Using AI to Empower System Development and Operations Team

Wahab Hamou-Lhadj, PhD
ECE, Concordia University
wahab.hamou-lhadj@concordia.ca

TNSBC, Montréal, QC
May 23, 2018
Software-intensive systems are everywhere!

- Health
- Energy
- Finance
- Manufacturing
- Education
- Public safety
- Telecom
- Aerospace
- Entertainment
- Hospitality
- Public administration
- Social interactions

“Our civilization runs on software”

B. Stroustrup
Facts

- From 1997 to 2012, software industry production grew from $149 billion to $425 billion.
- The software industry’s direct share of U.S. GDP went from 1.7% to 2.6%.
- Software accounted for 12.1% percent of all U.S. labor productivity gains from 1995 to 2004 and 15.4% from 2004 to 2012.

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)
Software Development

A set of activities for creating a software system including requirements analysis, architectural design, detailed design, coding, testing, maintenance, integration, acceptance testing, etc.

Software Operations

A set of activities for supporting end users of a software product in an operational environment. Typical activities include: installation, upgrade, monitoring, configuration, etc.

Software Development

A set of activities for creating a software system including requirements analysis, architectural design, detailed design, coding, testing, maintenance, integration, acceptance testing, etc.


Software Operations

A set of activities for supporting end users of a software product in an operational environment. Typical activities include: installation, upgrade, monitoring, configuration, etc.


Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)
SW Development Challenges

- Increased complexity
- High cost
- Heavy reliance on people
- Lack of automated tools
- Time to market pressure
- Maintaining quality
A NIST study* shows that defects in software cost the U.S. economy $56 billion annually.

A large percentage of software development costs are spent on identifying and correcting defects.

There is a need to invest in automated and intelligent solutions.

Active Research Community

- Change and defect management
- Continuous integration/deployment
- Release engineering
- Reverse engineering and re-engineering
- Run-time evolution and dynamic configuration
- Software and system comprehension
- Software migration and renovation
- Software refactoring and restructuring
- Empirical studies
- Evolution of non-code artefacts
- Human aspects of software maintenance and evolution
- Maintenance and evolution of model-based methods
- Maintenance and evolution processes
- Maintenance and evolution of mobile apps
- Maintenance versus release process
- Mining software repositories
- Etc.
Emergence of Software Analytics

- Data-driven software development and maintenance
- Big Data: source code, bug reports, test cases, logs, user feedback, etc.
- Predictive analytics using ML, DL, CI, and PR
- Information visualization of large-scale data
The Commit Assistant Project

- An NSERC project in collaboration with Ubisoft.
- Goal: To empower SW developers with an intelligent tool that detects defects as they write code, and proposes fixes.
Open Technologies behind CommitAssistant

**BIANCA**
Preventing Bug Insertion at Commit-Time Using Clone Detection

**CLEVER**
Combining Levels of Bug Prevention and Resolution Techniques

Bug Metarepository Search Engine for Developers and Researchers
Where do developers look for information when facing an unknown bug/crash?

89 participants
What do developers search for when facing an unknown bug/crash?

89 participants
Examples of Coding Websites

Learn, Share, Build

Each month, over 50 million developers come to learn, share their knowledge, and build their careers.

Join the world’s largest developer community.
Examples of Coding Websites

- Contain answers to coding questions
- Answers are based on developers’ experience
- Answers are in natural language
- Examples of code are inserted into text
- Useful, but not necessarily usable

Dr. Wahab Hamou-Lhadj  
(wahab.hamou-lhadj@concordia.ca)
BUMPER: Bug Metarepository Search Engine for Developers and Researchers

- Aggregates information from many bug report and code versioning systems
- Is an online search engine to millions of bug reports and fixes from open-source repositories
- Uses a query system for developers and advanced API for researchers
- Leverages the concept of collective coding → collective intelligence
Developers can search millions of lines of code and bug reports for a bug or crash they encountered.
User query

Bug reports where the same bug occurred

Fragments of code where the same bug was fixed
BIANCA: Preventing Bug Insertion at Commit-Time Using Clone Detection

- BIANCA learns known defects by mining BUMPER-indexed systems.
- It intercepts developer’s code and compares it to signatures of known defects.
- If a match exists, a flag is raised and a fix is proposed.
Subject systems: 42 open source projects
Precision = 90% and Recall: 37%
BIANCA fixes are accurate in 79% of the cases
CLEVER: Combining Levels of Bug Prevention and Resolution Techniques

- Combines multiple features to determine the defect signatures
- Uses domain expertise to create clusters of projects for improved accuracy
- Uses better code matching techniques
- Is evaluated on 12 Ubisoft systems
CLEVER Project Clustering

We can improve the detection accuracy if we search within inter-related projects
We can improve the detection accuracy if we search within inter-related projects
Evaluation of CLEVER at Ubisoft

- **Results:**
  - Subject systems: 12 Ubisoft systems
  - Precision = 79% and Recall = 65%
  - CLEVER recommends fixes in 67% of the cases

- **Impact on productivity:**
  - CommitAssistant (internal implementation of CLEVER) is designed to integrate well with developers’ workflow
  - Ubisoft announced in a press release that CommitAssistant can cut the bug fixing time by 20%
**Software Development**

A set of activities for creating a software system including requirements analysis, architectural design, detailed design, coding, testing, maintenance, integration, acceptance testing, etc.

**Software Operations**

A set of activities for supporting end users of a software product in an operational environment. Typical activities include: installation, upgrade, monitoring, configuration, etc.

Software Operations

- D2K Project: From Data To Knowledge for Better System Maintenance
- Collaborators: Ericsson, NSERC, and MITACS
- Objectives:
  - Improve the crash report (CR) handling process
  - Investigate automated solutions
  - Provide analysis capabilities to operators
  - Provide data governance solutions
Problem

- A large number of CRs
- Meeting contractual WLAs
- Severe problems have to be fixed quickly
- Cost is also an issue
Solution

Crash Reports

1st line support

2nd line support

nth line support

Operators/Analysts

Log data

Crash report

D2K Database

- ML model of CRs
- Domain knowledge

- ReCRAC: CR Duplicate Detection Tool
- TotalADS: Anomaly Detection System

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)
ReCRAC

Training Phase

Historical CRs → Preprocessing → Feature Extraction → Model Construction → D2K

Data mining
Natural Language Processing
CR description
Crash traces fingerprints

Machine learning /Deep Learning
Model combination

Testing Phase (ReCRAC in operation)

New CR → Preprocessing → Feature Extraction → Correlation → Similar CRs

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)
TotalADS: Total Anomaly Detection System

- Developed in an NSERC project with Defence R&D Canada and Ericsson

- Objectives:
  - Detection of abnormal behavior in computer hosts through the analysis of machine data
  - Combination of multiple machine learning techniques
  - Leverage of data abstraction, model combination, adaptive learning, and online learning
  - Tool development and integration
TotalADS: Total Anomaly Detection System Architecture

Software-Intensive Infrastructures

Machine Data Management
- Controller
- Loading
- Streaming

Machine Learning
- SVM
- NN
- KSM
- RNN
- Others

Control Center
- Plots
- Controllers
- Statistics
- Reports
- Analysis

Data Centers
Radio Stations
Smart Grids
IoT Devices

IBM CASCON 2014
PEOPLE’S CHOICE AWARD

Dr. Wahab Hamou-Lhadj (wahab.hamou-lhadj@concordia.ca)
TotalADS and Deployable System Units (DSUs)

The 4\textsuperscript{th} DSU prototype (PoC1)

Six Jetway industrial Mini-ITX computers + one manageable GB switch + six 4-TB hard disks (Intel's Haswell Core i7-4770TE 2.3 GHz processor, 8 GiB DDR3, 6 GB/sec mSATA, dual LAN) (The whole DSU needs less than 350 watts when used at full capacity)

(Next technology to be considered: the new NVIDIA Jetson TX-2 AI computing board)

Some thoughts on the use of DL/ML in SW Development and Operations
Powerful tool suite
Healthy analytics
Context matters
Domain expertise
Education
Impact on society
THANK YOU!