Genetic Algorithms as Shrinkers in Property-Based Testing

Fang-Yi Lo
National Chiao Tung University
Hsinchu City, Taiwan
fylo@nclab.tw

Chao-Hong Chen
Indiana University
Bloomington, IN, USA
chen464@indiana.edu

Ying-ping Chen
National Chiao Tung University
Hsinchu City, Taiwan
ypchen@cs.nctu.edu.tw

ABSTRACT

This paper proposes the use of genetic algorithms as shrinkers for shrinking the counterexamples generated by QuickChick, a property-based testing framework for Coq. The present study incorporates the flexibility and versatility of evolutionary algorithms into the realm of rigorous software development, in particular, making the results of property-based testing observable and comprehensible for human. The program code for merge sort is investigated as a showcase in the study. Due to the lack of similar proposals in the literature, random sample is used to compete with the proposal for comparison. The experimental results indicate that the proposed genetic algorithm outperforms random sample. Moreover, the minimal counterexamples, through which programmers are able to pinpoint the program mistakes with ease, can be successfully obtained by using genetic algorithms as shrinkers.

CCS CONCEPTS

• **Computing methodologies → Genetic algorithms**; • **Software and its engineering → Software testing and debugging**;

KEYWORDS

Genetic algorithms, Property-based testing, Shrinker, QuickChick, Coq

ACM Reference Format:
Fang-Yi Lo, Chao-Hong Chen, and Ying-ping Chen. 2019. Genetic Algorithms as Shrinkers in Property-Based Testing. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3319619.3322004

1 INTRODUCTION

Within the domain of software testing, techniques have been devised to respond to the fast-growing complexity of computer software. Among these techniques is property-based testing, nowadays widely used for quickly finding program mistakes, i.e., so-called *bugs*, in software development [1, 2, 7]. The idea of property-based testing is to specify certain property of a computer program and use random testing to discover counterexamples, if exist. Because the counterexamples found by random testing is usually complicated, barely comprehensible, *shrinking* is crucial and indispensable for property-based testing.

In the present work, QuickChick [4], originated from QuickCheck [3] for Haskell is adopted. QuickChick is a property-based testing framework for Coq [5], a proof assistant widely utilized to prove mathematical theorems [6] and to develop verified software [8]. In this study, QuickChick is utilized as the testing framework, and genetic algorithms are employed as general shrinkers, capable of handling a broad range of data types and structures, for shrinking counterexamples found by QuickChick. The program code for merge sort will be investigate as a showcase in this article to demonstrate that genetic algorithms are able to deliver satisfactory performance. Hence, this study may be considered presenting a step forward helping to build practical verified software.

2 SHOWCASE – MERGE SORT

Merge sort is a divide-and-conquer algorithm. The common practice of implementation separates into two parts: (a) an unsorted list is bisected into sublists recursively; (b) sublists are then merged by comparing their elements in order.

```
Fixpoint merge l1 l2 :=
  let fix merge_aux l2 :=
  match l1, l2 with
    | [],  _ ⇒ l2
    | _,  [] ⇒ l1
    | a1:: l1',  a2:: l2' ⇒ if a1 <? a2 then a1 ::  merge l1' l2
                            else if a2 <? a1 then a2 ::  merge_aux l2'
                            else a1 ::  merge l1' l2'
  end
  in merge_aux l2.
```

The property to test in this showcase is specified as

```
Conjecture LengthPreserved : forall l, length l = length (sort l).
```

The length of a list should remain unchanged after sorting. Here, as aforementioned, an implementation error is introduced on purpose. By running the generator and checker in QuickChick, a counterexample as a list which contains 748 integers can be found as

```
[ 5159922 ; 4532574 ; 34127068 ; 85259035 ; 18312552 ; 7119084 ; 9752785 ; 21510913 ; 54167003 ; 58476163 ; 61939581 ; 16079288 ; 74201198 ; 66271956 ; 30727362 ; 54008295 ; 34167376 ; 16134250 ; 6246051 ; 40071462 ; 23285501 ; 15898394 ; 1253473 ; 79615076 ; 64402078 ;
68496675 ; 29945601 ; 12273111 ; 88664484 ; 2437294 ; 71672561 ; 15327070 ; 80704006 ; 30966511 ; 22353129 ; 20742250 ; 26503141 ; 99567972 ; 76116861 ; 13994263 ; 10771900 ; 14596144 ; 7245963 ; 70831603 ; 31923208 ; 75877622 ; 41330583 ; 19829719 ; 66533945 ; 64513020 ;
333479 ; 6413382 ; 84147647 ; 24787344 ; 53755707 ; 74516351 ; 19392189 ; 98417271 ; 23948206 ; 52097737 ; 54117139 ; 3524857 ; 23041482 ; 31425905 ; 14883266 ; 80511432 ; 33735348 ; 92251102 ; 46498512 ; 50843369 ; 81651384 ; 56789675 ; 86729997 ; 95548128 ; 77131095 ;
20945678 ; 19461739 ; 14584597 ; 22737740 ; 76786131 ; 88997961 ; 35108843 ; 82078369 ; 20428335 ; 95340975 ; 81478546 ; 50358686 ; 47028577 ; 44690453 ; 78600878 ; 79943317 ; 5899921 ; 37035721 ; 42197799 ; 29479411 ; 84973782 ; 77601228 ; 91916754 ; 87829099 ; 5902992 ;
1481093 ; 49581829 ; 77501228 ; 95881609 ; 83544485 ; 31351522 ; 70507277 ; 68340322 ; 23512287 ; 22436500 ; 13673397 ; 85315228 ; 34176106 ; 74545458 ; 98865815 ; 4520024 ; 81784591 ; 61911840 ; 22418233 ; 85370491 ; 13678651 ; 98151045 ; 9792402 ; 93041155 ; 11130796 ;
41763453 ; 64331235 ; 13723079 ; 71518004 ; 35032955 ; 51622913 ; 16065689 ; 72319551 ; 11137772 ; 44911502 ; 39688094 ; 90935621 ; 89939490 ; 4990841 ; 72795211 ; 60701990 ; 56263839 ; 39035178 ; 31023921 ; 67813923 ; 50240581 ; 91391249 ; 33650195 ; 62971710 ; 31853600 ;
92537932 ; 23879674 ; 25202337 ; 12611904 ; 91345281 ; 41832457 ; 9604678 ; 81320680 ; 22130186 ; 10664206 ; 95831838 ; 2982018 ; 24811163 ; 30226724 ; 73864192 ; 91739239 ; 15063519 ; 9510854 ; 80837978 ; 48872031 ; 78682095 ; 24405665 ; 94425636 ; 77606326 ; 43281755 ;
28780222 ; 1671848 ; 53561148 ; 45576679 ; 90314117 ; 58912644 ; 76825251 ; 54935801 ; 89923030 ; 22377978 ; 47000628 ; 26232191 ; 32330759 ; 13065804 ; 603698 ; 27080853 ; 75777219 ; 11771353 ; 86490562 ; 12899370 ; 72486201 ; 46087440 ; 3582969 ; 31078557 ; 66521444 ;
38590485 ; 87377388 ; 52537424 ; 47971075 ; 27744184 ; 68484092 ; 93017152 ; 95393596 ; 71634798 ; 51555323 ; 98955425 ; 12013065 ; 25359996 ; 71811806 ; 56458912 ; 6194960 ; 31063221 ; 15667220 ; 36591712 ; 82873403 ; 60643390 ; 91430824 ; 63563224 ; 93170033 ; 1520618 ;
7447455 ; 618516 ; 95301663 ; 40262333 ; 40938536 ; 98114108 ; 86656846 ; 22586106 ; 9198360 ; 44029784 ; 28440180 ; 97867001 ; 83886082 ; 38279408 ; 92461889 ; 39093918 ; 27395407 ; 83583161 ; 69207675 ; 45352506 ; 3549413 ; 61754664 ; 66432870 ; 40210159 ; 27164829 ;
30021301 ; 50940599 ; 45707457 ; 35191639 ; 95360813 ; 62951981 ; 62590344 ; 79553628 ; 36178172 ; 64470373 ; 70522973 ; 14444287 ; 21211425 ; 95647991 ; 252000 ; 80584111 ; 16339205 ; 69342628 ; 34387271 ; 75685639 ; 36526982 ; 70484753 ; 60638488 ; 8915730 ; 96719448 ;
73880325 ; 40828670 ; 70283627 ; 18137331 ; 43821556 ; 21848465 ; 66928348 ; 85538072 ; 68736041 ; 32878297 ; 7993807 ; 34045157 ; 74614662 ; 35060551 ; 33164374 ; 60305344 ; 10093370 ; 49791772 ; 87949920 ; 22400117 ; 76737145 ; 97649595 ; 16917956 ; 86114494 ; 42426908 ;
9071962 ; 36189484 ; 11676659 ; 29821308 ; 80727691 ; 23375175 ; 76471425 ; 12577019 ; 37451333 ; 26240552 ; 89947465 ; 28526278 ; 8245824 ; 40887464 ; 90953881 ; 4403193 ; 67551641 ; 62249166 ; 80578131 ; 36688474 ; 23123812 ; 95869661 ; 11933563 ; 42257022 ; 17565830 ;
46374286 ; 13459757 ; 4251828 ; 12506647 ; 79318410 ; 30920403 ; 3117832 ; 97214708 ; 98557351 ; 39864173 ; 10516594 ; 87191676 ; 72511888 ; 59404923 ; 50252812 ; 77827878 ; 44597372 ; 16937442 ; 5161994 ; 3553417 ; 89951751 ; 30625846 ; 62565702 ; 4485932 ; 79864631 ;
9238315 ; 64554770 ; 77216863 ; 51328981 ; 80013483 ; 84309130 ; 69539277 ; 11237694 ; 35881835 ; 69755607 ; 66827365 ; 49713448 ; 14763006 ; 68390444 ; 65873011 ; 51404062 ; 4403765 ; 3173288 ; 26956547 ; 12501978 ; 63656445 ; 25175487 ; 53625940 ; 28285806 ; 25681278 ;
28143493 ; 72882869 ; 80131543 ; 77118393 ; 93324399 ; 39748308 ; 77061417 ; 82793454 ; 8507549 ; 83131327 ; 7781994 ; 94354811 ; 51101866 ; 93509027 ; 92564593 ; 56947271 ; 65597417 ; 9118360 ; 65015808 ; 68341504 ; 9565759 ; 32984881 ; 43166490 ; 43127824 ; 26222142 ;
55919619 ; 59748748 ; 37524136 ; 90236049 ; 5360363 ; 24211845 ; 86403204 ; 61427520 ; 89121873 ; 77244213 ; 12943240 ; 64841463 ; 78263154 ; 76216872 ; 57495446 ; 9117813 ; 45750480 ; 85640199 ; 70226861 ; 61351035 ; 68770722 ; 62333105 ; 77517774 ; 36132957 ; 61904065 ;
48225607 ; 37851565 ; 58666165 ; 71309982 ; 16010694 ; 32611845 ; 33038006 ; 77272450 ; 67297467 ; 88153812 ; 32719394 ; 69209514 ; 83476675 ; 28346071 ; 32710473 ; 57406046 ; 35210419 ; 41814292 ; 66613795 ; 74133626 ; 18226970 ; 22368753 ; 64826563 ; 70916797 ; 28013798 ;
31757803 ; 46903789 ; 17823679 ; 7974350 ; 75375717 ; 48902221 ; 2928519 ; 40222533 ; 19015050 ; 94465531 ; 68453200 ; 54516664 ; 6498885 ; 52995278 ; 10982545 ; 72009593 ; 26704955 ; 92594484 ; 354743 ; 18350665 ; 53891074 ; 22819656 ; 36843222 ; 77909576 ; 54359939 ;
10531017 ; 39310366 ; 8768357 ; 86679947 ; 3067809 ; 62914476 ; 21225144 ; 77119865 ; 87305652 ; 78906359 ; 42457864 ; 91082314 ; 72941925 ; 14807760 ; 89933220 ; 51929876 ; 9120948 ; 39209346 ; 36659530 ; 46192963 ; 13823881 ; 19576953 ; 92986318 ; 18581704 ; 66977609 ;
44975839 ; 93736466 ; 64857020 ; 24813357 ; 42973227 ; 28354720 ; 73542892 ; 10402280 ; 61840345 ; 32254521 ; 29395209 ; 40240179 ; 97843603 ; 59681237 ; 48956748 ; 70335760 ; 10182665 ; 86817345 ; 70782524 ; 83818189 ; 16808806 ; 69012619 ; 50667099 ; 74782067 ; 97154092 ;
99199848 ; 77508626 ; 84286212 ; 28763017 ; 51536951 ; 90030714 ; 13986005 ; 90456546 ; 31732763 ; 60848521 ; 89308633 ; 4298212 ; 23336823 ; 28595324 ; 81306131 ; 87776827 ; 33946899 ; 95060348 ; 88166110 ; 33253462 ; 26899401 ; 62473334 ; 90027008 ; 54315574 ; 50795798 ;
61784644 ; 89007259 ; 61744960 ; 48225626 ; 42173060 ; 32121540 ; 57829105 ; 39165285 ; 39635476 ; 22285234 ; 29576362 ; 42865560 ; 75248050 ; 53254955 ; 19930775 ; 38777991 ; 94381685 ; 30673183 ; 30987623 ; 30794210 ; 98242496 ; 50062265 ; 51646431 ; 3824709 ; 37255397 ;
87181091 ; 53515424 ; 28304431 ; 46070863 ; 17770263 ; 93319230 ; 62211613 ; 63854303 ; 44935808 ; 54732130 ; 43933685 ; 71880220 ; 98234704 ; 38261020 ; 17353205 ; 3070631 ; 70308399 ; 67679264 ; 53184237 ; 21411785 ; 21281073 ; 93146513 ; 21014632 ; 85109781 ; 49038280 ;
11846904 ; 66911862 ; 45606984 ; 99000361 ; 35406720 ; 21946674 ; 60425971 ; 3899977 ; 83193742 ; 32954706 ; 319088 ; 31850976 ; 89584512 ; 93430663 ; 14059582 ; 29980182 ; 23446812 ; 91811001 ; 37012710 ; 32642706 ; 74284582 ; 34058254 ; 1102608 ; 65921761 ; 3246505 ;
99316426 ; 25105776 ; 8195729 ; 64883406 ; 3600876 ; 57917778 ; 10940338 ; 38926225 ; 77034223 ; 23876533 ; 80834470 ; 64184077 ; 3146561 ; 98457515 ; 54103708 ; 37042030 ; 17130203 ; 26546775 ; 91126081 ; 71253417 ; 84682669 ; 53923782 ; 67431074 ; 55072887 ; 86212387 ;
41218237 ; 15093400 ; 66228294 ; 70712764 ; 55660440 ; 86830081 ; 90285006 ; 88075931 ; 94134060 ; 98139105 ; 3611475 ; 55727520 ; 96677179 ; 43969113 ; 89153554 ; 20422421 ; 26593367 ; 16526727 ; 93356592 ; 86746418 ; 11326603 ; 77446653 ; 27649391 ; 83306044 ; 74499910 ;
83999086 ; 5287723 ; 24091145 ; 89526806 ; 51095793 ; 44819098 ; 99269355 ; 77168629 ; 35897150 ; 32694672 ; 93737168 ; 40823442 ; 63169892 ; 1285667 ; 2068975 ; 12071423 ; 50780881 ; 81011144 ; 6283774 ; 46969196 ; 43977646 ; 87780146 ; 4297438 ; 66928390 ; 5015875 ;
32519524 ; 40119983 ; 24524273 ; 15851475 ; 76485582 ; 2258698 ; 6009760 ; 59387477 ; 64184077 ; 61362300 ; 87345271 ; 97560860 ; 60294447 ; 62564343 ; 75722886 ; 90599900 ; 84792556 ; 52323443 ; 4998736 ; 54137217 ; 44853882 ; 31686178 ; 63785473 ; 92337740 ; 83730429 ;
74618596 ; 27283684 ; 62280049 ; 88296507 ; 98474286 ; 55365198 ; 22046327 ; 91215751 ; 95601583 ; 15120820 ; 53654198 ; 64962596 ; 53124647 ; 58106818 ; 2185746 ; 76773346 ; 82137152 ; 61580134 ; 64633295 ; 87794320 ; 65803691 ; 84865415 ; 51967933 ]
```

It is obvious that although this counterexample proves the program wrong, the software developer is unlikely able to extract useful information by observation to correct the program code.

Algorithm 1 RANDOMDELETE for integer lists

```
procedure RANDOMDELETE($l$, $n$)
    $d \leftarrow$ RANDOM$(0, n)$
    for $i = 0$ to $d$-1 do
        $r \leftarrow$ RANDOM$(0, length(l) - 1)$
        $l \leftarrow$ DELETE$(l, r)$
    return $l$
```

Algorithm 2 Genetic Algorithm Shrinker

```
$ce \leftarrow \{OriginalCounterexample\}$
$initdmax \leftarrow$ SIZE$(ce)/n$
$penalty \leftarrow$ SIZE$(ce)$
$psize \leftarrow \{MaximumPopulation\}$
$mdmax \leftarrow \{MaximumElementsToDeleteInMutation\}$
$cemrate \leftarrow \{CounterexampleMutationRate\}$
$ncemrate \leftarrow \{NonCounterexampleMutationRate\}$
$survive \leftarrow \{IndividualsGuaranteeToSurvive\}$
$generation \leftarrow 0$
$eval \leftarrow 0$
$P[psize * 2] \leftarrow$ INITIALIZATION$(ce, initdmax)$
while $(eval < EvalMax)$ do {
    for $i = psize$ to $psize$ * 2 - 1 do
        $PA \leftarrow$ RANDOMSELECT$(P[0], P[psize - 1])$
        $PB \leftarrow$ RANDOMSELECT$(P[0], P[psize - 1])$
        $P[i] \leftarrow$ CROSSOVER$(PA, PB)$
    for $i = 0$ to $psize$ * 2 - 1 do
        $P[i] \leftarrow$ MUTATION$(P[i], mdmax, cemrate, ncemrate)$
    $P \leftarrow$ SELECTSURVIVOR$(P, survive)$
    $generation \leftarrow generation + 1$ }
```

3 SHRINKERS

Random sample and genetic algorithms are used as shrinkers. For the showcase of merge sort, Algorithm 1 shows the interface for both shrinkers to manipulate the data structure, integer lists.

3.1 Shrinker based on Random Sample

The shrinker based on random sample samples a random instance which is smaller than the original, given counterexample, by using the functionality of RANDOMDELETE. 500 random samples are tested.

3.2 Shrinker based on Genetic Algorithms

Algorithm 2 presents the pseudo code for the shrinker based on genetic algorithms, in which CROSSOVER generates a child by copying *PA* and deleting elements not in *PB*, and MUTATION deletes elements by using RANDOMDELETE. *EvalMax* is also 500 as for random sample.

4 RESULTS

For the counterexample introduced in section 2, the minimum counterexample, can be found by the proposed GA shrinker as

[64184077 ; 64184077].

By examining the counterexample, it is reasonable to speculate that the mistake of the program for merge sort is due to the inability to handle duplicate elements. Thus, the fix can be easily done as

Table 1: The mean and standard deviation of the minimal sizes over 30 runs are listed for random sample (RS) and genetic algorithm (GA) for merge sort.

Len.	RS	GA	Len.	RS	GA
259	43.3 ± 17	2 ± 0.2	748	134.3 ± 39.2	2 ± 0
443	57.1 ± 19.4	2 ± 0	856	109 ± 39.7	2 ± 0
520	72.2 ± 25.7	2 ± 0	874	143.1 ± 45.6	2.2 ± 0.9
629	112.4 ± 34	2 ± 0	932	119.5 ± 38.6	2 ± 0
705	112.3 ± 39.1	2 ± 0	985	127.3 ± 41.3	2 ± 0

```
Fixpoint merge l1 l2 :=
  let fix merge_aux l2 :=
    match l1, l2 with
      | [],  _ ⇒ l2
      | _, [] ⇒ l1
      | a1:: l1',  a2:: l2'  ⇒ if a1 <? a2 then a1 ::  merge l1'  l2
                                else if a2 <? a1 then a2 ::  merge_aux l2'
                                else a1 :: a2 ::  merge l1'  l2'
    end
  in merge_aux l2.
```

Further assessing the capability of shrinkers, Tables 1 show the average size and standard deviation of the minimal counterexamples found by the two shrinkers over the 30 runs for merge sort on 10 different counterexamples. The results clearly indicates that the proposed GA-based shrinker outperforms random sample.

ACKNOWLEDGMENTS

The work was supported in part by the Ministry of Science and Technology of Taiwan under Grant MOST 107-2221-E-009-101.

REFERENCES

[1] Bernhard K. Aichernig and Richard Schumi. 2016. Property-Based Testing with FsCheck by Deriving Properties from Business Rule Models. In *2016 IEEE Ninth International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*. 219–228. https://doi.org/10.1109/ICSTW.2016.24

[2] Clara Benac Earle, Lars-Åke Fredlund, and John Hughes. 2016. Automatic Grading of Programming Exercises Using Property-Based Testing. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '16)*. ACM, New York, NY, USA, 47–52.

[3] Koen Claessen and John Hughes. 2000. QuickCheck: A Lightweight Tool for Random Testing of Haskell Programs. In *Proceedings of the Fifth ACM SIGPLAN International Conference on Functional Programming (ICFP '00)*. ACM, New York, NY, USA, 268–279. https://doi.org/10.1145/351240.351266

[4] Maxime Dénès, Catalin Hritcu, Leonidas Lampropoulos, Zoe Paraskevopoulou, and Benjamin C Pierce. 2014. QuickChick: Property-based testing for Coq. In *The Coq Workshop*.

[5] The Coq development team. 2004. *The Coq proof assistant reference manual*. LogiCal Project. http://coq.inria.fr Version 8.0.

[6] Georges Gonthier. 2008. Formal proof–the four-color theorem. *Notices of the AMS* 55, 11 (2008), 1382–1393.

[7] Ronghui Gu, Zhong Shao, Hao Chen, Xiongnan Wu, Jieung Kim, Vilhelm Sjöberg, and David Costanzo. 2016. CertiKOS: An Extensible Architecture for Building Certified Concurrent OS Kernels. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation (OSDI'16)*. USENIX Association, Berkeley, CA, USA, 653–669. http://dl.acm.org/citation.cfm?id=3026877.3026928

[8] Xavier Leroy. 2006. Formal certification of a compiler back-end, or: programming a compiler with a proof assistant. In *33rd ACM symposium on Principles of Programming Languages*. ACM Press, 42–54.