ON Case-Based Representability and Learnability of Languages*

Christoph Globig[1] and Steffen Lange[2]

[1]University of Kaiserslautern, P.O. Box 3049, D-67653 Kaiserslautern, Germany
globig@informatik.uni-kl.de
[2]HTWK Leipzig, P.O. Box 66, D-04251 Leipzig, Germany
steffen@informatik.th-leipzig.de

Abstract. Within the present paper we investigate case-based representability as well as case-based learnability of indexed families of uniformly recursive languages. Since we are mainly interested in case-based learning with respect to an arbitrary fixed similarity measure, case-based learnability of an indexed family requires its representability, first.
We show that every indexed family is case-based representable by positive and negative cases. If only positive cases are allowed the class of representable families is comparatively small. Furthermore, we present results that provide some bounds concerning the necessary size of case bases.
We study, in detail, how the choice of a case selection strategy influences the learning capabilities of a case-based learner. We define different case selection strategies and compare their learning power to one another. Furthermore, we elaborate the relations to Gold-style language learning from positive and both positive and negative examples.

1 Introduction

Case-based reasoning is currently a booming subarea of artificial intelligence. In case-based reasoning knowledge is represented by a collection of typical cases in the case base and a similarity measure, instead of using any form of rules or axioms, for example. It is widely accepted that this approach may be considered as an reasonable model of how human experts structure their knowledge. Within case-based reasoning, case-based learning as understood in [1] seems to be of particular interest.

There are three possibilities to improve the knowledge representation in a case-based learning system (cf. [5]). The system can

- store new cases in the case base or remove cases from the case base,
- change the measure of similarity,
- or change both the case base and the similarity measure.

* This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) within the project Ind-Cbl.

In [7] a formalization of case-based learning in an Inductive Inference manner has been introduced. As it turns out case-based learning algorithms are of remarkable power, if effective classifiers should be learned, for instance. This power mainly results from one source, namely the ability of the case-based learner to change the underlying similarity measure within the learning task arbitrarily. Thereby, all knowledge will be more or less directly encoded within the similarity measure, no matter which cases have been stored in the case base. In order to overcome such undesirable encoding tricks we investigate case-based learning under the assumption that the underlying similarity measure cannot be changed during the whole learning task.

On a first glance, this approach seems to be too restrictive. Nevertheless, the results in [8] witness that also under this assumption interesting classes of formal languages are case-based learnable. Since we are mainly interested in investigating the problem of how the choice of a case selection strategy influences the learning capabilities of case-based learners, the above assumption seems to be particularly tailored.

In the sequel we confine ourselves to learning of indexed families of formal languages. Because of the underlying assumption that a case-based learner is not allowed to change its measure of similarity, case-based learnability of an indexed family requires its representability, first. In Section 3 we show that every indexed family is case-based representable by positive and negative cases. If positive cases are allowed, only, the class of representable families is comparatively small. Furthermore, the minimal size of case bases is discussed.

If we have a fixed measure of similarity the learning capability of a case-based learning system depends on the strategy used to select the cases for the case base, only. Section 4 discusses the influence of the following properties a case selection strategy may or may not have:

Access to case history: Is the case selection strategy allowed to store any case that is already presented or has the strategy access to the last one, only?
Deleting cases from the case base: Is the case selection strategy allowed to delete cases from the case base or does the case base grow monotonically?

The different case selection strategies define different types of case-based learning. We elaborate relations between these types of case-based learning and relate them to Gold-style language learning from positive and both positive and negative examples.

2 Preliminaries

The definitions of this section are adapted from the Inductive Inference literature (cf. [2]). Our target objects are (formal) languages over a finite alphabet A. By A^+ we denote the set of all non-empty strings over the alphabet A. Any subset L of A^+ is called a language. We set $\overline{L} = A^+ \setminus L$.

By $\mathbb{N} = \{1, 2, \dots\}$ we denote the set of all natural numbers. Let $c : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ denote *Cantor's pairing function*. We use $\mathbb{Q}_{[0,1]}$ to denote the set of all rational numbers between 0 and 1. We write $B \subseteq_{fin} C$, if B is a finite subset of C. Furthermore, by $card(B)$ we denote the cardinality of set B.

There are two basic ways to present information about a language to a learner. We can present positive data only or positive and negative data. These presentations are called *text* and *informant*, respectively. A *text* for a language L is an infinite sequence $t = (s_1, b_1), (s_2, b_2), \dots$ $((s_j, b_j) \in A^+ \times \{+\})$ such that $\{s_j \mid j \in \mathbb{N}\} = L$. $t[k]$ is the initial sequence $(s_1, b_1), (s_2, b_2), \dots, (s_k, b_k)$ of t. We set $t^+[k] = \{s_j \mid j \leq k\}$. Let $text(L)$ denote the set of all texts of L. An *informant* for a language L is an infinite sequence $i = (s_1, b_1), (s_2, b_2), \dots$ $((s_j, b_j) \in A^+ \times \{+, -\})$ such that $\{s_j \mid j \in \mathbb{N},\ b_j = +\} = L$ and $\{s_j \mid j \in \mathbb{N},\ b_j = -\} = A^+ \backslash L$. $i[k]$ is the initial sequence $(s_1, b_1), (s_2, b_2), \dots, (s_k, b_k)$ of i. Furthermore, we set $i^+[k] = \{s_j \mid j \leq k,\ b_j = +\}$ and $i^-[k] = \{s_j \mid j \leq k,\ b_j = -\}$. By $informant(L)$ we denote the set of all informants of L. Without loss of generality we assume that $t[k]$ ($i[k]$) is a natural number that represents the initial segment of the text (resp. informant).

We restrict ourselves to investigate the learnability of indexed families of recursive languages over A (cf. [2]). A sequence $\mathcal{L} = L_1, L_2, \dots$ is said to be an *indexed family* if all L_j are non-empty and there is a recursive function f such that for all indices j and all strings $w \in A^+$ holds

$$f(j, w) = \begin{cases} 1 & \text{if } w \in L_j \\ 0 & \text{otherwise} \end{cases}$$

So given an indexed family $\mathcal{L}$ the membership problem is uniformly decidable for all languages in $\mathcal{L}$ by a single function.

IF denotes the set of all indexed families.

The following definition is adapted from [2]. We use $f(x)\downarrow$ to denote that a function f is defined on input x.

Definition 1. Let $\mathcal{L} \in \mathbf{IF}$.
Then we say $\mathcal{L}$ is *learnable from text* (resp. *learnable from informant*) iff
$\exists M \in \mathbf{P}\ \forall L \in \mathcal{L}\ \forall t \in text(L)$ (resp. $\forall i \in informant(L)$)

(1) $\forall n \in \mathbb{N}\ M(t[n])\downarrow$ (*resp.* $\forall n \in \mathbb{N}\ M(i[n])\downarrow$),
(2) $\lim_{n\to\infty} M(t[n]) = a$ exists (*resp.* $\lim_{n\to\infty} M(i[n]) = a$ exists),
(3) $L_a = L$.

LIM.TXT (**LIM.INF**) is the set of all indexed families that are learnable from text (informant).

P denotes the set of the unary computable functions.

3 Case-Based Representability

This section summarizes the results obtained about case-based representability. First we have to define the language that is described by a set of cases and a similarity measure.[2] $\sigma : A^+ \times A^+ \to \mathbb{Q}_{[0,1]}$ is called a measure of similarity. Σ denotes the set of all totally defined and computable similarity measures. Let $\Sigma_{\{0,1\}} \subseteq \Sigma$ be the subset of all similarity measures that have the range $\{0, 1\}$.

We use the so called standard semantics L_{st} (cf. [8]).

Definition 2. Let $CB \subseteq_{fin} A^+ \times \{+,-\}$ and $\sigma \in \Sigma$ a similarity measure. Furthermore, let $CB^+ := \{s \mid (s,+) \in CB\}$, $CB^- := \{s \mid (s,-) \in CB\}$. Then we say CB and σ describe the language $L_{st}(CB,\sigma) = L_{st}(CB^+, CB^-, \sigma) := \{w \in A^+ \mid \exists c \in CB^+ \ (\sigma(c,w) > 0 \land \forall c' \in CB^- \sigma(c,w) > \sigma(c',w))\}$.

Definition 3. Let $\mathcal{L} \in \mathbf{IF}$ and $\sigma \in \Sigma$.
Then, $\mathcal{L} \in \mathbf{REPR}^+(\sigma)$ iff for every $L \in \mathcal{L}$ there is a $CB^+ \subseteq_{fin} L$ such that $L_{st}(CB^+, \emptyset, \sigma) = L$. Moreover, $\mathcal{L} \in \mathbf{REPR}^\pm(\sigma)$ iff for every $L \in \mathcal{L}$ there are $CB^+ \subseteq_{fin} L$ and $CB^- \subseteq_{fin} \overline{L}$ such that $L_{st}(CB^+, CB^-, \sigma) = L$.
Let $\mathbf{REPR}^+ := \bigcup_{\sigma\in\Sigma} \mathbf{REPR}^+(\sigma)$ and $\mathbf{REPR}^\pm := \bigcup_{\sigma\in\Sigma} \mathbf{REPR}^\pm(\sigma)$.

So $\mathcal{L} \in \mathbf{REPR}^+$ ($\mathcal{L} \in \mathbf{REPR}^\pm$) means that there is a σ such that $\mathcal{L} \in \mathbf{REPR}^+(\sigma)$ ($\mathcal{L} \in \mathbf{REPR}^\pm(\sigma)$).

If we allow only positive cases to be stored in the case base, we have the following lemma, which follows directly from Definition 2.

Lemma 4. *For any similarity measure $\sigma \in \Sigma$, there exists a measure $\sigma' \in \Sigma_{\{0,1\}}$ such that* $\mathbf{REPR}^+(\sigma) = \mathbf{REPR}^+(\sigma')$.

Applying the results of [8] we obtain:

Theorem 5. $\mathbf{REPR}^+ \subsetneq \mathbf{IF}$

In [8] it is proved that the family of pattern languages are not representable with positive cases.

On the other hand, it is possible to represent every indexed family if positive and negative cases can be stored within the case base.

Next, we introduce the concept of representative cases for languages. Let $L \subseteq A^+$, $w \in L$, and $\sigma \in \Sigma$. w is said to be a *representative case for L w.r.t. σ* provided that $\sigma(w,v) > 0$ iff $v \in L$. The notion of representative cases will be used subsequently in order to simplify some of the proofs.

[2] These definitions are adapted from [7].

Theorem 6. $\mathbf{REPR}^{\pm} = \mathbf{IF}$

Proof. Let $\mathcal{L} = L_1, L_2, \dots$ be an indexed family. Instead of $\mathcal{L}$ we use another enumeration $\tilde{\mathcal{L}}$ which includes the range of $\mathcal{L}$. We set $\tilde{\mathcal{L}} := \tilde{L}_1, \tilde{L}_2, \dots$ such that for all $j \in \mathbb{N}$, $\tilde{L}_{2j} = L_j$ and $\tilde{L}_{2j-1} = A^+$. Furthermore, let $w_1, w_2, \dots$ be the lexicographical enumeration of the strings over A^+.

We define the following total recursive function $r : \mathbb{N} \to \mathbb{N}$. Initially, we set $r(1) = 1$. We proceed inductively. Let $i > 1$. We set $r(i) = j$, if j is the least index satisfying $w_i \in \tilde{L}_j$ and $r(k) \neq j$, for all $k < i$.

Since for all $j \in \mathbb{N}$, $\tilde{L}_{2j-1} = A^+$, r is indeed total recursive. If $r(i) = j$, then w_i is a representative case for $\tilde{L}_j$. Moreover, we can easily conclude:

Claim 1: For every $j \in \mathbb{N}$, if $\tilde{L}_j$ is infinite, then there is a $k \in \mathbb{N}$ such that $r(k) = j$.

Now, we use r to define the desired similarity measure σ. Let $k, j \in \mathbb{N}$.

$$\sigma(w_k, w_j) = \begin{cases} 1 & \text{if } w_k = w_j \\ 1 - \frac{1}{k+1} & \text{if } r(k) = l,\ w_j \in \tilde{L}_l \setminus \{w_k\} \\ 0 & \text{otherwise} \end{cases}$$

Claim 2: $\mathcal{L} \in \mathbf{REPR}^{\pm}(\sigma)$.

Let $j \in \mathbb{N}$. First, suppose L_j to be infinite. By Claim 1 there is an $i \in \mathbb{N}$ such that $r(i) = 2j$. Note that $\tilde{L}_{2j} = L_j$. Moreover, $w_i \in \tilde{L}_{2j}$ by construction. Since w_i is a representative case for $\tilde{L}_{2j}$, we obtain $L_{st}(\{w_i\}, \emptyset, \sigma) = \tilde{L}_{2j} = L_j$.

It remains to handle the case that L_j is finite. Now, let $z \in \mathbb{N}$ such that w_z is the maximal element in L_j. By Claim 1 there are infinitely many representative cases for A^+. Hence, there is a $k > z$ such that $r(k) = 2m - 1$ for some $m \in \mathbb{N}$. By the choice of k, $w_k \notin L_j = \tilde{L}_{2j}$. Furthermore, for all $w \in \tilde{L}_{2j}$ and all $v \in A^+ \setminus \{w\}$, it holds $\sigma(w_k, v) > \sigma(w, v)$. Thus, $L_{st}(\tilde{L}_{2j}, \{w_k\}, \sigma) = \tilde{L}_{2j} = L_j$.

Hence, the theorem is proved. □

As we have seen, every $\mathcal{L} \in \mathbf{REPR}^{+}$ can be represented by a measure from $\Sigma_{\{0,1\}}$. Recently, Billhardt (cf. [4]) has shown that this results remains valid, if case bases containing both positive and negative cases are admissible. The underlying idea is quite similar to that used in the definition of the representatives in the last proof.

Lemma 7. *For every indexed family $\mathcal{L} \in \mathbf{REPR}^{\pm}$, there exists a $\sigma \in \Sigma_{\{0,1\}}$ such that $\mathcal{L} \in \mathbf{REPR}^{\pm}(\sigma)$.*

Notice that Billhardt's as well as our proof mainly exploit the fact that a case base used in order to represent a language has not necessarily to be computable itself. If we assume that the finiteness of L_j is decidable for all $j \in \mathbb{N}$, then the case bases are computable.

Furthermore, from a practical point of view it seems to be rather natural to choose the corresponding case bases as small as possible. Applying the construction underlying the proof of Theorem 6, at least some finite languages will be represented by putting all their elements into the corresponding case base. As we will see, we can do better.

Theorem 8. *Let $\mathcal{L} \in \mathbf{IF}$. Then there exists a $\sigma \in \Sigma$ such that for every $L \in \mathcal{L}$, there are $CB^+ \subseteq L$ and $CB^- \subseteq \overline{L}$ with $card(CB^+) = 1$ and $card(CB^-) \leq 1$ such that $L_{st}(CB^+, CB^-, \sigma) = L$.*

Proof. We use a slightly modified version of the concept of representatives introduced in the proof of Theorem 6. Let $\mathcal{L} = L_1, L_2, \ldots$ be an indexed family.

In order to obtain the desired result the following similarity measure is used. Again, let $(w_i)_{i\in\mathbb{N}}$ be the lexicographical enumeration of the strings over A^+. Let $j, k \in \mathbb{N}$ and $v \in A^+$.

Case 1: $w_{c(j,k)} \in L_j$.

Then we set:

$$\sigma(w_{c(j,k)}, v) = \begin{cases} 1 & \text{if } w_{c(j,k)} = v \\ 1 - \frac{1}{c(j,k)+2} & v \in L_j \setminus \{w_{c(j,k)}\} \\ \frac{1}{c(j,k)+2} & v \in \overline{L_j} \end{cases}$$

Case 2: $w_{c(j,k)} \notin L_j$.

Now we set:

$$\sigma(w_{c(j,k)}, v) = \begin{cases} 1 & \text{if } w_{c(j,k)} = v \\ 1 - \frac{1}{c(j,k)+2} & v \in \overline{L_j} \setminus \{w_{c(j,k)}\} \\ \frac{1}{c(j,k)+2} & v \in L_j \end{cases}$$

By definition, σ belongs to Σ. Moreover, for every $j, k \in \mathbb{N}$, the string $w_{c(j,k)}$ serves as a representative case for either L_j or $\overline{L_j}$. Now, let $j \in \mathbb{N}$. In order to show that every $L_j \in \mathcal{L}$ can be represented with respect to σ, we distinguish the following cases.

- Case 1: $L_j = A^+$.
  We simply choose $CB^+ = \{w_1\}$ and $CB^- = \emptyset$. By definition, $\sigma(w_1, v) > 0$ for all $v \in A^+$. Thus, $L_{st}(CB^+, CB^-, \sigma) = A^+$.
- Case 2: $L_j \neq A^+$.
  We proceed as follows. Choose any two strings $w_z, w_{\hat{z}}$ satisfying $w_z \in L_j$ and $w_{\hat{z}} \notin L_j$, respectively. Consider the string $w_{c(j,z+\hat{z})}$.
  - Subcase 2.1: $w_{c(j,z+\hat{z})} \in L_j$.
    By definition $w_{c(j,z+\hat{z})}$ is a representative case for L_j. Therefore, set $CB^+ = \{w_{c(j,z+\hat{z})}\}$ and $CB^- = \{w_{\hat{z}}\}$. By definition of σ we have for all $v \in L_j$, $\sigma(w_{c(j,z+\hat{z})}, v) > \sigma(w_{\hat{z}}, v)$ because $c(j, z+\hat{z}) > \hat{z}$ (cf. Case 1 in the definition of σ). On the other hand, applying the same argument yields $\sigma(w_{c(j,z+\hat{z})}, v) < \sigma(w_{\hat{z}}, v)$ for all $v \in \overline{L_j}$. Therefore, $L_{st}(CB^+, CB^-, \sigma) = L_j$.

- Subcase 2.2: $w_{c(j,z+\hat{z})} \notin L_j$.
  In contrast to Subcase 2.1, now $w_{c(j,z+\hat{z})}$ serves as a representative case for $\overline{L_j}$. Consequently, let $CB^+ = \{w_z\}$ and $CB^- = \{w_{c(j,z+\hat{z})}\}$. Finally, $L_{st}(CB^+, CB^-, \sigma) = L_j$ can be shown in a similar manner as above. We omit the details.

By construction we have $card(CB^+) = 1$ and $card(CB^-) \le 1$ in each of the discussed cases. This finishes the proof. □

Again, the corresponding case bases are not computable. Nevertheless, if we can assume that "$L_j = A^+$" is decidable for all $j \in \mathbb{N}$, the case bases themselves are computable as well.

4 Case-Based Learnability

Based on the representability results of the last section we now study case-based learnability of indexed families.

Definition 9. An indexed family $\mathcal{L}$ is said to be *case-based learnable from text* by the case selection strategy $S : \mathbb{N} \to A^+ \times \{+\}$ iff
$\exists \sigma \in \Sigma \; \forall L \in \mathcal{L} \; \forall t \in text(L)$

(1) $\forall n \in \mathbb{N} \; CB_n = S(t[n]) \downarrow$, and $S(t[n]) \subseteq t^+[n] \times \{+\}$,
(2) $CB = \lim_{n\to\infty} CB_n$ exists,
(3) $L_{st}(CB, \sigma) = L$.

Definition 10. An indexed family $\mathcal{L}$ is said to be *case-based learnable from informant* by the case selection strategy $S : \mathbb{N} \to A^+ \times \{+, -\}$ iff
$\exists \sigma \in \Sigma \; \forall L \in \mathcal{L} \; \forall i \in informant(L)$

(1) $\forall n \in \mathbb{N} \; CB_n = S(i[n]) \downarrow$, and $S(i[n]) \subseteq (i^+[n] \times \{+\}) \cup (i^-[n] \times \{-\})$,
(2) $CB = \lim_{n\to\infty} CB_n$ exists,
(3) $L_{st}(CB, \sigma) = L$.

By our underlying assumption the learner is not allowed to change the measure of similarity during the learning process. Therefore, its learning capability depends on the case selection strategy, only.

Let us first informally describe possible dimensions that characterize our case selection strategies.

Access to case history: Is the case selection strategy allowed to store any case that is already presented or has the strategy access to the last one, only?
Deleting cases from the case base: Is the case selection strategy allowed to delete cases from the case base or does the case base grow monotonically?

With respect to these dimensions we can define types of case selection strategies. Let CB_k be the case base constructed when a learner has seen an initial sequence of length k.

Definition 11. Let S be a case selection strategy. Then S is said to be of type[3] **MO-LC**, **MO-RA**, **DE-LC**, and **DE-RA**, respectively, iff the corresponding condition holds for all $k \in \mathbb{N}$ ($CB_0 := \emptyset$).

MO-LC $\quad CB_{k-1} \subseteq CB_k \subseteq CB_{k-1} \cup \{(s_k, b_k)\}$
MO-RA $\quad CB_{k-1} \subseteq CB_k \subseteq \{(s_1, b_1), \dots, (s_k, b_k)\}$
DE-LC $\quad CB_k \subseteq CB_{k-1} \cup \{(s_k, b_k)\}$
DE-RA $\quad CB_k \subseteq \{(s_1, b_1), \dots, (s_k, b_k)\}$

We use these abbreviations as prefixes to **CBL.TXT** and **CBL.INF**. For example, $\mathcal{L} \in$ **DE-RA-CBL.TXT** means that there is a case selection strategy $S \in$ **DE-RA** such that $\mathcal{L}$ can be learned by S in the sense of Definition 9.

Strategies of type **MO-RA** and **DE-RA**, respectively, may store multiple cases in a single learning step. If we demand that strategies of both types store at most a single case in every learning step their learning capabilities will not change.

Because many existing systems simply collect all presented cases, we model this approach, too. A case selection strategy S is said to be of type[4] **CA**, if $CB_k = \{(s_j, b_j) \mid j \leq k\}$ for all $k \in \mathbb{N}$.

It is possible that a **CA-CBL.TXT**-strategy leads to a case base of infinite size, for instance, if the language that is described by a text is infinite. So we have to define what it means that such a strategy learns successfully.

Definition 12. Let $\mathcal{L}$ be an indexed family. We say $\mathcal{L} \in$ **CA-CBL.TXT** iff $\exists \sigma \in \Sigma\ \forall L \in \mathcal{L}\ \forall t \in \mathit{text}(L)$

$$(1)\ \forall n \in \mathbb{N}\ CB_n = t^+[n] \times \{+\},$$
$$(2)\ \exists j \in \mathbb{N}\ L_{st}(CB_k, \sigma) = L \text{ for all } k > j.$$

CA-CBL.INF is defined analogously.

We say $\mathcal{L} \in$ **CA-CBL.TXT** if for all texts of L, $(L_{st}(CB_n, \sigma))_{n \in \mathbb{N}}$ converges *semantically*. This is somehow comparable to the notion of convergence underlying the identification type **BC** in Inductive Inference of recursive functions [3]. All other case-based learning types demand that the sequence $(CB_n)_{n \in \mathbb{N}}$ itself has to converge.

4.1 Learning from Text

In this section we study case-based language learning from positive cases. The first theorem shows that representability and learnability are incomparable. # denotes set incomparability.

Theorem 13. LIM.TXT # REPR$^+$

[3] *MO* stands for "monotonically", *DE* for "delete", *RA* for "random access" and *LC* for "last case"

[4] *CA* stands for "collect all"

To prove that there are representable classes that cannot be learned from text look at the indexed family $\mathcal{L}$ with $L_1 := \{a\}^+$ and for $j > 1$, $L_j := \{a^k \mid 1 \leq k \leq j\}$. $\mathcal{L}$ is representable but is not learnable from text. On the other hand the family of all pattern languages is learnable but not representable (cf. [8]).

Theorem 14. LIM.TXT $\cap$ REPR$^+$ = DE-RA-CBL.TXT

Proof. For "$\subseteq$" let $\mathcal{L} = L_1, L_2, \ldots$ be any indexed family with $\mathcal{L} \in$ **LIM.TXT** $\cap$ **REPR**$^+$. In order to show $\mathcal{L} \in$ **DE-RA-CBL.TXT**, we try to simulate a learning strategy M which **LIM.TXT**-identifies $\mathcal{L}$. To do so, we have to interleave two limiting processes.

Let $\sigma \in \Sigma$ such that $\mathcal{L} \in$ **REPR**$^+(\sigma)$. Moreover, assume any effective enumeration $(F_k)_{k \in \mathbb{N}}$ of all finite subsets of A^+. Since $\sigma \in \Sigma$ and membership is uniformly decidable in $\mathcal{L}$, we may conclude:

Claim: There exists a total recursive function $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ such that for all $j \in \mathbb{N}$:

(1) $\forall x \; f(j,x) \downarrow$,
(2) $\lim_{x \rightarrow \infty} = a$ exists,
(3) $F_a \subseteq L_j$ and $L_{st}(F_a, \emptyset, \sigma) = L_j$.

Thus, on input j f may be used to compute in the limit a finite case base for L_j.

On the other hand, assume any $M \in$ **P** which **LIM.TXT**-identifies $\mathcal{L}$. Now let $L \in \mathcal{L}$, $t = (s_1, +), (s_2, +), \ldots$ be any text for L, and $x \in \mathbb{N}$. The desired case-based learner S will be defined as follows:

- Compute $j_x = M(t[x])$.
- Compute $z_x = f(j_x, x)$.
- If $F_{z_x} \subseteq t^+[x]$, then set $CB_x^+ = F_{z_x}$. Otherwise, set $CB_x^+ = \{s_1\}$.

Finally, taking into consideration that M infers L on text t it follows by the claim above that S converges to a correct case base for L. By definition S is indeed a case selection strategy of type **DE-RA**. □

Theorem 15.
(1) **MO-LC-CBL.TXT $\subsetneq$ DE-LC-CBL.TXT**
(2) **DE-LC-CBL.TXT $\subsetneq$ DE-RA-CBL.TXT**
(3) **DE-RA-CBL.TXT $\subsetneq$ LIM.TXT**

Proof. We only prove Assertion (1). The remaining part can be handled in a similar manner.

By definition **MO-LC-CBL.TXT $\subseteq$ DE-LC-CBL.TXT**. Let $L_1 = \{a\}^+$, $L_2 = \{a, b\}$ and for all $j > 2$, $L_j = \{a^{j-1}\}$. We show that $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$ witnesses the desired separation.

Claim 1: $\mathcal{L} \notin$ **MO-LC-CBL.TXT**

Suppose the converse, i.e., $\mathcal{L} \in$ **MO-LC-CBL.TXT**. Let σ denote the underlying similarity measure. Obviously, for every $k > 1$, the string a^k has to serve as a representative case for the singleton language $\{a^k\} \in \mathcal{L}$. Furthermore, $L_1 = \{a\}^+$ has to be representable w.r.t. σ, too. Consequently, the string a has to be the only representative case for L_1 w.r.t. σ. Now, suppose S is a **MO-LC-CBL.TXT**–strategy for $\mathcal{L}$. Assume that, initially, $(a,+)$ is presented in a text for L_1 and L_2, respectively. Now, it is easy to verify that S, when putting $(a,+)$ into the case base, definitely fails to learn L_2 because S is not allowed to delete $(a,+)$ subsequently. On the other hand, if $(a,+)$ will not be included in the case base, S fails to learn L_1 on its text $t = (a,+),(a^2,+),\dots$ because $\sigma(a^k,a) = 0$ for all $k > 2$. Thus, S is fooled, a contradiction.

Claim 2: $\mathcal{L} \in$ **DE-LC-CBL.TXT**.

Recall that a **DE-LC-CBL.TXT**–strategy S is allowed to delete cases from its actual case base. Obviously, S can be easily defined provided that the underlying similarity measure σ fulfills the following requirements:

- a is representative for L_1 w.r.t. σ,
- b is representative for L_2 w.r.t. σ,
- for all $k > 1$, a^k is representative for L_{k+1} w.r.t. σ.

We omit the details. □

Theorem 16.
(1) **CA-CBL.TXT** $\subsetneq$ **MO-LC-CBL.TXT**
(2) **MO-LC-CBL.TXT** $\subsetneq$ **MO-RA-CBL.TXT**
(3) **MO-RA-CBL.TXT** $\subsetneq$ **DE-RA-CBL.TXT**

Proof. Again we present only the proof of the first Assertion. First, we show **CA-CBL.TXT** $\subseteq$ **MO-LC-CBL.TXT**. As we will see, this is the most interesting part of the proof. Let $\mathcal{L}$ be an indexed family of languages over the alphabet A that is learnable by a **CA-CBL.TXT**–strategy using σ. In order to show $\mathcal{L} \in$ **MO-LC-CBL.TXT** we define a different similarity measure $\tilde{\sigma}$. This will be done in two steps.

Without loss of generality we assume $\sigma \in \Sigma_{\{0,1\}}$. Moreover, assume for all $w \in A^+$, $\sigma(w,w) = 1$. Furthermore, let $w_1, w_2, \dots$ be an effective enumeration of all strings in A^+.

- Step 1: For all $j,k \in \mathbb{N}$, we set $\hat{\sigma}(w_j,w_k) = \sigma(w_j,w_k)$, if $j \le k$. Otherwise, set $\hat{\sigma}(w_j,w_k) = 0$. It is easy to verify that $\mathcal{L}$ is learnable by a **CA-CBL.TXT**–strategy S using $\hat{\sigma}$.
- Step 2: The definition of $\tilde{\sigma}$ is based on $\hat{\sigma}$. Let $j,k \in \mathbb{N}$. Then we set $\tilde{\sigma}(w_j,w_k) = 1$, if there are indices $j_1 < j_2 < \dots < j_n$ such that $j_1 = j$, $j_n = k$ as well as $\hat{\sigma}(w_{j_m}, w_{j_{m+1}}) = 1$ for all $m \le n-1$. Otherwise, set $\tilde{\sigma}(w_j,w_k) = \hat{\sigma}(w_j,w_k)$.
  Now, we may conclude:
  - Observation A: Let $j \in \mathbb{N}$, and $B \subseteq L_j$. Then $L_{st}(B,\emptyset,\hat{\sigma}) \subseteq L_{st}(B,\emptyset,\tilde{\sigma}) \subseteq L_j$.
    To see this, assume any $v \in L_{st}(B,\emptyset,\tilde{\sigma})$ which does not belong to L_j. By definition there has to be a $w \in B$ such that $\tilde{\sigma}(w,v) = 1$. Hence,

there are strings $w = s_1, ..., s_n = v$ such that $1 = \hat{\sigma}(s_1, s_2) = \cdots = \hat{\sigma}(s_{n-1}, v)$. Since $\mathcal{L} \in$ **CA-CBL.TXT** by S w.r.t. $\hat{\sigma}$, $w \in L_j$ together with $\hat{\sigma}(w, s_2) = 1$ directly implies $s_2 \in L_j$. By iterating this argument we obtain $s_n = v \in L_j$. This contradicts our assumption that $v \notin L_j$.

- Observation B: Let $j \in \mathbb{N}$, $B \subseteq L_j$, and $w \in A^+$. Then, $w \in L_{st}(B, \emptyset, \tilde{\sigma})$ implies $L_{st}(B, \emptyset, \tilde{\sigma}) = L_{st}(B \cup \{w\}, \emptyset, \tilde{\sigma})$.
  This observation can be proved in a similar manner.

Now, we are ready to define a strategy T witnessing $\mathcal{L} \in$ **MO-LC-CBL** w.r.t. $\tilde{\sigma}$. Let $L \in \mathcal{L}$ and let $t = (s_1, +), (s_2, +), \ldots$ be any text for L. Initially, set $CB_1 = (s_1, +)$. Let $k \in \mathbb{N}$.

- If $s_{k+1} \notin L_{st}(CB_k, \tilde{\sigma})$, then set $CB_{k+1} = CB_k \cup \{(s_{k+1}, +)\}$
- Otherwise, set $CB_{k+1} = CB_k$

Since S learns L from t by simply collecting all cases, there is an $x \in \mathbb{N}$ such that $L_{st}(t^+[x], \emptyset, \hat{\sigma}) = L_j$. By Observations A and B it follows $L_j = L_{st}(t^+[x], \emptyset, \hat{\sigma}) \subseteq L_{st}(t^+[x], \emptyset, \tilde{\sigma}) = L_{st}(CB_x, \tilde{\sigma}) \subseteq L_j$. Consequently, $L_{st}(CB_x, \tilde{\sigma}) = L_j$. By the definition of T we have $CB_x = CB_{x+r}$ for all $r \in \mathbb{N}$. Thus, T works as required.

Finally, we show **MO-LC-CBL.TXT** $\setminus$ **CA-CBL.TXT** $\neq \emptyset$.

Let $L := \{a^{3k} \mid k \in \mathbb{N}\}$, and for $k \in \mathbb{N}$, $L_k := \{a^{3k}, a^{3k+1}\}$. Let $\mathcal{L}$ be an enumeration of L and all L_k. As we will see, $\mathcal{L}$ witnesses the separation.

Claim 1: $\mathcal{L} \notin$ **CA-CBL.TXT**

Suppose there is a $k \in \mathbb{N}$ such that $\sigma(a^{3k}, w) > 0$ for infinitely many $w \in A^+$. Then L_k cannot be learned by a **CA-CBL.TXT**-strategy for the following reason. If the text $t = (a^{3k}, +), (a^{3k+1}, +), (a^{3k+1}, +), \ldots$ is presented, then $card(L_{st}(t^+[j], \emptyset, \sigma)) = \infty$ for all $j \in \mathbb{N}$. But the language described by the text t is finite.

Suppose there is no $k \in \mathbb{N}$ such that $\sigma(a^{3k}, w) > 0$ for infinitely many $w \in A^+$. Then L is not representable by finitely many cases. Therefore, $\mathcal{L} \notin$ **CA-CBL.TXT** is proved.

Claim 2: $\mathcal{L} \in$ **MO-LC-CBL.TXT**

We need a similarity measure that fulfills the following requirements. Each a^{3k} is a representative for L and each a^{3k+1} is a representative for L_k. Such a measure exists, because the set of representatives for the languages are pairwise disjoint. The corresponding **MO-LC-CBL.TXT**-strategy waits for a a^{3k+1} and the second a^{3k}, respectively, and stores it in the case base. □

Theorem 17. MO-RA-CBL.TXT # DE-LC-CBL.TXT

Proof. **MO-RA-CBL.TXT** $\setminus$ **DE-LC-CBL.TXT** $\neq \emptyset$: Let $L_1 := \{a\}^+$ and $L_j := \{a, a^j\}$ for all $j \geq 2$. Let $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$. Using similar ideas as in the proof of Theorem 15 one can show that $\mathcal{L} \in$ **MO-RA-CBL.TXT**$\setminus$**DE-LC-CBL.TXT**.

DE-LC-CBL.TXT $\setminus$ **MO-RA-CBL.TXT** $\neq \emptyset$: Let $\mathcal{L}$ be an enumeration of $L_1 = \{a\}^+$, $L_k = \left(L_1 \cup \{b^k\}\right) \setminus \{a^k\}$ for $k > 1$, $\widehat{L}_k = \{a^k\}$ for $k > 1$. Then $\mathcal{L} \in$ **DE-LC-CBL.TXT** $\setminus$ **MO-RA-CBL.TXT**. □

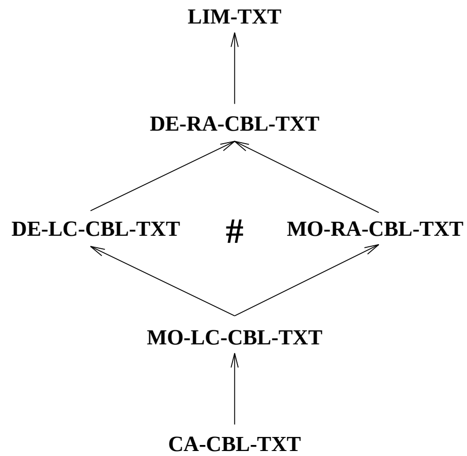

Fig. 1. Relationship between the learning types

These theorems show that both random access to the already presented cases and the ability to delete cases from the actual case base increase the learning power of a case-based learning system. But neither subsumes the other.

Figure 1 illustrates the relationships between all types of case-based learning from text. A path from type T_1 to type T_2 indicates that T_1 is a proper subset of T_2.

4.2 Learning from Informant

Learning from informant is more powerful than learning from text. It is known that every indexed family is learnable from informant. The main result of this section is that every indexed family is case-based learnable with an appropriate fixed measure.

Theorem 18. **IF = LIM.INF = DE-RA-CBL.INF**

Proof. **IF = LIM.INF** follows from [6]. **LIM.INF = DE-RA-CBL.INF** can be shown using the same idea as in the proof of Theorem 14. □

If we consider case-based learning from informant the relations between the case-based learning types change. If the case selection strategy is either allowed to store any case from the informant or to drop cases from the case base, every indexed family is learnable.

Theorem 19.
(1) **MO-RA-CBL.INF = IF**
(2) **DE-LC-CBL.INF = IF**

Proof. It suffices to show **IF** $\subseteq$ **MO-RA-CBL.INF** as well as **IF** $\subseteq$ **DE-LC-CBL.INF**. The main idea is to define the target case selection strategies w.r.t. the similarity measure σ introduced in the proof of Theorem 8. We omit the details. □

From the last theorem we can easy conclude that the learning capability will not increase if we combine both free access to the case history and the ability to delete cases from the case base.

Corollary 20. DE-RA-CBL.INF = IF

Theorem 21. MO-LC-CBL.INF $\subsetneq$ IF

Proof. Let $A = \{a\}$ be an alphabet. $\mathcal{L}$ is an indexed family that contains A^+, all finite languages and $A^+ \setminus \{a^k\}$ for all $k \in \mathbb{N}$. Then $\mathcal{L} \notin$ **MO-LC-CBL.INF**.

Suppose to the contrary that there is a strategy S and a similarity measure σ such that $\mathcal{L} \in$ **MO-LC-CBL.INF** by S w.r.t. σ. Because S learns A^+ there has to be an informant i for A^+ and an $x \in \mathbb{N}$ such that for all $w \in A^+$, $S(i[x]) = S(i[x], (w, +)) = CB_x \subseteq i^+[x] \times \{+\}$ and $L_{st}(CB_x, \sigma) = A^+$. Let $L = i^+[x]$.

Now, taking into account that S has to learn L, in particular, when fed any informant $\hat{\imath}$ that has $i[x]$ as prefix, we can conclude: There are distinct $w, v \in \overline{L}$ such that for all $u \in CB_x^+$, $\sigma(u, v) \leq \sigma(w, v)$. Otherwise, S would fail to learn L when fed $\hat{\imath}$ because for all $y > x$, $CB_x \subseteq S(\hat{\imath}[y])$ and, therefore, $v \in L_{st}(S(\hat{\imath}[y]), \sigma)$, but $v \in \overline{L}$

Now, choose $w, v \in \overline{L}$ such that for all $u \in CB^+$, $\sigma(u, w) \leq \sigma(v, w)$. Finally, consider S when fed the initial segment $i[x], (w, +), (v, -)$. We distinguish two cases.

- Let $S(i[x], (w, +), (v, -)) = CB_x$. Then S fails to learn $\hat{L} = \{a\}^+ \setminus \{v\}$ on each of its informants that have the prefix $i[x], (w, +), (v, -)$ and contain the case $(v, -)$ exactly once.
- Now, let $S(i[x], (w, +), (v, -)) = CB_x \cup \{(v, -)\}$. Obviously, this implies that S fails to learn $\hat{L} = L \cup \{w\}$ on every extension of $i[x], (w, +), (v, -)$ which yields an informant for $\hat{L}$ containing the case $(w, +)$ exactly once.

Hence, in each of the above cases S can be easily fooled, a contradiction. □

Theorem 22. CA-CBL.INF $\setminus$ MO-LC-CBL.INF $\neq \emptyset$

Proof. Let $\mathcal{L}$ be the family of all finite and co-finite languages. We show that there exists a $\sigma \in \Sigma$ such that $\mathcal{L} \in$ **CA-CBL.INF** w.r.t. σ. Let $m, n \in \mathbb{N}$.

$$\sigma(a^m, a^n) = \begin{cases} 1 & \text{if } m = n \\ 1 - \frac{1}{m} & m < n \\ 0 & \text{otherwise} \end{cases}$$

By definition for all $m \in \mathbb{N}$, it holds $L_{st}(\{a^m\}, \emptyset, \sigma) = \{a^n \mid n \geq m\}$ as well as $L_{st}(\emptyset, \{a^m\}, \sigma) = \emptyset$. Furthermore, assume any sets $B, C \subseteq \{a\}^+$ such that $B \cap C = \emptyset$ and $B \cup C = \{a^n \mid n < m\}$. Then, $L_{st}(\{a^m\} \cup B, C, \sigma) = B \cup \{a^n \mid n \geq m\}$ as well as $L_{st}(B, \{a^m\} \cup C, \sigma) = B$. Finally, taking both properties of σ into consideration it is easy to verify that the family of all finite and co-finite languages is learnable by a case-selection strategy which simply collects all cases.

It remains to prove that $\mathcal{L}$ is not in **MO-LC-CBL.INF**. This follows directly from Theorem 21 where we have already shown that a proper subfamily of $\mathcal{L}$ does not belong to **MO-LC-CBL.INF**. This completes the proof. □

While the learning power of **CA-CBL.TXT** is very limited **CA-CBL.INF** contains remarkably rich indexed families like that used in the proof above.

5 Conclusion

Within the present paper we studied different types of case-based learning of indexed families from positive data and both positive and negative data. Following the approach in [8], we considered case-based learning with respect to an arbitrary fixed similarity measure. Thereby, we focused our attention on the problem of how the underlying case selection strategies influence the capabilities of case-based learners. In order to answer this question a couple of new results concerning case-based representability of indexed families have been achieved. As it turns out, the choice of the case selection strategy is of particular importance, if case-based learning from text is investigated. If both positive and negative data are provided, even quite simple case selection strategies are sufficient in order to exhaust the full power of case-based learning.

From our point of view, further investigations concerning case-based learning of indexed families should be oriented in the following way. On the one hand, it seems to be rather natural to give up the assumption that a case-based learner is allowed to use the whole history of the learning task in order to determine its next hypothesis. This may lead to the notion *iteratively working* case-based learning strategies (cf. [7]). On the other hand, when formalizing case-based learning one has to take into consideration that in existing systems a case-based learner has the freedom to change the underlying similarity measure during the learning task, too.

References

1. David W. Aha. Case-Based Learning Algorithms. In Ray Bareiss, editor, *Proc. CBR91*, pages 147 – 158. Morgan Kaufmann Publishers, 1991.
2. D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
3. D. Angluin and C. H. Smith. Inductive Inference: Theory and Methods. *Computing Surveys*, 15(3):237–269, 1983.
4. H. Billhardt. On case-based representability of classes of formal languages. Diploma thesis, HTWK Leipzig, June 1994.

5. Christoph Globig and Stefan Wess. Symbolic Learning and Nearest-Neighbor Classification. In P. Bock, W. Lenski, and M. M. Richter, editors, *Information Systems and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 17–27. Springer Verlag, 1994.
6. E. Mark Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
7. Klaus P. Jantke. Case-Based Learning in Inductive Inference. In *Proceedings of the 5th ACM Workshop on Computational Learning Theory (COLT'92)*, pages 218–223. ACM-Press, 1992.
8. K.P. Jantke and Steffen Lange. Case-based representation and learning of pattern languages. In *Proceedings of the 4th International Workshop on Algorithmic learning Theory (ALT'93)*, volume 744 of *LNAI*, pages 87–100. Springer-Verlag, 1993.