p_{hor} – horizontal pace (min/km of distance)

p_{vert} – vertical pace (min/km of altitude gain)

p_3, p_4 – 2nd order fitting coefficients

Starting values:

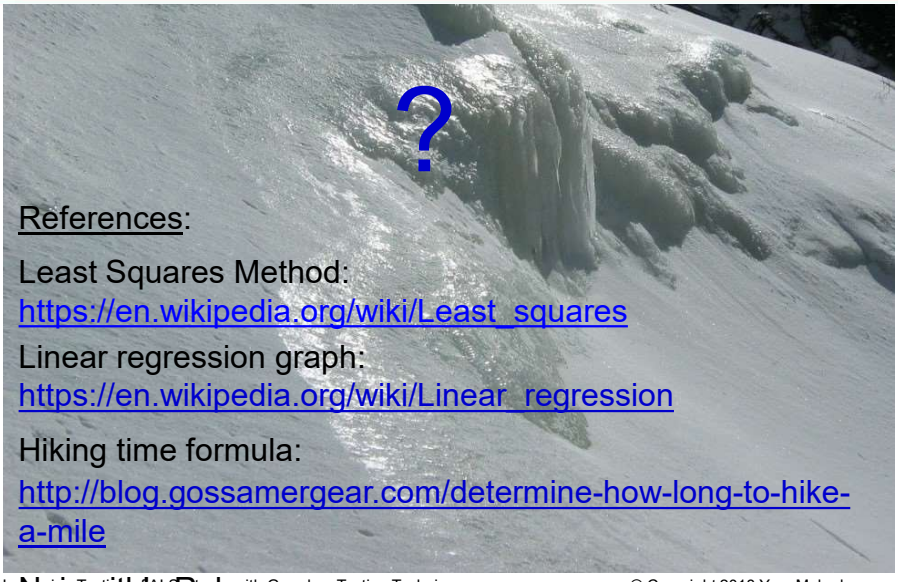
p_{hor} = 0.3 hours/km of distance

p_{vert} = 2 hours/km of altitude gain

p_3 = -0.0001

p_4 = -0.000001

Q & A



References:

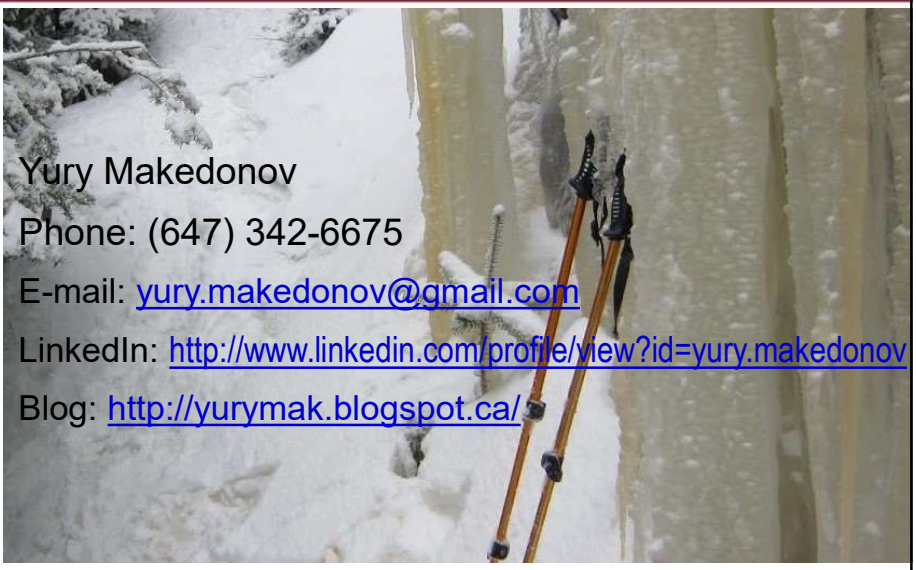
Least Squares Method:
https://en.wikipedia.org/wiki/Least_squares

Linear regression graph:
https://en.wikipedia.org/wiki/Linear_regression

Hiking time formula:
http://blog.gossamergear.com/determine-how-long-to-hike-a-mile

Naismith's Rule:

Contact information



Yury Makedonov

Phone: (647) 342-6675

E-mail: yury.makedonov@gmail.com

LinkedIn: http://www.linkedin.com/profile/view?id=yury.makedonov

Blog: http://yurymak.blogspot.ca/

