

HOW TO MARKET YOURSELF & YOUR RESEARCH

2020.05.25

SHIN Yoo

(MOST OF THE MATERIAL ORIGINALLY DEVELOPED BY PROF. JUHO KIM)



ASSIGNMENT #6: WRITING A RESUME (DUE 8TH JUNE)

- X Write a customized resume for a specific job posting. Job postings will be provided. Selected resumes will be discussed during the resume workshop. We welcome volunteers.
- X Assume that you're writing your resume for a specific job posting:
 - Undergraduates:
https://nmsl.kaist.ac.kr/courses/i2r/resources/undergradute_research_internship_MSR.pdf
 - Graduate students: pick one from <https://www.microsoft.com/en-us/research/careers/>

FULL CAVEAT

- X This lecture was originally done by Prof. Juho Kim: I took it over for 2020
- X A paper deadline on 9pm Saturday messed up my brain and...
- X I spent Sunday night preparing ANOTHER lecture (which you will see on Wednesday) 🤔
- X Which is why this is mostly Prof. Kim's material (with a few drops of my anecdotes)



WHY DO I NEED TO MARKET MYSELF AND MY RESEARCH?

- X “Do good work & let people know about it.” (David Karger on [Quora](#))
- X Self-promotion does matter.
 - Affects jobs, grants, awards, citations, ...
- X Helps you revise research vision & reassess impact.
- X Broadens research impact & audience.
- X Norms & culture vary between areas/communities.
- X Nothing replaces good research.

PROMOTING YOUR RESEARCH



"MARKETING" CHANNELS

- | | |
|-----------------------------|-----------------------------|
| X Paper | X Media exposure |
| X Grant proposal | X Productization / startup |
| X CV | X Elevator pitch |
| X Talk | X Networking |
| X Demo, poster | X Teaching |
| X Video | X Academic service |
| X Open data, code, material | X Job market |
| X Website | X Letters of recommendation |
| X Social media | X Collaboration |



- X “The curriculum vitae, also known as a CV or vita, is a comprehensive statement of your educational background, teaching, and research experience. It is the standard representation of credentials within academia.”
- X CV vs Résumé
 - CV: can be long, academic purposes, focus on academic accomplishments
 - Résumé: 1 or 2 pages, industry jobs, focus on skills
- X Many researchers keep their CV up-to-date and make it publicly available, even if they are not looking for jobs.



CV: WHAT TO INCLUDE?

X Required components

- Name, contact information
- Education
- Honors & Awards
- Research interests
- Experience (positions held)
- Publications
- Teaching
- Talks
- Services

X Optional components

- Grants
- Patents
- Students
- Press
- Languages
- Skills
- References



CV: WHAT NOT TO INCLUDE?

- X Anything personal beyond contact info
 - Photo
 - Age, date of birth
 - Marital status
- X Long descriptions of projects and papers



CV: EXAMPLES

X There are tons of available examples online.

X Examples

- [Minsuk Chang](#) (Ph.D. student)
- [Prof. Kim](#) (junior faculty)
- [Me](#) (somewhat older faculty)
- [Scott Klemmer](#)
- [Geoffrey Hinton](#)

X Resources

- https://www.oise.utoronto.ca/orss/UserFiles/File/Creating_your_academic_cv_handout_1.pdf
- <https://grad.illinois.edu/sites/default/files/PDFs/CVsamples.pdf>



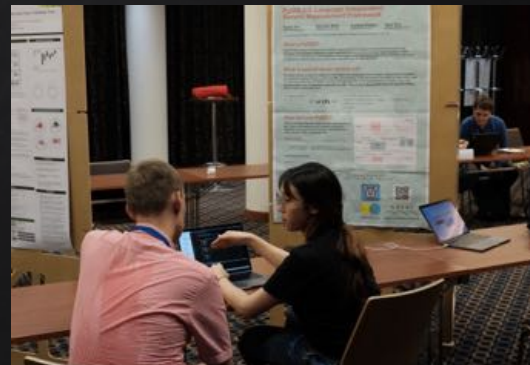
DEMO & POSTER

- X Great way to show off your work & talk to people
 - Remember the “let people know about it” part?
- X Many conferences have non-archival / lightly reviewed tracks
 - In ML, CV, etc., full papers get poster slots.



← UIST 2017
demos

ESEC/FSE 2019 →
posters





POSTER TIPS

- X Avoid making it text-heavy.
 - Use as talking points, not a printed version of the paper.
 - Many people will look at it from a distance.
- X Add your contact info.
 - Many people might see it even when you're not at the booth.
- X Acknowledge sponsors.
- X Bring a laptop for additional info and/or demo.
- X Examples & Resources
 - [UIST poster gallery](#)
 - [Research Posters 101](#)

Interaction Peaks and Data-Driven Interfaces for Online Lecture Videos

Juho Kim (MIT CSAIL)

with Rob Miller (MIT CSAIL)



Understanding In-Video Dropouts and Interaction Peaks in Online Lecture Videos.

Juho Kim, Philip J. Guo, Daniel T. Seaton, Piotr Mitros, Krzysztof Z. Gajos, Robert C. Miller. *Learning at Scale 2014*, to appear.

Video interaction data from MOOCs

Motivation: How do students learn from videos on Massive Open Online Courses (MOOCs)? We analyze video interaction data (pause, play, scrubbing).

Dataset: interaction log from 4 edX courses in Fall 2012

Course	Subject	Enrollment	Students	Videos	Video Length	Viewed Events
6.034	Intro CS & Programming	MIT	59,129	141	7:40	4,401,448
6.037a	Statistics for Public Health	Harvard	50,742	50	10:44	11,611,068
CS-439-1x	Artificial Intelligence	Stanford	11,496	140	4:44	14,174,201
10.001a	Intro to Chemistry	MIT	11,261	275	30:19	4,451,437
Total			127,628	562	7:46	30,638,152

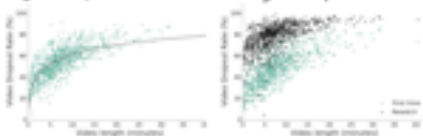
Video dropout analysis

Video dropout: percentage of students navigating away from a video before completion.

Overall dropout rate: 55.2% (36.6% within the first 3%)

Longer videos have a higher dropout rate.

Re-watching sessions have a higher dropout rate.



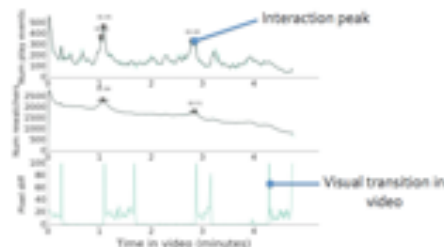
Interaction peak analysis

Interaction peaks occur when a significant number of students play, pause, or replay at the time of the video.

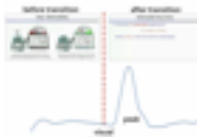
- 3.7 peaks per video on average
- Tutorial videos show more frequent and stronger peaks than lecture videos.
- Re-watching sessions show more frequent and stronger peaks than first-time sessions.

What causes interaction peaks to occur?

Observation: interaction peaks often accompany visual transitions in the video.



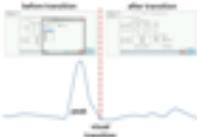
1. Beginning of new material



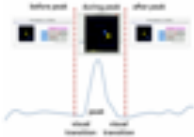
2. Returning to content



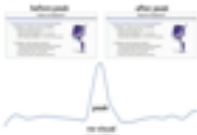
3. Tutorial step



4. Replaying a segment



5. Non-visual explanation



More content channels for future analysis

- Transcript: text analysis
- Acoustic: speech analysis

Implications

For video editors & instructors

- Avoid sudden visual transitions
- Make shorter videos

For video interfaces

- Provide interactive links and screenshots for highlights.
- Consider video summarization for selective watchers.
- Enable one-click access for steps in tutorial videos.

Data-driven video interface

Can video interaction data be used to improve students' learning experience?

Video interface dynamically generated by learner data



Acknowledgements

- This work is supported in part by Quanta Computer & edX.
- Juho Kim is supported by the Samsung Fellowship.

PyGGI 2.0: Language Independent Genetic Improvement Framework

Gabin An¹, Aymeric Blot², Justyna Petke², and Shin Yoo¹
agb94@kaist.ac.kr, a.blot@cs.ucl.ac.uk, j.petke@ucl.ac.uk, shin.yoo@kaist.ac.kr

¹: Korea Advanced Institute of Science and Technology, Republic of Korea | ²: University College London, UK

What is PyGGI?

PyGGI (Python General Framework for Genetic Improvement) is an easy-to-use Genetic Improvement (GI) framework that has been designed for quick experimentation and proof-of-concept implementation. PyGGI provides lower level program representations and interfaces to actual source code, so that users can only focus on algorithmic aspects of GI such as the search algorithm and the edit operators.

What is special about version 2.0?

The latest release, version 2.0, supports srcML (<https://www.srcml.org>) as the program representation, in addition to its default language, Python. srcML is an XML representation of Abstract Syntax Tree (AST) for Java, C/C++, and C#. This means that users of PyGGI can define edit operators for srcML representations and apply them to GI experiments for all languages supported by srcML. Since srcML is also just XML, powerful tools like XPATH can be used to define operators.

```
// Source Code
if ( j > i ) {
    x = j;
}

<!-- srcML Representation -->
<stmt>if <condition>( j &gt; i )</condition> {
    <stmt >x = j;</stmt>
}</stmt>
```

How do I use PyGGI?

PyGGI already provides the default srcML representation, a set of basic edit operators, and a default local search algorithm. All you need to do in order to get started is to write a simple JSON configuration file:

[pyggi.config]

```
{
  "target_files": ["triangle.py"],
  "test_command": "pytest -s test_triangle.py"
}
```

and initiate the search:

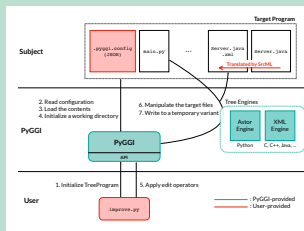
[repair.py]

```
program = MyTreeProgram("../sample/Triangle_bug_python")
tabu_search = MyTabuSearch(program)
tabu_search.operators = [StmReplacement, StmInsertion, StmDeletion]
result = tabu_search.run(warmup_reps=1, epoch=5, max_iter=100, timeout=10)
```

Where can I get PyGGI?

PyGGI is publicly available from <https://github.com/coincse/pyggi> based on MIT license. The only requirement is Python 3.5+ and srcML binary.

Have fun!



<https://coincse.io>



<https://crestweb.cs.ucl.ac.uk>

Ahead of Time Mutation Based Fault Localisation using Statistical Inference

Jinhan Kim
KAIST
jinhankim@kaist.ac.kr

Gabin An
KAIST
agb94@kaist.ac.kr

Robert Feldt
Chalmers University
robert.feldt@chalmers.se

Mike Papadakis
University of Luxembourg
michail.papadakis@uni.lu

Shin Yoo
KAIST
shin.yoo@kaist.ac.kr

The Problem

- Mutation Based Fault Localisation (MBFL) techniques such as MUSE, Metallaxis suffer from **high cost** of mutant execution **after the observation of faults**.
- How about doing mutation analysis **in advance**? (i.e., before the faults are observed)
- **Assumption**: for a test that has killed the mutants located on a specific program element, the same program element should be identified as the suspicious location when the same test later fails again.

Methodology of SIMFL

Approximating the probability of a test t killing a mutant

$$\Pr(F_t | M_t) \approx \frac{| \{ m \in X_e | t \in K_m \} |}{|X_e|}$$

a set of tests that kill mutant m

a set of mutants located on a program element e

an event that e is mutated

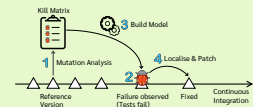
an event that a test case t fails on a given program.

Using Bayes' rule

$$\Pr(M_t | F_t) = \frac{\Pr(F_t | M_t) \Pr(M_t)}{\Pr(F_t)}$$
$$\approx \Pr(\text{fault exists in } e | F_t)$$

Based on this probability, we build several ranking models

Usage Scenario of SIMFL

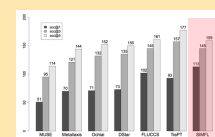


- (1) Perform mutation analysis for a version of a SUT, and produce the kill matrix. The version is called the reference version.
- (2) While testing a subsequent version, a failure is observed.
- (3) Using the information of which test case(s) failed, as well as the kill matrix, build a predictive model for fault localisation.
- (4) Produce the ranking.

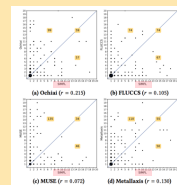
Research Questions

- RQ1. Localisation Effectiveness: Does the models of SIMFL produce accurate fault localisation compared to the state-of-the-art FL techniques?
- RQ2. Model Viability: How well does SIMFL hold up when applied using models built earlier?
- RQ3. Sampling Impact: What is the impact of mutation sampling to the effectiveness of SIMFL?

Experimental Result



RQ1. Comparison to other FL techniques



RQ2. Relation with other FL techniques

61	PCB Rank	7	4	5	5	6
	f529d4 (3)	7	4	5	5	6
	3d30d4 (4)	7	4	5	5	6
	2d30d4 (7)	7	4	5	5	7
	8d1d19 (8)	7	4	5	5	8
	37de13 (9)	7	4	5	5	8

RQ2. only 19 out of 140 cases show degraded performance since we used less recent mutation analysis results

55.00	105.00	133.25	172.90
44.05	121.45	133.75	193.75
70.00	128.10	156.50	193.10
74.50	127.05	162.05	200.15
68.70	128.20	158.15	196.40
62.30	140.05	160.05	199.35
61.55	130.15	155.55	194.60
60.55	62.40	69.05	81.40
160.70	118.60	134.60	148.90
16.25	138.20	156.00	165.90

RQ3. Using uniform sampling, on average 80% of the faults ranked at the top without sampling can still be localised at the top

TMI TIP

- X You can print your posters on wrinkle-free, cloth-like material
- X Makes travelling so much easier!
- X Printing on location is usually difficult and much more expensive: prepare in advance
- X DO NOT EVER put up multiple A4 sheets as your poster!



- X Visual demonstration of your technique with storytelling
- X It scales! (it's like demoing to the whole world)
- X It lasts! (even if your demo cannot run live anymore, video is likely to still play)
- X It engages! (who doesn't like watching cool videos?)
- X Common in HCI/Graphics/Vision
- X Examples
 - [Not Going to Take This Anymore: Multi-objective Overtime Planning for Software Engineering Projects](#)
 - [inFORM – Interacting With a Dynamic Shape Display](#)
 - [Revealing Invisible Changes In The World](#)



OPEN SCIENCE

- X More emphasis on open science (data, code, reviews, questionnaires, etc.)
- X Important for greater impact & replicability
- X Helps to keep a project website with all resources available
- X Examples
 - [RecipeScape: An Interactive Tool for Analyzing Cooking Instructions at Scale](#)
 - [MoSculp: Interactive Visualization of Shape and Time](#)
 - [CMU Panoptic Dataset](#)



OPEN SCIENCE: ARTIFACT EVALUATION

- X Gaining popularity in PL/SE
- X ACM supports artifact review & badging

Artifact Review and Badging:

A variety of research communities have embraced the goal of reproducibility in experimental science.

[\[more information\]](#)

Artifacts Evaluated Functional

The artifacts associated with the research are found to be documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation.

Artifacts Evaluated Reusable

The artifacts associated with the paper are of a quality that significantly exceeds minimal functionality.

Artifacts Available

Author-created artifacts relevant to this paper have been placed on a publically accessible archival repository.

Results Replicated

The main results of the paper have been obtained in a subsequent study by a person or team other than the authors, using, in part, artifacts provided by the author.

Results Reproduced

The main results of the paper have been independently obtained in a subsequent study by a person or team other than the authors, without the use of author-supplied artifacts.



WEBSITE

- X Various people might want to learn about you and your work.
 - They mostly start by Googling your name.
- X Convenient to route people to various resources: “It’s on my website.”
- X Could be seen as a visual, creative, accessible version of your CV.
 - Videos, images, news items, travel schedule, etc. can be shown.
- X More important for junior researchers who need more visibility.
- X Most CS academics nowadays have one.
- X Examples
 - [Minsuk Chang](#)
 - [Stefanie Mueller](#)
 - [Frans Kaashoek](#)
 - [Me](#)
 - [Maneesh Agrawala](#)
 - [Barbara Liskov](#)



- X You can also create a project-specific website
 - Github is a popular choice, but you can also host at your institution
 - Anything you had to leave out due to space restriction can live here
 - You can get DOI for online pages: this is the PROPER way of doing open repository (because of versioning). See <https://zenodo.org> for example.
- X Making it interactive is a good way to engage people
 - <https://coinse.kaist.ac.kr/projects/sbflvis/>
 - [RecipeScape: An Interactive Tool for Analyzing Cooking Instructions at Scale](#)
 - <https://rise4fun.com/z3>



NETWORKING & CONFERENCES

- X You're there as an active participant, not as a consumer.
- X Listen to talks, but more importantly, talk to people.
- X List up people you want to talk to.
 - People who do relevant research regardless of their fame
- X Prepare elevator pitches.
 - You'll be asked "Are you presenting anything?" numerous times.
- X Try NOT to hang out with labmates (even worse: continuously speaking in Korean or other non-English languages).
- X Make introductions & ask for introductions.



NETWORKING & CONFERENCES: RESOURCES

- X [Attending Professional Conferences as a Newcomer](#) by Philip Guo
- X [Networking Tips for Younger PhD Students](#) by Jean Yang
- X [Advice for Social Interactions & Relationships](#) by Philip Guo



PC: Philip Guo @
CHI 2017

JOB MARKET

SOME NOTES

X N=1 (Prof. Kim)... + 0.5 (me)

- Quick tour of what the last year of Ph.D. looks like (if you target the U.S. job market)
- On the U.S. (& KAIST) job market in 2014–2015

MY ULTIMATE VISION FOR KAIST STUDENTS

- X Faculty at top schools anywhere in the world! MIT, Stanford, CMU...
- X Because... you can.
- X But we need more leadership in the field, and students can and should be a driving force.

THE TIMELINE

- X 2014.8: discussion with advisors
- X 2014.9–11: research statement
- X 2014.10: letters of recommendation
- X 2014.11: teaching & diversity statement
- X 2014.11–12: school / company selection
- X 2014.12: application
- X 2015.2–4: interview
- X 2015.4–5: offer
- X 2015.5: decision

1. DISCUSSION WITH ADVISORS

X The talk: “Can I graduate?”

- Am I ready?
- Does my advisor think I’m ready?
- Does the market think I’m ready?

X Three pillars of research

X Market fluctuation >> my productivity increase

2. RESEARCH STATEMENT

- X Your identity as a researcher
- X Past & future with focus on results, impact, and vision
- X At least 2 months
- X 10+ readers at different stages
- X 2-3 complete rewrites

3. LETTERS OF RECOMMENDATION

- X One of the most important parts, partly because letters can't be easily made up and take time & effort to get & write
- X Strong letter, well known > Weak letter, well known
> Strong letter, less known > Weak letter, less known
- X Every letter says "best": it's the descriptor that matters.
- X Plan your writers YEARS in advance.

4. TEACHING & DIVERSITY STATEMENT

X Often not a crucial factor in R1 universities

X Teaching statement

- Need a philosophy
- Possible courses you can teach
- Something beyond TAing helps.

X Diversity statement

- How you plan to contribute to diversity
- Equal vs. Equitable

5. SCHOOL / COMPANY SELECTION

- X ~15 places
- X 10–30 is quite common in HCI/CS job market
- X Rule of thumb: $1/3$ interview offers, $1/3^2$ final offers

6. APPLICATION

- X RS, TS, DS, Letters
- X CV, Cover letter, Website
- X Advisors' gentle prod
- X Now pray... at least focus on something ELSE
- X Don't try to overthink what's happening at these places.

7. INTERVIEW

X Job talk

- Most people spend months on it
- 5–10 practice talks, video recordings
- The most organized / prepared talk you'll listen to
- Attend & watch many job talks, especially ones outside your area.

X Interview day

- 1–2 days, 10–20 1-on-1s, 3+ uncomfortable meals
- Faculty, students, dean / chair / head, potential collaborators from other departments

8. OFFER

- X You're 갑 for the first time, and only for a short time.
- X Enjoy the negotiation process.
- X Salary, startup fund, space, teaching load, etc.
- X Deferring for 0.5 or 1 year might be a compelling option, especially if you're a fresh Ph.D.

9. DECISION

- X People in your area, department, school
- X Community, life, location, family
- X Industry vs Academia, Korea vs US vs ...
- X Probably a more painful decision than you think.
- X Make sure you stay in good terms with people at places you turn down.

NOW ABOUT THAT 0.5

- X US/UK systems differ a bit: PhDs are relatively shorter (although they are getting longer now) and post-doc experience is almost expected
- X I did a few things that people tell you not to do
 - I stayed on at the same group that I did my PhD as a postdoc
 - Then I got hired by the same group (which moved to another institution) as a lecturer (= assistant professor)
- X In general, you still should not do these... but the lesson is also that job market is complicated and everyone has their own story...

NOW ABOUT THAT 0.5

X Things I did to gain independence

- Had a webpage! 🤪 (with a “Peer Esteem” section)
- Wrote a sole author paper near the end of my PhD
- (although the PI was my supervisor) wrote and received two Google Faculty Award grants
- Launched a new international workshop series on the topic of my PhD
- Took up reviewing and PC roles

JOB MARKET LESSONS

- Job market prep starts TODAY.
 - Community building in at KAIST is a great first step.
- “What do you want to be known as?”
 - “That person who did X”
- You are the product. Advertise yourself.
 - Getting people to know you is difficult.
 - You can easily write 10 papers nobody cares about.
 - Engage in community discussions & volunteer to lead efforts.
 - Collaborate, give talks, & visit places.

RESOURCES

- X [Tomorrow's Professor mailing list](#)
- X [Philip Guo's blog & Ph.D. Grind](#)
- X [Jean Yang's blog](#)
- X [mcpanic.com](#): 박사과정을 돌아보며
- X [UCSD job talk videos](#)
- X [Juho Kim's faculty app material](#)