PAC Analyses of a 'Similarity Learning' IBL Algorithm

A.D. Griffiths and D.G. Bridge

Department of Computer Science, University of York, York YO1 5DD, UK
Email: {tony|dgb}@minster.york.ac.uk

Abstract. *VS-CBR* [14] is a simple instance-based learning algorithm that adjusts a weighted similarity measure as well as collecting cases. This paper presents a 'PAC' analysis of *VS-CBR*, motivated by the PAC learning framework, which demonstrates two main ideas relevant to the study of instance-based learners. Firstly, the hypothesis spaces of a learner on different target concepts can be compared to predict the difficulty of the target concepts for the learner. Secondly, it is helpful to consider the 'constituent parts' of an instance-based learner: to explore separately how many examples are needed to infer a good similarity measure and how many examples are needed for the case base. Applying these approaches, we show that *VS-CBR* learns quickly if most of the variables in the representation are irrelevant to the target concept and more slowly if there are more relevant variables. The paper relates this overall behaviour to the behaviour of the constituent parts of *VS-CBR*.

1 Introduction

Instance-based learning (IBL) algorithms may learn by accumulating exemplars in a case base. However, empirical studies [1] [15] show that an instance-based learner that also tunes its similarity measure is generally more efficient than one that does not. More formal studies [8] [6] [7] indicate similar conclusions. For example, in [7] we show that an instance-based learner, which learns monomial target concepts (defined in Section 2) using a fixed but 'optimal' similarity measure, is less efficient than a learner with a simple rule for adjusting a weighted similarity measure. The ability to alter the similarity measure is therefore an important part of IBL.

This paper describes the formal analysis of a similarity-learning IBL algorithm within the PAC learning framework. Few publications describe analyses of this kind, and the most similar work to our own is that of Satoh and Okamoto [11]. These authors study the problem of similarity learning by defining a learning problem where the learner must infer a similarity measure from 'qualitative distance information'. In contrast, this paper analyses the problem of choosing, from positive and negative instances of the target concept, a case base and a similarity measure that together approximate the target concept. In other words, Satoh and Okamoto study instance-based learning only indirectly, via a model that abstracts the problem of learning weights for the similarity measure, while we make an actual instance-based learner the object of our study. In particular, our paper demonstrates two ideas that we suggest might be generally useful for the analysis of specific instance-based learners.

2 Definitions

For the purposes of this paper we assume that examples are represented by a number of binary (0 or 1) valued features. The *example space*, or space of possible problem descriptions, is therefore the space of N-bit binary vectors, referred to as D_N and defined $D_N \hat{=} \{0,1\}^N$.

The paper is concerned with a binary classification task, or the task of learning $\{0,1\}$-valued functions, or *concepts*, defined on D_N. The set of all such concepts is called B_N; $B_N \hat{=} D_N \to \{0,1\}$. In particular, this paper will study the behaviour of instance-based learning algorithms on *monomial* or conjunctive target concepts. A monomial concept can be represented in the propositional calculus by a simple conjunction of literals. The set of monomial concepts is referred to as M_N. Furthermore, $M_{N,k}$ is defined as the set of monomials with exactly k literals; u_1 represents a concept in $M_{N,1}$, while $u_1\overline{u}_2u_3\overline{u}_4$ represents a concept in $M_{N,4}$. The i-th bit of the representation is said to be *relevant* to a monomial concept $t \in M_N$ if the literal u_i or $\overline{u}_i$ appears in the expression representing t, and *irrelevant* if not.

Simple IBL algorithms learn by adding cases to a case base CB and by adjusting a similarity measure σ. A case base CB is a set of *exemplars*, each of which is a pair $(d,n) \in (D_N \times \{0,1\})$. Normally, a case base is *compatible* with some target concept $t \in B_N$ such that for each exemplar $(d,n) \in D_N$, $t(d) = n$. This is written $CB \subseteq t$:

$$CB \subseteq t \hat{=} (\forall (d,n) \in CB \cdot t(d) = n)$$

The similarity measure σ is a total function in $D_N \times D_N \to [0,1]$ which returns a real value indicating the degree of similarity between its two arguments. The pair $\langle CB, \sigma \rangle$ is interpreted as the *representation* of a $\{0,1\}$-valued function defined on D_N as follows:

$$h_{\langle CB,\sigma\rangle}(d) = \begin{cases} 1 \; \textit{if} \;\; \exists (d_{pos},1) \in CB \cdot \forall (d_{neg},0) \in CB \cdot \sigma(d,d_{pos}) > \sigma(d,d_{neg}) \\ 0 \; \textit{otherwise} \end{cases} \tag{1}$$

In other words, a point $d \in D_N$ is positively classified by $h_{\langle CB,\sigma\rangle}$ if and only if there is a stored positive exemplar d_{pos} that is strictly more similar to d according to the similarity measure σ than any of the stored negative exemplars d_{neg}.

Like many other IBL algorithms (e.g. [12] [5] [1]), the learner studied here uses a *weighted* similarity measure; here, this measure is simply a sum of the weights of the bits of the representation on which two descriptions agree:

$$\sigma_{\overline{w}}(d_1,d_2) = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \times (1 - |(d_1)_i - (d_2)_i|) \tag{2}$$

If the weight vector $\overline{w}$ has weight 1 in all elements then $\sigma_{\overline{w}}$ treats all dimensions of the representation equally and is analogous to the Hamming distance between the two descriptions. This special case will be written σ_H.

Finally, if $\overline{x} \in (D_N)^m$ is a sequence, or *sample*, of m descriptions from D_N, then let $\overline{x}_t$ stand for the sequence $\overline{x}_t = \langle (x_i, t(x_i)) \rangle_{i=1}^m$ where x_i is the i-th element of $\overline{x}$. In other words, for each element x_i from $\overline{x}$, $\overline{x}_t$ contains both the element x_i and also $t(x_i)$, the value of the concept t on that example. $\overline{x}_t$ is called a *training sample* for t since it provides a partial definition of t through the labels $t(x_i)$. The *hypothesis* of a learner L on a training sample $\overline{x}_t$, written $L(\overline{x}_t)$, is the concept chosen from B_N by L to approximate the target concept t while the *hypothesis space* of L with respect to t, written H_t^L, is the set of hypotheses that might be output by L on some training sample for t:

$$H_t^L \triangleq \{h \in B_N | \exists \overline{x} \in (D_N)^* \cdot L(\overline{x}_t) = h\} \tag{3}$$

The hypothesis space of a learner L with respect to a *set* of concepts or *concept space* $C \subseteq B_N$ is similarly written H_C^L and defined $H_C^L = \bigcup_{t \in C} H_t^L$.

3 Instance-Based Learning Algorithm *VS-CBR*

```
forall 1 ≤ i ≤ N, n ∈ {0,1} set f[i,n] = 1
set CB = ∅
for i = 1 to m do
    if n_i = 1 then
        if ¬∃d ∈ D_N · (d,1) ∈ CB then set CB = CB ∪ {(d_i,1)}
        for j = 1 to N do
            set f[j, 1 − (d_i)_j] = 0
    else
        set CB = CB ∪ {(d_i,0)}
forall 1 ≤ i ≤ N
    if f[i,0] = 1 ∨ f[i,1] = 1 then
        set w_i = 1
    else
        set w_i = 0
RETURN VS-CBR(s̄) = h_⟨CB,σ_w̄⟩
```

Fig. 1. ***VS-CBR*** **Learning Algorithm for Concepts in** M_N **[14, Fig 4].** $\overline{s} = \langle (d_i, n_i) \rangle_{i=1}^m$ is a training sample from $(D_N \times \{0,1\})^m$.

This paper studies *VS-CBR*, an IBL algorithm that has a simple rule for choosing weights for the weighted similarity measure $\sigma_{\overline{w}}$ (Figure 1). This algorithm learns only monomial target concepts and operates in the following fashion:

- Only the first positive example in the training sample is added to the case base. All other positive examples are discarded.
- All negative examples in the training sample are added to the case base.
- Only binary weights (0 or 1) are assigned to $\sigma_{\overline{w}}$.

- All weights are 1 initially. A weight changes to zero iff two *positive* examples are observed that disagree on that bit of the representation.

Wess and Globig [14] explain the workings of *VS-CBR* with reference to Mitchell's Version Space algorithm [10]. In contrast, we find a closer analogy between the method used by *VS-CBR* to calculate the weights for the similarity measure and the (non case-based) 'standard learning algorithm for monomials' [13] [3]. We call the standard algorithm M for convenience. On the first positive example in the training sample, M sets its hypothesis to the monomial expression representing the concept whose only positive instance is the positive example. On subsequent positive examples, M deduces that if bit j of the positive example is 1 then $\overline{u}_j$ cannot appear in the hypothesis and if bit j is 0 then u_j cannot appear. This rule correctly identifies whether a bit of the representation is relevant, as long as the target concept is a monomial. Literals judged irrelevant can then be deleted by the learner.

VS-CBR operates in exactly the same way, using the array f to calculate which bits of the representation are irrelevant. After processing any sample $\overline{s}$, $f[i,n] = 1$ only if no positive exemple d_{pos} has been processed such that $(d_{pos})_i = (1-n)$ and therefore all observed positive examples have value n on bit i of the representation. *VS-CBR* must then convert f to the weight vector $\overline{w}$ to be used in the similarity measure. This is straightforward; a bit of the representation is irrelevant to the definition of the concept whenever both possible values have been seen in positive exemplars and hence $f[i,0] = 0 \wedge f[i,1] = 0$. In this case, the corresponding weight w_i is set to 0 so that bit i is ignored by $\sigma_{\overline{w}}$; otherwise, it is set to 1.

4 Direct Analysis of VS-CBR

The PAC Learning Framework [3] provides a means of evaluating learning algorithms and in particular defines a quantity called the *sample complexity* which serves as the measure of efficiency of a learning algorithm:

Definition 1. Sample Complexity [3]. The sample complexity $m_L(t,\delta,\epsilon)$ of a learning algorithm L with respect to a target concept t is the least value of m such that, for any degree of confidence and accuracy $0 < \delta, \epsilon < 1$, the hypothesis inferred by L from a training sample of size m will, with probability greater than $1-\delta$, have an error less than ϵ with respect to the target concept t, using any underlying distribution.

Additionally the sample complexity $m_L(C,\delta,\epsilon)$ of a learner L with respect to a *concept space* C is defined $m_L(C,\delta,\epsilon) = \max_{t \in C} m_L(t,\delta,\epsilon)$ and stands for the minimum size of sample sufficient for *probably* (δ) *approximately* (ϵ) correct learning of *any* target concept in C.

An algorithm with a small sample complexity will therefore require fewer examples to choose a hypothesis that is probably approximately correct than an algorithm with a high sample complexity. Key results in the PAC framework [4] link the sample complexity of a learner to its hypothesis space. For example, an upper bound on sample complexity, in terms of the cardinality of the hypothesis space of the learner, is easily proven:

Theorem 2. **[7, Prop 6.5.11] c.f. [4, Thm 2.2]** *The sample complexity of any consistent learning algorithm L that learns a target concept $t \in B_N$ is bounded above by a quantity of the order of* $(\frac{1}{\epsilon}\log\frac{1}{\delta} + \frac{\log|H_t^L|}{\epsilon})$.

This result can also be expressed in terms of a concept space $C \subseteq B_N$, giving a bound on $m_L(C, \delta, \epsilon)$ in terms of $|H_C^L|$.

Our approach therefore is to explore the hypothesis space of $VS\text{-}CBR$ to make predictions about the relative efficiency of the learner on different target concepts. Firstly, Proposition 3 shows that the discarded positive examples are redundant since they are equivalent, from the point of view of the weighted similarity measure, to the stored, 'prototypical' exemplar [14]. Hence the hypothesis space $H_t^{VS\text{-}CBR}$ is simply the set of concepts with case based representation $\langle CB, \sigma_{\overline{w}_{CB}}\rangle$, where CB is a case base compatible with t and the similarity measure $\sigma_{\overline{w}_{CB}}$ has weights $\overline{w}_{CB} \in \{0,1\}^N$ such that a bit of the representation has value 0 iff that bit can be proven to be irrelevant to the target concept from the positive exemplars in the case base:

$$(w_{CB})_i = \begin{cases} 1 \text{ if } \exists b \in \{0,1\} \cdot \forall (d,1) \in CB \cdot (d)_i = b \\ 0 \text{ otherwise} \end{cases} \quad (4)$$

Proposition 3. **[7, Propn 6.3.2]**[1] *A concept f is a member of the hypothesis space of $VS\text{-}CBR$ with respect to a target concept $t \in M_N$ if and only if there is a case base $CB \subseteq t$ such that $h_{\langle CB, \sigma_{\overline{w}_{CB}}\rangle} = f$, where $\sigma_{\overline{w}_{CB}}$ is defined as in equation (4):*

$$\forall t \in M_N \cdot \forall f \in B_N \cdot f \in H_t^{VS\text{-}CBR} \leftrightarrow \exists CB \subseteq t \cdot h_{\langle CB, \sigma_{\overline{w}_{CB}}\rangle} = f$$

While it is reassuring to know that the positive exemplars discarded by $VS\text{-}CBR$ are redundant, the problem with Proposition 3 is that the hypothesis is represented using different similarity measures at different times. It is the constantly changing relationship between the case base and the similarity measure that makes the analysis of similarity learning IBL algorithms difficult. Fortunately, in the case of $VS\text{-}CBR$, changes in the similarity measure are unidirectional; weights can be changed from one to zero but not vice versa. The similarity measure therefore converges monotonically toward the ideal as more examples are read from the training sample. As a result, it is possible to express the hypothesis space $H_t^{VS\text{-}CBR}$ as a set of concepts representable by a *single* similarity measure:

Proposition 4. **[7, Propn 6.4.2]** *The effective hypothesis space of $VS\text{-}CBR$ w.r.t. any target concept $t \in M_N$ is the set of concepts $h_{\langle CB, \sigma_H\rangle}$ where CB is any*

[1] The proof of Proposition 3 and of all the other new results in this paper is ommitted but can be found in [7].

case base compatible with t and that in addition has no more than one positive exemplar:

$$\forall t \in M_N \cdot H_t^{VS\text{-}CBR} = \{h_{\langle CB,\sigma_H\rangle} | CB \subseteq t \wedge \#\{d_{pos} \in D_N | (d_{pos}, 1) \in CB\} \leq 1\}$$

A concept in $H_t^{VS\text{-}CBR}$ can therefore be represented by σ_H and by a case base containing some subset of the negative instances of the target concept and no more than one positive instance. Since the concepts in $M_{N,1}$ are the most general monomial concepts, with the largest number of positive instances and therefore the smallest number of negative instances, there will be at least as many concepts in $H_{M_{N,N}}^{VS\text{-}CBR}$ as in $H_{M_{N,1}}^{VS\text{-}CBR}$. For example, the concepts in $H_{M_{N,1}}^{VS\text{-}CBR}$ are those concepts with a case-based representation where the negative exemplars are drawn from one half of the example space and the single positive exemplar lies in the opposite half. If, instead, a target concept that also has that positive exemplar as a positive instance is taken from $M_{N,2}$, then there will be an additional quarter of the example space whose descriptions can appear as negative exemplars in the representation of hypothesis concepts. In the limit, if the target concept is from some space $M_{N,N}$, then there is only one description in D_N that is a positive instance of the target concept, and all other descriptions can appear as negative exemplars. In addition, for each monomial concept $t \in M_{N,k}$, there will be more specific monomial target concepts that will have all the negative instances of t as negative instances and will still be positive on some single description. The result below therefore follows immediately as a corollary of Proposition 4:

Corollary 5. **[7, Cor 6.4.3]** *The hypothesis space $H_{M_{N,k}}^{VS\text{-}CBR}$ of $VS\text{-}CBR$ w.r.t. the concept space $M_{N,k}$* is a subset of *the hypothesis space $H_{M_{N,k'}}^{VS\text{-}CBR}$ w.r.t. $M_{N,k'}$, for all $N \geq k' \geq k$.*

$$\forall 1 \leq k \leq k' \leq N \cdot H_{M_{N,k}}^{VS\text{-}CBR} \subseteq H_{M_{N,k'}}^{VS\text{-}CBR}$$

This result shows that the size of the hypothesis space of $VS\text{-}CBR$ on the concept space $M_{N,k}$ increases in the value of k. We would therefore expect $VS\text{-}CBR$ to learn target concepts in $M_{N,1}$ most easily, since this is when it has the smallest hypothesis space, while target concepts in $M_{N,k}$ are learnt more slowly. Simple experiments reported in [7] confirm this. Corollary 5 therefore allows predictions to be made about the relative efficiency of $VS\text{-}CBR$ on different monomial target concepts and also gives some kind of explanation of why target concepts defined by the smallest monomial expressions are learnt more easily than target concepts represented by larger monomial expressions.

5 Constituent Analysis of VS-CBR

In Section 4, we showed that, in this instance, the size of the hypothesis space H_C^L correlates with the efficiency of the learner on target concepts $t \in C$. The link between hypothesis space and efficiency is only a heuristic one, however.

In particular, Theorem 2 is only an upper bound rather than an expression for the sample complexity itself. We therefore attempted to validate the results of the previous section by adopting a different line of analysis, and considering *VS-CBR* as two separate processes, or *constituents*, one of which tunes the similarity measure, while the other is responsible for populating the case-base.

1. We noted in our description of *VS-CBR* that the learner assigns a zero weight to a bit of the representation iff that bit is judged to be irrelevant by the monomial learner M; *VS-CBR* will choose the correct similarity measure on precisely those training samples where M correctly identifies the target concept. The convergence rate of the process by which *VS-CBR* chooses $\overline{w}$ can therefore be determined by considering the sample complexity of M.
2. On the other hand, the convergence rate of the process by which *VS-CBR* populates its case base can be studied through an IBL algorithm that does not have to learn a measure of similarity, but is given, *a priori,* a similarity measure whose weights are 1 iff that bit of the representation is relevant to the target concept. This new algorithm therefore *starts* learning with all the information needed for the similarity measure, and any subsequent learning is due to the exemplars added to the case base.

These two views are developed in the following subsections.

5.1 Learning the Similarity Measure for *VS-CBR*

We noted previously that *VS-CBR* chooses the correct similarity measure on precisely those training samples where M, the 'standard learning algorithm' for monomial target concepts [13], exactly identifies the target concept. M starts with a *specific* hypothesis corresponding to the first positive exemplar read from the training sample, and then generalises the hypothesis by deleting literals until the monomial expression that exactly represents the target concept is reached. For each target concept $t \in M_N$, each of the monomials more specific than t might have been chosen as an 'intermediate' hypothesis by M before t is correctly identified:

Proposition 6. [7, Propn 6.5.13] *The hypothesis space of M, the standard learning algorithm for monomial concepts, w.r.t. a k-literal monomial target concept $t \in M_{N,k}$, contains all concepts in M_N which specialise the target concept t along with the concept f_0 that has value 0 on all descriptions:*

$$H_t^M = \{h \in M_N | h \sqsubseteq t\} \cup \{f_0\}$$

where $h \sqsubseteq t$ is read 'h specialises t' $(\forall d \in D_N \cdot h(d) = 1 \rightarrow t(d) = 1)$ *and f_0 is the concept such that* $\forall d \in D_N \cdot f_0(d) = 0$.

[7, Propn 6.5.13] also argues that $|H_t^M| = 3^{N-k} + 1$, which agrees with what is already known about M. Langley and Iba claim in passing that the number of examples that M needs *in the average case* increases with the number of irrelevant attributes [9]. Proposition 6 and Theorem 2, correspondingly, show that the sample complexity of M (a *worst-case* measure of efficiency) can increase no more than linearly in $N-k$. The more irrelevant bits, the more examples will be needed by M, and therefore by $VS\text{-}CBR$ before it chooses a good similarity measure. This contrasts with the overall picture of $VS\text{-}CBR$ developed in Section 4, where target concepts with the most irrelevant variables appeared to be the easiest to learn. The reason for this difference must lie in the requirements of $VS\text{-}CBR$ for examples to populate the case base.

5.2 Populating the case base for $VS\text{-}CBR$

```
set CB = ∅
for i = 1 to m do
     set CB = CB ∪ {(d_i, n_i)}
RETURN CB2(s̄) = h_⟨CB, σ_w̄t⟩
```

Fig. 2. CB2 Learning Algorithm for Concepts in M_N. $\overline{s} = \langle (d_i, n_i) \rangle_{i=1}^{m}$ is a training sample from $(D_N \times \{0,1\})^m$ and weight vector $\overline{w}_t$ has value 1 iff that bit of the representation is relevant to t.

Figure 2 shows the instance-based learning algorithm $CB2$. $CB2$ collects all the examples from the training sample into the case base CB and outputs the hypothesis represented by $\langle CB, \sigma_{\overline{w}_t} \rangle$, where $\sigma_{\overline{w}_t}$ is the instance of $\sigma_{\overline{w}}$ that is ideally weighted according to whether or not a bit of the representation is relevant to the target concept t ($(w_t)_i = 1$ iff bit i is relevant to t).

Whereas $CB2$ uses the ideal measure $\sigma_{\overline{w}_t}$ (as if this was known in advance), $VS\text{-}CBR$ infers a similarity measure $\sigma_{\overline{w}_{CB}}$ from the available exemplars that *approximates* $\sigma_{\overline{w}_t}$ and will eventually converge to the 'ideal' weighting if sufficient positive exemplars are available. In this way, $CB2$ can be seen as a limiting approximation of $VS\text{-}CBR$ that illustrates the maximum contribution that can be made by the best possible choice of similarity measure from the class defined by binary weight vectors $\overline{w} \in \{0,1\}^N$.

The fact that $CB2$ *starts off* with the correctly weighted similarity measure $\sigma_{\overline{w}_t}$ means that it has only to populate the 2^k classes of descriptions which are treated as equivalent by $\sigma_{\overline{w}_t}$ [14]. The sample complexity of $CB2$ therefore can be established to be a function of k and not of N. For example, by using a straightforward simplification of the analysis of Aha *et al* [2], the following upper bound can be established:

Proposition 7. **[7, Cor 6.5.9]** *The sample complexity of $CB2$ with respect to a target concept represented by a k-literal monomial expression, $t \in M_{N,k}$, is no more than $\frac{2^k}{\epsilon} \log_e \frac{2^k}{\delta}$, independent of the value of N.*

$$m_{CB2}(t, \delta, \epsilon) \leq \frac{2^k}{\epsilon} \log_e \frac{2^k}{\delta}$$

Having established that the sample complexity of $CB2$ must be independent of N, the following result, similar to Corollary 5, can also be proven which suggests that the sample complexity of $CB2$ increases in k:

Proposition 8. **[7, Propn 6.5.4]** *The effective hypothesis space* $H^{CB2}_{M_{N,k}}$ *of* $CB2$ *w.r.t. the concept space* $M_{N,k}$ *is a subset of the hypothesis space* $H^{CB2}_{M_{N,k'}}$ *w.r.t.* $M_{N,k'}$ *for any* $N \geq k' \geq k$.

$$\forall 1 \leq k \leq k' \leq N \cdot H^{CB2}_{M_{N,k}} \subseteq H^{CB2}_{M_{N,k'}}$$

In contrast to M (Section 5.1), $CB2$ has a sample complexity apparently increasing in the number of *relevant* bits of the representation and independent of the overall size N of the representation. (As with *VS-CBR* in Section 4, these statements can also be demonstrated in the average-case by simple experiment [7].) The *order* of the sample complexity of $CB2$ is not known, unlike that of M which we could state to be $O(N-k)$. To obtain such a result, we would need to characterise and count the number of different hypotheses in H_t^{CB2} in order to apply Theorem 2. We have made partial progress towards this; [7] gives some necessary conditions on the representation of concepts in the propositional calculus in order for them to be members of H_t^{CB2}. Counting the number of such concepts for small values of k [7, Table 6.2] makes it clear that, although $\log |H_t^{CB2}|$ increases more than linearly in k, the upper bound indicated by Theorem 2 would be greatly less than the exponential bound indicated by Proposition 8. We can say no more than this, however, since we have no general expression for $|H_t^{CB2}|$.

6 Conclusions

Section 4 showed that it is possible to re-express the hypothesis space of *VS-CBR*, an IBL algorithm that tunes its similarity measure as part of learning, as a set of concepts representable w.r.t. a single similarity measure. This characterisation is sufficient to show that the hypothesis space of *VS-CBR* is smallest w.r.t. the concept space $M_{N,1}$, and becomes larger for the concept space $M_{N,k}$ as k increases. From this we infer, using standard results in the PAC learning framework, an upper bound on the sample complexity of *VS-CBR* that also increases in k.

Section 5 then described *VS-CBR* as two processes that operate in tandem, each independently manipulating an element of the representation $\langle CB, \sigma_{\overline{w}} \rangle$. This 'constituent' analysis showed that increasing the number of relevant variables (the parameter k) makes it harder for *VS-CBR* to get sufficient exemplars to cover the example space but easier to infer a good similarity measure. However, decreasing k makes the similarity measure harder to infer but reduces the sample complexity of building a suitable case base. Since a target concept $M_{N,1}$ is more easily learnt by *VS-CBR* than a target concept in $M_{N,N}$, it seems that populating the case base is the harder task. In an experiment where the accuracy of *VS-CBR* is measured on target concepts in $M_{N,k}$ for decreasing values of k, the extra cost of inferring a similarity measure that ignores *more* irrelevant features must be compensated for by a greater reduction in the sample complexity of building the case base. This suggests that the sample complexity of $CB2$ increases more than linearly in k and, conversely, agrees with the results of

[11], which also suggest that the problem of finding suitable weights may have a relatively low sample complexity.

Our work aims to develop ways of evaluating and comparing IBL algorithms. We believe that theoretical analyses are a valuable complement to empirical comparisons such as [15] because they can define the limits of the performance of a learner, and characterise when one instance-based learner outperforms another. However, more work is needed to refine the theoretical tools that are available so that directly usable results can be derived for realistic learning algorithms. The work we describe here has explored two different ideas which seem promising for investigating instance-based learners:

1. We have shown that it is useful to define the hypothesis space of a learning algorithm (w.r.t. different target concepts) in order to predict the relative efficiency of the learner on those target concepts.
2. We have also shown that it can be useful to think of IBL algorithms in terms of their constituent parts.

In this paper we have considered only the straightforward world of conjunctive (monomial) target concepts. Further work must explore whether these principles are also useful for more general instance-based learners.

References

1. D. W. Aha. Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36:267–287, 1992.
2. D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.
3. M. Anthony and N. Biggs. *Computational Learning Theory.* Cambridge University Press, 1992.
4. A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, Oct 1989.
5. J. P. Callan, T. E. Fawcett, and E. L. Rissland. CABOT: An adaptive approach to case-based search. In *Proceedings of IJCAI-91*, pp 803–808, Morgan Kaufmann, 1991.
6. C. Globig, K. P. Jantke, S. Lange, and Y. Sakakibara. On case-based learnability of languages. *New Generation Computing*, 15(1), 1997.
7. A. D. Griffiths. *Inductive Generalisation in Case-Based Reasoning Systems.* PhD thesis, Published as Technical Report YCST-97-02, Department of Computer Science, University of York, York YO1 5DD, UK, 1997.
8. K. P. Jantke. Case-based learning in inductive inference. In *Proceedings of COLT92*, pp 218–223. ACM Press, 1992.
9. P. Langley and W. Iba. Average-case analysis of a nearest neighbour algorithm. In R Bajcsy, editor, *Proceedings of IJCAI-93*, pp 889–894. Morgan Kaufmann, 1993.
10. T. M. Mitchell. Generalisation as search. *Artificial Intelligence*, 18(2):203–226, 1982.
11. K. Satoh and S Okamoto. Towards PAC-learning of weights from qualitative distance information. In D W Aha, editor, *Case-based reasoning: Papers from the 1994 AAAI Workshop*, Technical Report WS-94-01. AAAI Press, 1994.
12. C. Stanfill and D. L. Waltz. Towards memory-based reasoning. *Communications of the ACM*, 29(12).1213–1228, 1986.
13. L. G. Valiant. Deductive learning. *Philosophical Transactions of the Royal Philosophical Society of London A*, 312:441–446, 1984.
14. S. Wess and C. Globig. Case-based and symbolic classification - A case study. In S Wess, K-D Althoff, and M M Richter, editors, *Topics in CBR: Selected Papers from EWCBR93*, LNCS 837, pp 77–91. Springer-Verlag, 1994.
15. D. Wettschereck, D. W. Aha, and T. Mohri. A review and comparative evaluation of feature weighting methods for lazy learning algorithms. Technical Report AIC-95-012, Navy Center for Applied Research in AI, Naval Research Laboratory, Washington, DC 20375-5337, USA, 1995.