A-022

Relational Properties Expressible with One Universal Quantifier are Testable (Extended Abstract)

Charles Jordan[†] Thomas Zeugmann[‡]

Abstract. Property testing is an application of induction in which we take a small, random sample of an object and wish to distinguish with high probability between the case where it has a desired property and the case where it is far from having the property. Although much of the recent work has focused on graphs, we outline some of our recent work on testing properties of relational structures. We introduce three generalized models for relational testing and use these models to consider the logical classification problem for testability, where we state one of our recent results: Ackermann's class with equality is testable.

1. Introduction

In property testing, we are given access to a large object such as a graph or database and wish to state some conclusion regarding whether this object has a desired property after examining only a small, random sample of the object.

Property testers are probabilistic approximation algorithms that examine only a small part of their input. Our goal is to distinguish inputs that have a desired property from inputs that are *far* from having it. We focus on *classification*, i.e., the testability of large classes of properties. There are several surveys of property testing, see e.g., Fischer [7] or Ron [13].

Property testing began in the context of program verification (see Blum *et al.* [4] and Rubinfeld and Sudan [14]). Goldreich *et al.* [9] extended the idea to testing properties of graphs, which much of the recent research has focused on. Alon and Shapira [3] have surveyed some of these recent results in graph testing.

We are interested in the *classification problem for testability*, i.e., in determining the testability of fragments of first-order logic. This line of research began with Alon *et al.* [2]. They proved that first-order sentences in the language of graphs with quantifier prefixes[1] matching the pattern $\exists^*\forall^*$ are testable, while there exist untestable sentences (in the language of graphs) with the prefix $\forall^*\exists^*$.

These prefix patterns are familiar from the classification problem for decidability; Ramsey [12] proved that validity is decidable for all first-order sentences without function symbols where the quantifier prefix matches $\exists^*\forall^*$. This result is *not* restricted to the language of graphs, and so there may be any number of distinct predicate symbols with possibly different (but finite) arities. Skolem [16] showed that the set of sentences with prefixes of the pattern $\forall^*\exists^*$, without the restriction to graphs, forms a reduction class (which then implies undecidability for the validity problem).

These similarities (and others, see Jordan [11]) naturally lead us to consider the classification problem for testability. However, a generalization of property testing to relational structures is required to consider the problem in a meaningful way. We present three possible such generalizations in this paper. Due to page limitations, we do not review classification problems or first-order logic in detail (see, e.g., Börger *et al.* [5] and Enderton [6] respectively).

This paper summarizes some of our recent results; for a longer version see Jordan [11] or Jordan and Zeugmann [10]. The summary is organized as follows. We briefly introduce necessary definitions in Section 2. Then, we briefly state recent results in Section 3. The major result described there is the testability of all relational properties expressible by sentences in Ackermann's class with equality, i.e., with quantifier prefixes of the form $\exists^*\forall\exists^*$.

2. Definitions

We seek a generalization of property testing for relational structures. We begin by defining vocabularies and (relational) structures.

Definition 1. *A* vocabulary τ *is a tuple of distinct predicate symbols* R_i *together with their arities* a_i,

$$\tau := (R_1^{a_1}, \dots, R_s^{a_s}) \ .$$

Definition 2. *A* structure A of type τ *is an* $(s+1)$*-tuple*

$$A := (U, \mathcal{R}_1^A, \dots, \mathcal{R}_s^A)$$

[†]Graduate School of Information Science and Technology, Hokkaido University. Supported by a Grant-in-Aid for JSPS Fellows.

[‡]Graduate School of Information Science and Technology, Hokkaido University.

[1]Without loss of generality, we assume all sentences are in prenex normal form.

consisting of a finite universe U and where each $\mathcal{R}_i^A \subseteq U^{a_i}$ *is a predicate corresponding to the predicate symbol* R_i *of* τ.

We identify U with the non-negative integers $\{0, \ldots, n-1\}$ and write $n = \#(A)$ for the size of the universe of a structure A. The set of all structures of type τ and universe size n is $STRUC^n(\tau)$ and the set of all structures of type τ is $STRUC(\tau) := \bigcup_{0 \le n} STRUC^n(\tau)$. A *property of type* τ is any subset of $STRUC(\tau)$. For $A \in P$, we say *A has P*.

Our goal is to distinguish structures that have a given property from structures that are far from having it. This requires a distance measure; we provide the following variations, where $\oplus$ denotes exclusive-or.

Definition 3. *Let* $A, B \in STRUC(\tau)$ *be any structures such that* $\#(A) = \#(B) = n$. *The* distance *between structures A and B is*

$$\text{dist}(A,B) := \frac{\sum_{1 \le i \le s} |\{\mathbf{x} \mid \mathbf{x} \in U^{a_i} \text{ and } \mathcal{R}_i^A(\mathbf{x}) \oplus \mathcal{R}_i^B(\mathbf{x})\}|}{\sum_{i=1}^{s} n^{a_i}} .$$

Definition 3 is based on the Hamming distance between (natural) representations of relational structures as binary strings. However, the number of high-arity tuples asymptotically dominates the number of low-arity tuples. The importance of relations that are not of maximal arity is then diminished to (nearly) zero. Definition 3 essentially weights each *tuple* equally, while the next definition essentially weights each *relation* equally.

Definition 4. *Let* $A, B \in STRUC^n(\tau)$ *be structures. Then, the* r-distance *is*

$$\text{rdist}(A,B) := \max_{1 \le i \le s} \frac{|\{\mathbf{x} \mid \mathbf{x} \in U^{a_i} \text{ and } \mathcal{R}_i^A(\mathbf{x}) \oplus \mathcal{R}_i^B(\mathbf{x})\}|}{n^{a_i}} .$$

Fischer *et al.* [8] gave a model that is roughly equivalent to that resulting from Definition 4. Definition 4 equally weights each relation, but tuples with repeated elements are still dominated. For example, the number of possible loops in a graph is asymptotically insignificant compared to the number of non-loops.

The mrdist definition resolves this and is similar to rdist, with the addition that it treats each *subtype*[2] of each relation as a separate relation. Full details are omitted, see Jordan [11].

For each of these definitions, we build a model for property testing. The following definitions are for the dist definition; the remaining two cases are analogous.

[2]For example, tuples of the form (x, x, y) and (x, y, x) belong to two different *binary* subtypes of a ternary relation.

Definition 5. *Let P be a property of structures with vocabulary* τ *and let A be such a structure with a universe of size n. Then,*

$$\text{dist}(A, P) := \min_{A' \in P \cap STRUC^n(\tau)} \text{dist}(A, A') .$$

Definition 6. *An* ε-tester *for property P is a randomized algorithm given an oracle which answers queries for the universe size and truth values of relations on desired tuples in a structure A. The tester must accept with probability at least 2/3 if A has P and must reject with probability at least 2/3 if* $\text{dist}(A, P) \ge \varepsilon$.

Definition 7. *Property P is* testable *if for all* $\varepsilon > 0$ *there are* ε*-testers making a number of queries which is upper-bounded by a function depending only on* ε.

Definition 7 is non-uniform in the sense that the ε-testers may not be computable given ε. The addition of a uniformity condition results in uniform testability. Our results hold in both cases and so we do not distinguish between them. Our logic is a pure predicate logic with equality. We omit all definitions related to logic and classification.

3. Recent Results

We denote the sets of properties that are testable under the dist, rdist and mrdist definitions with $\mathcal{T}_d$, $\mathcal{T}_r$ and $\mathcal{T}_{mr}$ respectively.

Theorem 1. *The following* strict *inclusions hold;*

$$\mathcal{T}_{mr} \subset \mathcal{T}_r \subset \mathcal{T}_d .$$

That is, testing with the mrdist definition is strictly harder than testing with the rdist definition, which in turn is strictly harder than with the dist definition. For a proof see Jordan [11].

However, there exist certain classes of properties, including Ackermann's class with equality, where testability in the rdist and mrdist senses coincide in a nice way. Continuing a parallel with the classical decidability results, we have shown the following.

Theorem 2. *All properties definable by first-order sentences in prenex normal form, with equality, where the quantifier prefix is of the form* $\exists^*\forall\exists^*$ *are in* $\mathcal{T}_{mr}$.

This class consists of all first-order sentences in prenex normal form, where there is at most one universal quantifier and any number of existential quantifiers. The sentences may contain equality and any number of predicate symbols of any (finite) arities but must not contain function symbols or arithmetic relation symbols like *PLUS*.

Given the inclusions of Theorem 1, Ackermann's class with equality is testable in all three variations of relational testability and implies the corresponding results in dense graph and hypergraph testing. A proof of a weaker result is in Jordan [11], while a proof of Theorem 2 is in Jordan and Zeugmann [10].

Ackermann [1] proved the decidability of this class without equality. Adding equality and a unary function symbol results in Shelah's class, which Shelah [15] proved decidable. Shelah's class is a decidable class that does not have the finite model property, and it would be interesting to consider its testability.

4. Conclusion

We have introduced a generalization of (dense) graph property testing, which we call *relational property testing*. In particular, we focused on three possible distance measures and showed the relationship between the resulting models of testing.

Relational databases are an example of massive relational structures and it would be interesting to consider applications of testability to such databases. Relational property testing is a natural way to characterize this type of problem. In addition, properties of databases are often given as queries written in formal languages such as SQL. It is then natural to consider the testability of syntactic restrictions of formal languages. This type of classification problem is our primary topic.

Our major result is the testability of all properties expressible in Ackermann's class with equality (Theorem 2). This provides an additional similarity between the known classification for testability and the classical one for decidability.

References

[1] Wilhelm Ackermann. Über die Erfüllbarkeit gewisser Zählausdrücke. *Math. Annalen*, 100:638–649, 1928.

[2] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[3] Noga Alon and Asaf Shapira. Homomorphisms in graph property testing. In M. Klazar, J. Kratochvíl, M. Loebl, J. Matoušek, R. Thomas, and P. Valtr, editors, *Topics in Discrete Mathematics*, volume 26 of *Algorithms and Combinatorics*, pages 281–313. Springer, 2006.

[4] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. of Comput. Syst. Sci.*, 47(3):549–595, 1993.

[5] Egon Börger, Erich Grädel, and Yuri Gurevich. *The Classical Decision Problem*. Springer-Verlag, 1997.

[6] Herbert B. Enderton. *A Mathematical Introduction to Logic*. Academic Press, second edition, 2000.

[7] Eldar Fischer. The art of uninformed decisions. *Bulletin of the European Association for Theoretical Computer Science*, 75:97–126, October 2001. Columns: Computational Complexity.

[8] Eldar Fischer, Arie Matsliah, and Asaf Shapira. Approximate hypergraph partitioning and applications. *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2007*, pages 579–589, 2007.

[9] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[10] Charles Jordan and Thomas Zeugmann. Relational properties expressible with one universal quantifier are testable. To appear in *Fifth Symposium on Stochastic Algorithms, Foundations and Applications, SAGA 2009, Proceedings*. Lecture Notes in Computer Science, 2009.

[11] Charles Harold Jordan. Master's thesis: The classification problem in relational property testing. Technical Report TCS-TR-B-09-6, Hokkaido University, Division of Computer Science, April 2009.

[12] F. P. Ramsey. On a problem of formal logic. *Proc. London Math. Soc. (2)*, 30:264–286, 1930.

[13] Dana Ron. Property testing. In Sanguthevar Rajasekaran, Panos M. Pardalos, John H. Reif, and José Rolim, editors, *Handbook of Randomized Computing*, volume II, chapter 15, pages 597–649. Kluwer Academic Publishers, 2001.

[14] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[15] Saharon Shelah. Decidability of a portion of the predicate calculus. *Israel J. Math.*, 28(1-2):32–44, 1977.

[16] Th. Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen. *Videnskapsselskapets skrifter, I. Matematisk-naturvidenskabelig klasse*, (4):1–26, 1920.