Hölder Norms and a Hierarchy Theorem for Parameterized Classes of CCG

Christophe Costa Florêncio[1] and Henning Fernau[2]

[1] Department of Computer Science, K.U. Leuven, Leuven, Belgium
Chris.CostaFlorencio@cs.kuleuven.be
[2] Universität Trier, FB IV—Abteilung Informatik, D-54286 Trier, Germany
fernau@uni-trier.de

Abstract. We develop a framework based on Hölder norms that allows us to easily transfer learnability results. This idea is concretized by applying it to Classical Categorial Grammars (CCG).

1 Introduction

The question to present new, potentially interesting language classes that can be learnt in certain scenarios has been always in the focus of Grammatical Inference. To the knowledge of the authors, no systematic way to design or define new learnable language classes, starting out from known learnability results, is known, possibly apart from the results exhibited in [1, 2, 3] where it is shown how formal language results can be used to transfer learnability results by using a kind of preprocessing reduction. We will here show another way of such a transfer, namely by using elementary mathematical properties of Hilbert spaces. This works in particular for learning from text, also known as learning from positive data, a setting where such transfer results have been hitherto unknown.

One way of obtaining learnable classes of grammars is by imposing bounds on their descriptive complexity. This approach has been applied in the past to the Classical Categorial Grammar (CCG) formalism [4, 5] resulting in, among others, the learnable class of k-valued grammars, the grammars of which assign at most k categorial types to any single symbol in the (fixed) alphabet.[1] Using ideas exhibited in [6], we might first associate to any categorial grammar G over an n-letter alphabet $\{a_1, \ldots, a_n\}$ an n-dimensional vector $\boldsymbol{v}(G)$, where the ith component counts that number of categorial types associated to a_i. From this viewpoint, a categorial grammar G is k-valued iff the maximum norm (also known as L_∞ norm) of $\boldsymbol{v}(G)$ is bounded by k. We will therefore term the corresponding class k-max-valued in what follows. This association of G to a normed Hilbert space poses a natural mathematical question: How do formal language (hierarchy) and learnability results transfer when changing the norm?

[1] For reasons of space, we are not giving definitions of CCG here, but refer to the given references. In particular, range(G) denotes the types occurring in the CCG G and Pr is the set of primitive types.

J.M. Sempere and P. García (Eds.): ICGI 2010, LNAI 6339, pp. 280–283, 2010.

2 *k*-Sum-Valued CCG

We first study the sum norm: a categorial grammar G is k-sum-valued iff the sum norm (also known as L_1 norm) of $\boldsymbol{v}(G)$ is bounded by k.

For a crisp notation, let $\mathcal{G}_{k-\text{max-val}}(\Sigma)$, $\mathcal{L}_{k-\text{max-val}}(\Sigma)$, and $\mathcal{FL}_{k-\text{max-val}}(\Sigma)$ denote that class of k-max-valued grammars, their languages and functor-argument structure languages, respectively; sometimes, the basic alphabet Σ is made explicit. Similarly, the notations $\mathcal{G}_{k-\text{sum-val}}(\Sigma)$, $\mathcal{L}_{k-\text{sum-val}}(\Sigma)$, and $\mathcal{FL}_{k-\text{sum-val}}(\Sigma)$ are understood.

Kanazawa showed that both the hierarchies $\mathcal{L}_{k-\text{max-val}}(\Sigma)$ and $\mathcal{FL}_{k-\text{max-val}}(\Sigma)$ have finite elasticity, hence entailing text learnability.

Theorem 1. *Fix some finite alphabet Σ. Families from both the hierarchies $\mathcal{L}_{k-sum\text{-}val}(\Sigma)$ and $\mathcal{FL}_{k-sum\text{-}val}(\Sigma)$ have finite elasticity.*

Proof. The maximum number of types that can be assigned to any symbol in a k-max-valued grammar is k. Thus, $\mathcal{G}_{k-\text{sum-val}}(\Sigma) \subset \mathcal{G}_{k-\text{max-val}}(\Sigma)$, and $\mathcal{L}_{k-\text{sum-val}}(\Sigma) \subseteq \mathcal{L}_{k-\text{max-val}}(\Sigma)$.

Since $\mathcal{L}_{k-\text{max-val}}$ is known to have finite elasticity for every k, every subclass, including $\mathcal{L}_{k-\text{sum-val}}$, has finite elasticity (for every k), and is thus learnable.

An analogous argument can be given for $\mathcal{FL}_{k-\text{sum-val}}$. □

This yields the easy corollary:

Corollary 2. *Fix some finite alphabet Σ. Both families from the hierarchies $\mathcal{L}_{k-sum\text{-}val}(\Sigma)$ and $\mathcal{FL}_{k-sum\text{-}val}(\Sigma)$ are identifiable in the limit.*

For the maximum norm, a hierarchy theorem was proved in [4], i.e., for any $k \geq 1$, $\mathcal{L}_{k-\text{max-val}} \subsetneq \mathcal{L}_{(k+1)-\text{max-val}}$. Such a result also holds for the k-sum-val grammars, as we will now demonstrate.

Theorem 3. *For any $k \geq 1$, $\mathcal{L}_{k-sum\text{-}val} \subsetneq \mathcal{L}_{(k+1)-sum\text{-}val}$.*

Proof. This proof parallels Theorem 5.5 as presented in Section 8.1.1 of [4]: By definition, $\mathcal{G}_{k-\text{sum-val}} \subset \mathcal{G}_{(k+1)-\text{sum-val}}$, which immediately implies $\mathcal{L}_{k-\text{sum-val}} \subseteq \mathcal{L}_{(k+1)-\text{sum-val}}$. It remains to show that $\mathcal{L}_{(k+1)-\text{sum-val}} - \mathcal{L}_{k-\text{sum-val}} \neq \emptyset$.

Let $L_n = \{a^i \mid 1 \leq i \leq n\}$, and $\mathcal{L}_n = \{L_n \mid n \in \mathbb{N}^+\}$. Note that $\mathcal{L}_n$ is an infinite ascending chain, so, since $\mathcal{L}_{k-\text{sum-val}}$ has finite elasticity (Theorem 1), $\mathcal{L}_n \nsubseteq \mathcal{L}_{k-\text{sum-val}}$ for any fixed k. Thus, for every k, there is an n such that $L_n \notin \mathcal{L}_{k-\text{sum-val}}$. We will show that for the least such n, $L_n \in \mathcal{L}_{(k+1)-\text{sum-val}}$. Let G_{n-1} be a k-sum-valued grammar such that $\mathrm{L}(G_{n-1}) = L_{n-1}$. There must exist a type $A \in \mathrm{range}(G_{n-1}) - \mathrm{Pr}$ such that A is not a proper subtype of any type in $\mathrm{range}(G_{n-1})$. Let $B = \underbrace{(\ldots(t/A)/\ldots)/A}_{n}$, and let $G_n = G_{n-1} \cup \{\langle a, B\rangle\}$. The reader can verify that $\mathrm{L}(G_n) = L_n$. □

From this proof, it immediately follows that:

Theorem 4. *For any $k \geq 1$, $\mathcal{FL}_{k-sum\text{-}val} \subsetneq \mathcal{FL}_{(k+1)-sum\text{-}val}$.*

3 Hölder Norms

The bounds defining the k-max-valued and k-sum-valued classes are special cases of *Hölder norms*. These are of the form $\| \boldsymbol{x} \|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$ for $\boldsymbol{x} \in \mathbb{R}^n$ and $p \geq 1$. For $p = 1$, this is the sum bound, for $p \to \infty$ this is the maximum bound, and for $p = 2$ we obviously obtain the Euclidian distance bound.

Proposition 5 (Folklore).

1. $\| \boldsymbol{x} \|_\infty \leq \| \boldsymbol{x} \|_p$ *for any* $p \geq 1$.
2. *For 1-dimensional spaces,* $\| \boldsymbol{x} \|_\infty = \| \boldsymbol{x} \|_p$ *for any* $p \geq 1$.

Thus, natural questions present themselves: Do learnability and hierarchy theorems hold for every class of CCGs defined in terms of Hölder norms? More specifically, let $\mathcal{G}_{H(k,p)}(\Sigma)$ collect those grammars G over the alphabet Σ which satisfy $\| \boldsymbol{v}(G) \|_p \leq k$. Accordingly, the notations $\mathcal{L}_{H(k,p)}(\Sigma)$ and $\mathcal{FL}_{H(k,p)}(\Sigma)$ are understood. By definition, we have:

Lemma 6. *For every* $k, p \geq 1$, $\mathcal{G}_{H(k,p)} \subseteq \mathcal{G}_{H(k+1,p)}$.

Proposition 5 (1) entails, completely analogous to Theorem 1:

Theorem 7. *Fix some finite alphabet* Σ *and some* $p \geq 1$. *Families from both the hierarchies* $\mathcal{L}_{H(k,p)}(\Sigma)$ *and* $\mathcal{FL}_{H(k,p)}(\Sigma)$ *have finite elasticity.*

Theorem 8. *For every* $k, p \geq 1$, $\mathcal{L}_{H(k,p)}(\Sigma) \subsetneq \mathcal{L}_{H(k+1,p)}(\Sigma)$.

Proof. From Lemma 6, it immediately follows that $\mathcal{L}_{H(k,p)} \subseteq \mathcal{L}_{H(k+1,p)}$. Since the strictness (for the case $p = 1$) was shown in Theorem 3 by a unary example, based on the finite elasticity of that class, Proposition 5 (2), together with Theorem 7 yield the claim. □

From this proof, it immediately follows that:

Theorem 9. *For every* $k, p \geq 1$, $\mathcal{FL}_{H(k,p)}(\Sigma) \subsetneq \mathcal{FL}_{H(k+1,p)}(\Sigma)$.

More generally, if $\mathcal{L}_k$ now denotes any language class that has finite elasticity and whose definition is obtained by restricting the L_∞ norm of some vector $\boldsymbol{v}(G)$ associated to some grammar G for $L \in \mathcal{L}_k$, then, for any $p \geq 1$, $\mathcal{L}_{k,p}$ is a hierarchy with finite elasticity, as well, where $\mathcal{L}_{k,p}$ collects all languages that can be described by grammars G with $\| \boldsymbol{G} \|_p \leq k$.

4 Further Extensions and Questions

We have indicated a quite general framework that allows us to transfer learnability results within the scenario of identification in the limit. We exemplified this with results on classical categorial grammars. In passing, the interplay between learnability and formal languages allowed us to prove a hierarchy theorem based on previously shown learnability results (which seems to be a new way of argumentation).

Our reasoning was based on the notion of finite elasticity. Can such arguments be extended to similar notions like finite thickness? Can we find other interesting special instances of our approach? The learning algorithms underlying this framework are not very efficient in general (although they are in the special case of CCG). So, is there a good way of transferring efficient learnability?

To give a better known example: We might associate to any nondeterministic n-state automaton A an n-dimensional vector $\boldsymbol{v}(A)$, indicating in component i the amount of lookahead needed to disambiguate any nondeterminism in state s_i. Then, a deterministic automaton A is k-reversible [7] iff its reversal A^r satisfies: $\| \boldsymbol{v}(A^r) \|_\infty \leq k$.[2] From the viewpoint exhibited in this paper, it appears natural to study the learnability of classes of regular languages defined by the restriction $\| \boldsymbol{v}(A^r) \|_p \leq k$ for any $p \geq 1$.

In general terms, we believe that there are further quite interesting connections between the areas of Grammatical Inference and that of Descriptional Complexity that deserve further studies.

References

[1] Takada, Y.: A hierarchy of language families learnable by regular language learning. Information and Computation 123, 138–145 (1995)

[2] Fernau, H., Sempere, J.M.: Permutations and control sets for learning non-regular language families. In: Oliveira, A.L. (ed.) ICGI 2000. LNCS (LNAI), vol. 1891, pp. 75–88. Springer, Heidelberg (2000)

[3] Fernau, H.: Even linear simple matrix languages: formal language properties and grammatical inference. Theoretical Computer Science 289, 425–489 (2002)

[4] Kanazawa, M.: Learnable Classes of Categorial Grammars. PhD, CSLI (1998)

[5] Costa Florêncio, C., Fernau, H.: Finding consistent categorial grammars of bounded value: a parameterized approach. In: Dediu, A.-H., Fernau, H., Martín-Vide, C. (eds.) LATA 2010. LNCS, vol. 6031, pp. 202–213. Springer, Heidelberg (2010)

[6] Dassow, J., Fernau, H.: Comparison of some descriptional complexities of 0L systems obtained by a unifying approach. Information and Computation 206, 1095–1103 (2008)

[7] Angluin, D.: Inference of reversible languages. Journal of the ACM 29(3), 741–765 (1982)

[2] We omit here the particularity on the disambiguation of final states that can be treated analogously.