Testing Juntas

Eric Blais
University of Waterloo, Waterloo, ON, Canada

Years aud Authors of Summarized Original Work

2004; Fischer, Kindler, Ron, Safra, Samorodnitsky
2009; Blais

Keywords

Juntas; Property testing; Sublinear-time algorithms; Dimension reduction

Problem Definition

Fix positive integers n and k with $n \geq k$. The function $f : \{0,1\}^n \to \{0,1\}$ is a *k-junta* if it depends on at most k of the input coordinates. Formally, f is a k-junta if there exists a set $J \subseteq \{1, 2, \ldots, n\}$ of size $|J| \leq k$ such that for all inputs $x, y \in \{0,1\}^n$ that satisfy $x_i = y_i$ for each $i \in J$, we have $f(x) = f(y)$. Juntas play an important role in different areas of computer science. In machine learning, juntas provide an elegant framework for studying the problem of learning with datasets that contain many irrelevant attributes [9; 10]. In the analysis of Boolean functions, they essentially capture the set of functions of low complexity under natural measures such as total influence [19] and noise sensitivity [12].

How efficiently can we distinguish k-juntas from functions that are far from being k-juntas? We can formalize this question in the setting of property testing. Define the distance between two functions $f, g : \{0,1\}^n \to \{0,1\}$ to be the fraction of inputs on which f and g take different values: $\mathrm{dist}(f, g) := \frac{1}{2^n}|\{x \in \{0,1\}^n : f(x) \neq g(x)\}$. When $\mathrm{dist}(f, g) \geq \epsilon$ for every k-junta g, we say that f is *ϵ-far* from being a k-junta; otherwise we say that f is *ϵ-close* to being a k-juntas. An *ϵ-test* for k-juntas is a randomized algorithm that queries the value of $f : \{0,1\}^n \to \{0,1\}$ on some of its inputs and then with probability at least $\frac{2}{3}$

1. accepts if f is a k-junta, and
2. rejects if f is ϵ-far from being a k-junta.

(The algorithm is free to output anything when f is not a k-junta but is ϵ-close to being a k-junta.)

Problem 1. What is the minimum number of queries to $f : \{0,1\}^n \to \{0,1\}$ required to ϵ-test if f is a k-junta?

Key Results

Testing 1-juntas. One important class of functions related to junta testing is *dictator* functions—the functions $f : \{0,1\}^n \to \{0,1\}$ of the form $f(x) = x_i$ for some $i \in [n]$. Bellare, Goldreich, and Sudan [3], in work that was stated in terms of testing the *long code* and part of their analysis of probabilistically-checkable proofs (PCPs), showed that dictator functions can be ϵ-tested with $O(1/\epsilon)$ queries. (See the Locally Testable

Codes entry for more details.) This result was later extended by Parnas, Ron, and Samorodnitsky [21]. The class of 1-juntas includes dictator functions, their negations (known as *anti-dictator* functions), and the constant functions; using the algorithms in [3; 21], we can test 1-juntas with $O(1/\epsilon)$ queries.

Testing k-juntas. The first result on testing k-juntas for values $k > 1$ followed from related work on the problem of *learning* juntas. Blum, Hellerstein, and Littlestone [11] introduced an algorithm that queries a k-junta $f : \{0,1\}^n \to \{0,1\}$ on $O(k \log n + k/\epsilon + 2^k)$ inputs and with probability at least $\frac{5}{6}$ returns a k-junta $h : \{0,1\}^n \to \{0,1\}$ such that $\mathrm{dist}(f,h) \leq \epsilon$. Shortly afterwards, Goldreich, Goldwasser, and Ron [20] gave a general reduction showing that a proper learning algorithm with query complexity q for a class C of functions can be used to ϵ-test the class C with $q + O(1/\epsilon)$ queries. This result, combined with the Blum–Hellerstein–Littlestone algorithm, shows that k-juntas can be tested with $O(k \log n + 2^k + 1/\epsilon)$ queries.

Fischer, Kindler, Ron, Safra, and Samorodnitsky [18] showed that, remarkably, it is possible to test k-juntas with a number of queries that is *independent* of n. Specifically, they introduced ϵ-tests for k-juntas with query complexity $O(k^2/\epsilon^2)$. This result was sharpened in [4; 5], leading to the following theorem.

Theorem 1 ([5]). *It is possible to ϵ-test if $f : \{0,1\}^n \to \{0,1\}$ is a k-junta with $O(k \log k + k/\epsilon)$ queries.*

Chockler and Gutfreund [16] showed that $\Omega(k)$ queries are required to test k-juntas, so the bound in Theorem 1 is nearly optimal. (See also [4; 7; 13] for related lower bounds.)

Theorem 1 can be generalized to apply to the setting where $X_1, \ldots, X_n$, and Y are arbitrary finite sets and we wish to test whether a function $f : X_1 \times \cdots \times X_n \to Y$ is a k-junta. Interestingly, the query complexity of the k-junta test remains unchanged in this general setting as well. See [5] for the details.

Junta-testing algorithm. The proof of Theorem 1 contains two main ingredients.

The first ingredient is a simple modification of the Blum–Hellerstein–Littlestone learning algorithm. The original learning algorithm proceeds in two stages: first, the algorithm learns the k relevant coordinates of the junta; then, it queries f for all 2^k different values of the k relevant coordinates. When we test k-juntas, the second stage is unnecessary and can be replaced with a simpler test that checks whether the (at most) k relevant coordinates that have been identified completely determine the value of f or not. With this modification, we obtain an ϵ-test for k-juntas with query complexity $O(k \log n + k/\epsilon)$. Note that this result already yields the desired bound in Theorem 1 when $n = \mathrm{poly}(k)$.

The second ingredient in the proof of Theorem 1 is a dimension reduction argument. Consider a random partition of the n coordinates into $m = \mathrm{poly}(k)$ parts $S_1, \ldots, S_m$. A function $f : \{0,1\}^n \to \{0,1\}$ is isomorphic to a function $f' : X_1 \times \cdots \times X_m \to \{0,1\}$ where $X_i = \{0,1\}^{|S_i|}$. The function f' is defined over a domain with much smaller dimension and it satisfies two useful properties. First, when f is a k-junta, then so is f'. Second, when f is ϵ-far from k-juntas and $m = \Omega(k^2)$, then with high probability f' is $\frac{\epsilon}{2}$-far from k-juntas as well. The second fact is far from obvious. It was established in [5] using Fourier analysis and in [8] using a combinatorial argument. These two properties let us complete the algorithm for testing k-juntas by applying the modified Blum–Hellerstein–Littlestone algorithm on the function f'. More details on the algorithm itself can be found in the original papers [18; 5] and the survey [6].

Applications

Feature Selection

Feature selection is the general machine learning task of identifying the features (also known as *attributes* or *variables*) in a dataset that suffice to describe the model being studied. This task is formalized within the junta framework as follows: given a function $f : \{0,1\}^n \to \{0,1\}$, the algorithm seeks to identify a set $J \subseteq [n]$ of size $|J| = k$ where (i) k is as small as possible, and (ii) there is a k-junta $h : \{0,1\}^n \to \{0,1\}$ on the set J that is close to f.

The junta testing algorithm can be used to approximate the minimal value of k for which these two conditions can be satisfied. For example, by executing the junta testing algorithm with $k = 1, 2, 4, 8, \ldots$ until it accepts, we obtain the following estimation result.

Corollary 1. *There is an algorithm that, given query access to $f : \{0,1\}^n \to \{0,1\}$, outputs an estimate $\widehat{k}$ such that f is ϵ-close to a k-junta and such that f is not an ℓ-junta for any $\ell < k/2$. Furthermore, this algorithm makes $O(k \log k + k/\epsilon)$ queries to f.*

Testing by Implicit Learning

Let $\mathcal{C}$ be any class (i.e., family) of Boolean functions where every function in $\mathcal{C}$ is close to a being a k-junta. Many natural classes of Boolean functions that have been studied in learning theory and computational complexity fall into this framework. For example, functions with bounded, decision tree complexity, DNF complexity, circuit complexity, and sparse polynomial representation all satisfy this condition. (See the) Diakonikolas et al. [17] gave a general result showing that for each of these classes $\mathcal{C}$, we can ϵ-test the property of being in the class $\mathcal{C}$ efficiently. This result has since been sharpened by Chakraborty et al. [14], yielding the following bounds.

Theorem 2 ([14]). *Fix $s > 0$ and $\epsilon > 0$. We can ϵ-test whether $f : \{0,1\}^n \to \{0,1\}$ can be represented by*

1. *a DNF with s terms,*
2. *a size-s Boolean formula,*
3. *an s-sparse polynomial over $\mathbb{F}_2^n$, or*
4. *a decision tree of size s*

with $O(s/\epsilon^2 \cdot \mathrm{polylog}(s/\epsilon))$ queries.

The proof of Theorem 2 is remarkable in that the ϵ-test algorithm in [17; 14] learns the function $f : \{0,1\}^n \to \{0,1\}$ when f is a k-junta, but without identifying which of the k coordinates of f are part of the junta. This technique is called *testing by implicit learning*, and it is obtained by using and building on the junta testing algorithm.

Testing Function Isomorphism

Two functions $f, g : \{0,1\}^n \to \{0,1\}$ are *isomorphic* to each other when they are identical up to relabeling of the input variables. In the *function isomorphism testing* problem, we are given query access to (an unknown function) f and must determine whether it is isomorphic to (the known function) g or whether it is ϵ-far from being so. How many queries to f do we need to perform this task? The answer, it turns out, depends on the choice of the function g. The functions g for which we can test isomorphism to g with a constant number of queries are called *efficiently isomorphism-testable.*

Every symmetric function is efficiently isomorphism-testable. Using the junta testing algorithm, Fischer et al. [18] showed that for any constant $k \geq 0$, every k-junta is also efficiently isomorphism-testable. An important open problem in property testing is to characterize the set of functions that are efficiently isomorphism-tesable. The state of the art on this question is a recent result—also building on the junta testing algorithm—showing that every partially symmetric function is also efficiently isomorphism-testable. A function $f : \{0,1\}^n \to \{0,1\}$ is *k-partially symmetric* if there is a function $g : \{0,1\}^k \times \{0,1,2,\ldots,n\} \to \{0,1\}$ and a mapping $\rho : [k] \to [n]$ such that $f(x) = g(x_{\rho(1)}, \ldots, x_{\rho(k)}, \|x\|)$ where $\|x\| = \sum_i x_i$ the Hamming weight of x.

Theorem 3 ([8; 15]). *For every constant $k \geq 0$, every k-partially symmetric function is efficiently isomorphism-testable.*

Open Problems

There are two particularly appealing open problems related to the junta testing problem that are motivated by its application to the feature selection problem.

Distance approximation. Theorem 1 shows that we can distinguish k-juntas from functions that are ϵ-far from k-juntas with few queries. Can we also approximate the distance of a function to its closest k-junta with a small number of queries?

Problem 2. What is the minimum number of queries to $f : \{0,1\}^n \to \{0,1\}$ required to approximate the distance of f to its closest k-junta within an additive error of $\pm\epsilon$, where $\epsilon \in [0, \frac{1}{2}]$ is a parameter given to the algorithm?

In some cases, property testing algorithms can also be used directly for the corresponding distance approximation problem. This is the case, for example, for the BLR linearity test in the Testing Linearity chapter. But it is currently not known whether the junta testing algorithms in [18] or [5] can be extended to yield distance approximators or not.

Testing with random samples. The query model we have discussed throughout this chapter—where the algorithm is free to query the target function on any input of its choosing—is known as the *membership query model* in machine learning. In some applications, however, we must consider weaker query models where we restrict the queries that the algorithm can make in some ways. Can we also test k-juntas efficiently in restricted query models?

Problem 3. In which restricted query models can we test whether $f : \{0,1\}^n \to \{0,1\}$ is a k-junta with a number of queries that is asymptotically smaller than the number of queries required to learn k-juntas in the same settings?

Two examples of restricted query models include the passive sampling model (where each query is drawn independently at random from some fixed distribution) and the active query model (where the algorithm can choose its queries from a larger set of inputs drawn from some distribution). Some initial results on this problem can be found in [2; 1].

Cross-References

Testing Linearity; Testing by Implicit Learning.

Recommended Reading

1. Alon N, Hod R, Weinstein A (2013) On active and passive testing. arXiv preprint arXiv:13077364
2. Balcan MF, Blais E, Blum A, Yang L (2012) Active property testing. In: Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on, IEEE, pp 21–30
3. Bellare M, Goldreich O, Sudan M (1998) Free bits, PCPs, and nonapproximability—towards tight results. SIAM J Comput 27(3):804–915
4. Blais E (2008) Improved bounds for testing juntas. In: Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques, Springer, pp 317–330
5. Blais E (2009) Testing juntas nearly optimally. In: STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing, ACM, New York, pp 151–157
6. Blais E (2010) Testing juntas: A brief survey. In: Goldreich O (ed) Property Testing - Current Research and Surveys, pp 32–40
7. Blais E, Brody J, Matulef K (2012) Property testing lower bounds via communication complexity. Comput Complexity 21(2):311–358
8. Blais E, Weinstein A, Yoshida Y (2012) Partially symmetric functions are efficiently isomorphism-testable. In: Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on, pp 551–560
9. Blum A (1994) Relevant examples and relevant features: thoughts from computational learning theory. In: AAAI Fall Symposium on 'Relevance'
10. Blum A, Langley P (1997) Selection of relevant features and examples in machine learning. Artificial Intelligence 97(2):245–271
11. Blum A, Hellerstein L, Littlestone N (1995) Learning in the presence of finitely or infinitely many irrelevant attributes. J of Comp Syst Sci 50(1):32–40
12. Bourgain J (2002) On the distribution of the fourier spectrum of boolean functions. Israel Journal of Mathematics 131(1):269–276
13. Buhrman H, García-Soriano D, Matsliah A, de Wolf R (2013) The non-adaptive query complexity of testing k-parities. Chic J Theoret Comput Sci pp Article 6, 11
14. Chakraborty S, García-Soriano D, Matsliah A (2011) Efficient sample extractors for juntas with applications. In: Automata, Languages and Programming, Springer, pp 545–556
15. Chakraborty S, Fischer E, García-Soriano D, Matsliah A (2012) Junto-symmetric functions, hypergraph isomorphism, and crunching. In: 2012 IEEE 27th Conference on Computational Complexity—CCC 2012, IEEE Computer Soc., Los Alamitos, CA, pp 148–158
16. Chockler H, Gutfreund D (2004) A lower bound for testing juntas. Information Processing Letters 90(6):301–305
17. Diakonikolas I, Lee HK, Matulef K, Onak K, Rubinfeld R, Servedio RA, Wan A (2007) Testing for concise representations. In: Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on, IEEE, pp 549–558
18. Fischer E, Kindler G, Ron D, Safra S, Samorodnitsky A (2004) Testing juntas. J Comput System Sci 68(4):753–787
19. Friedgut E (1998) Boolean functions with low average sensitivity depend on few coordinates. Combinatorica 18(1):27–35
20. Goldreich O, Goldwasser S, Ron D (1998) Property testing and its connection to learning and approximation. J ACM 45(4):653–750
21. Parnas M, Ron D, Samorodnitsky A (2002) Testing basic boolean formulae. SIAM J Discrete Math 16(1):20–46