Comments on Daniel Steel's "Mind Changes and Testability"

Matthew C. Kopec

ABSTRACT

In "Mind Changes and Testability," Steel argues that for all inductive problems that are what he terms '*strongly VC ordered,'* formal and statistical learning theory will yield the same results. I argue that this conclusion is not yet established. Due to a flaw in his definition of *VC ordering*, an ambiguity arises, and one promising way to remove the ambiguity causes problems for his proofs. The most worrisome problem arises for his proposition 3(c) and its corollary, which Steel asserts is needed for Goodman's riddle to be answered. I suggest the modifications required to salvage the conclusion. Though the modifications succeed, they do so at the cost of limiting the applicability of Steel's conclusion.

Introduction

Steel's central thesis is that for all inductive problems that are strongly VC-ordered, even though formal and statistical learning theory utilize different methods for selecting hypotheses, both methods select the same hypothesis given any string of data. This is an interesting result. Here two very different methods of hypothesis selection that trace their intellectual heritages back to quite distinct thinkers (Pierce, James, and Reichenbach for formal learning theory and Popper for statistical) converge. This result suggests future work to search for convergence between the two fields in new areas where hypothesis selection is required. In addition, it offers a way to put to rest some disagreements between formal and statistical learning theorists concerning which theory is superior. If the inductive problem under examination is strongly VC ordered, then the two theories will converge, and practitioners of either can save themselves the hassle of defending their methods against their opponents.

The linchpin of this conclusion is the notion of *VC ordering*, which is to my knowledge a novel addition to the literature, as well as *strong VC ordering* which depends on the former. I believe there to be a flaw with the definition of VC order, which causes difficulties for Steel's argument. The definition as stated is:

> VC Order: Let P be an inductive problem defined by the set of data sequences Ω and partition of hypotheses Π. Then P is *VC ordered by* $\{C_0, ..., C_k, N\}$ if and only if $(C_0, \cup ... \cup C_k \cup N) = \Pi$, and, for every C_i in $\{C_0, ..., C_k\}$, the VC dimension of C_i equals i.

The problem is that the condition that $(C_0 \cup ... \cup C_k \cup N) = \Pi$ cannot be met. Since Π is a partition of Ω, it is a mutually exclusive and jointly exhaustive set of subsets of Ω, i.e. it is a set containing sets of data sequences. Requiring that $(C_0 \cup ... \cup C_k \cup N) = \Pi$ entails that the former set must be a subset of the latter and vice versa. But $(C_0 \cup ... \cup C_k \cup N)$ is a set containing hypotheses as elements (i.e. data sequences), while Π contains sets as elements. So it's not the case that Π will be a subset of $(C_0 \cup ... \cup C_k \cup N)$, and therefore $(C_0 \cup ... \cup C_k \cup N) \neq \Pi$ for any inductive problem. So a modification must be made to the definition if the term '*VC order'* is to apply to some inductive problem and some partition of hypotheses.[1]

Modification and Objection 1

Though the above definition is flawed strictly speaking, I believe there are ways to modify it to capture the intended spirit. It seems to me that what Steel intends to say is that a set of hypothesis groupings can be said to *VC order* an inductive problem only if every hypothesis to be found within at least one of the groupings can also be found somewhere in the partition that frames the problem and vice versa. So the condition would be that $(C_0 \cup ... \cup C_k \cup N) = \cup\Pi$ where '$\cup\Pi$' stands for the union of all of the members within Π. Since Π is a partition of Ω, $\cup\Pi = \Omega$. So the problematic condition could be changed to '$(C_0 \cup ... \cup C_k \cup N) = \Omega$.' This gets around the problem from before that the one set contained hypotheses as elements and the other sets as elements. Though this seems a natural and trivial modification, it makes a difficulty for his proof of proposition 3(c) and its corollary.

As Steel states, "proposition 3(c) entails that an inductive problem cannot have distinct strong VC orderings that make a difference to what the natural projection rule

[1] I owe Peter Vranas credit for pointing out that this equality won't hold.

conjectures"(23), which he calls the "corollary" of the proposition. I will show by way of a counterexample that once the above modification is made, this claim is false. I'll use the *generalized Goodman problem* as the framework for the counterexample, where we are attempting to decide between different hypotheses concerning the colors of the marbles. Notice that we could describe the observations in a different language, and that this change might make different groupings more natural. Let's say we have another group of scientists that speak in terms of objects being rue and bled instead of blue and red. An object is 'rue' for these individuals just in case it is observed up to the 99^{th} observation and is red, or after the 99^{th} and is blue, and an object is 'bled' just in case it is observed up to the 99^{th} observation and is blue or after the 99^{th} observation and is red. These scientists decide to group the hypotheses according to the number of "color" changes that occur. For instance G_0 would be {all are rue}, G_1 would be {n: all are rue_n-bled}, G_2 would be {n, m: all are rue(n)-bled(m)-rue; m > n }, and so on. The way the data are encoded into 1s and 0s can remain the same. For instance, the only hypothesis in G_0 stands for a string of 99 1s followed by an infinite string of 0s, the first member of G1 stands for an infinite string of 1's, while the second member stands for a string of 99 1s, followed by a 0, followed by an infinite string of 1's, and so on. Take the partition of the hypotheses that consist of {G_0, G_1, G_2, N}, where N will stand for the complement of the union the other members of the partition. Since we know that the partition {H_0, H_1, H_2, N} strongly VC orders the inductive problem, the corollary rules out the possibility that {G_0, G_1, G_2, N} strongly VC orders the problem as well, unless this makes no difference to what the natural projection rule conjectures.

It turns out that the problem is in fact strongly VC ordered by {G_0, G_1, G_2, N}, that it will make a difference to what the natural projection rule conjectures, and therefore something has gone wrong. First, the inductive problem *P* is VC ordered by {G_0, G_1, G_2, N}. The set of data sequences is Γ (the set of all infinite sequences of 1s and 0s), ($G_0 \cup G_1 \cup G_2 \cup N) = \Gamma$, and for every G_i in {G_0, G_1, G_2}, the VC dimension of G_i equals *i*. The latter is true because the set {all are rue} only shatters an extension of the data of length 0, since it is not consistent with the next observation being bled (red in our language, i.e., $e = 1$). So the VC dimension of G_0 is 0. The set {n: all are rue_n-bled} shatters an extension of length one, since 'rue$_{99}$-bled' and 'rue$_{100}$-bled' covers both possibilities, but the set does not shatter an extension of length two. It isn't consistent with the 100th observation being bled and the 101st being rue (i.e., e = 10). So the VC dimension of G1 is 1. Similarly, G_2 shatters an extension of length 2, but isn't consistent with the 100th being bled, the 101st being rue and the 102nd being bled (i.e., e = 101). So this qualifies under the modified definition of VC order. Similarly, for every G_i in $\{G_0, G_1, G_2\}$, any extension e of the data in *P,* and any $m \geq 0$, if G_i is the set with the lowest VC dimension consistent with e, then G_{i+m} (if it exists) shatters $e*m$ but no further extension of e. Therefore, the problem *P* is strongly VC ordered by the set $\{G_0, G_1, G_2, N\}$. So the problem is strongly VC ordered by $\{G_0, G_1, G_2, N\}$ and $\{H_0, H_1, H_2, N\}$.

Now notice that the natural projection rule will choose different hypotheses depending on which of these two partitions it is acting upon given the data so far. The natural projection rule instructs us to choose the hypothesis consistent with the data from within the set with lowest VC dimension. Given the data so far (i.e., 99 1s), the natural projection rule acting upon the H partition tells us to choose a hypothesis from H_0, since this is the set with the lowest VC dimension (i.e., 0). Given the data so far, the natural projection rule when acting upon the G partition tells us to choose a hypothesis from G_0 since it is the set with the lowest VC dimension (i.e., 0). The problem is that the two sets, H_0 and G_0, do not contain the same hypothesis. In the former case the natural projection rule tells us to choose 'all are red' while in the latter it tells us to choose 'all are rue.' Since these two partitions are both strongly VC ordered, and the ordering does make a difference to what the natural projection rule conjectures, the corollary to proposition 3(b) seems to be false.

Modification and Objection 2

So what went wrong in the above case, and does this show that the proof of proposition 3(c) or its corollary is flawed? I think that the above problem doesn't show that either proof fails, but rather that the definition of *VC order* needs further modification. The problem arises because what N stands for in $\{H_0, H_1, H_2, N\}$ and $\{G_0, G_1, G_2, N\}$ are actually different. I will call the first N_h and the second N_g. The trouble in the proofs stems from Steel's reliance on the reasoning used in the proof of proposition

3(b) to prove proposition 3(c). Proposition 3(b) is proven by showing that if a method of hypothesis selection is logically reliable then some extension of the data can be made to conjecture at least one hypothesis from each C_i. The proof of 3(c) relies on this reasoning to show that if a method chooses from a set that doesn't have the lowest VC dimension, then it can be made to choose again from the same set, thus increasing the maximum number of mind changes by one. But the move from the former case, where only one partition is under consideration, to the latter is not yet justified.

Since N_h and N_g amount to different catch-all hypotheses, there is at least one member that is in the latter but not in the former or vice versa. It is true that if the method given the H partition chooses a hypothesis from a set with a non-minimal VC dimension then the method can be made to choose again from that H set. But it does not follow that if the method chooses a member who appears in an H set with non-minimal VC dimension relative to the H partition then it can be made to draw again from the G set in which it appears. The reason is that some of the hypotheses found in H_2 are only to be found in N_g and some in G_2 are only found in N_h. The appendix shows a representation of this. Just to make this explicit, let's take the method to be the natural projection rule acting upon the G partition. First, the method conjectures {all are rue}, which is a member of H_1. If the next observation is red, then the method will conjecture {all are rue$_{99}$-bled} which corresponds to the hypothesis in H_0. For the method to be forced to choose a hypothesis that appears in H_1 again, one of the future observations must be blue. But notice that once this happens, the method operating on the G partition will choose a corresponding hypothesis from G_2. And though future observations would be able to make the method choose a hypothesis that appears in H_2 and then N_h if it were acting on the H partition (making the total of mind changes on the H side one more than the maximum for the natural projection rule acting upon the H partition), no future observation can make it budge to anything besides N_g, on the G side. So some data strings that cause the method to choose twice from a set on the H side cannot do so on the G side, since that string will cause the method acting upon the G partition to posit N_g, which will count as the true hypothesis relative to that partition.

Now I believe that a modification can be made to block this problem. The problem arose because $(H_0 \cup H_1 \cup H_2) \neq (G_0 \cup G_1 \cup G_2)$, or equivalently because $N_h \neq$

N_g. The modification would be to specify that if the partition Π contains a set containing a catch-all, then the problem can only be VC ordered by a partition with the same catch-all. Though this bypasses the above objection, I am worried that this modification won't be enough. When Steel gives the definition of 'VC ordered' he frames it in finite terms, where the number of possible subsets C_1 through C_k is finite. He then mentions that "The definition of VC order can also apply in cases where there is no upper bound on k"(21). Then when stating the definition for 'strongly VC ordered' the only portion of the definition that involves the finitude of the subsets is the portion that notes that a strongly VC ordered problem must be VC ordered by $\{C_0, ..., C_k, N\}$. So I take it that the transfinite extension is legitimate in the latter case as well. So I assume that according to Steel it is possible for an inductive problem to be strongly VC ordered by the set $\{C_0, C_1, C_3,....\}$ where the C_i go on indefinitely.

Now consider what I'll call the *more generalized Goodman problem*, where the basic structure is kept the same, but where the number of color changes can be infinite. First, notice that the crucial claims about Steel's *generalized Goodman problem* seem to be true of the *more generalized Goodman problem* as well. If we were to form a partition of the hypotheses $\{H_0, H_1, H_2, ...\}$ all grouped according to the number of color changes (in our observation language), the VC dimensions of these sets will still climb from 0, 1, 2... in an orderly fashion. It seems clear that the *more generalized Goodman problem* is strongly VC ordered by this partition. The reasoning from the previous section could be used to show that the problem is also strongly VC ordered by the partition $\{G_0, G_1, G_2, ...\}$. The possible data sequences are the same, as are the observations thus far. More importantly the condition just added concerning the catch-all is met, since in the *more generalized Goodman problem* there is no catch all. Since we again have two distinct strong VC orderings, and it will similarly make a difference to what the natural projection rule conjectures, the problem arises again.

Given this last objection, I would argue that Steel needs to either assert that an inductive problem cannot be VC ordered by an infinite partition, or at least that it cannot be strongly VC ordered by one.

Conclusion

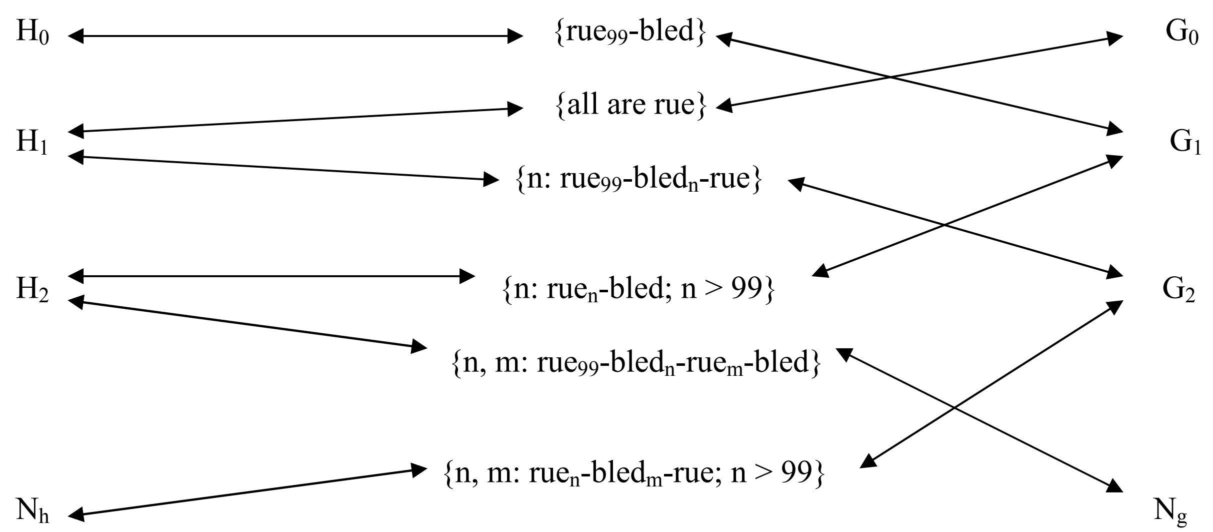
I hope to have shown that the concept of *VC order* must be modified for at least some of Steel's conclusions to follow. First, due to the condition that $(C_0 \cup ... \cup C_k \cup N) = \Pi$, there are no inductive problems that are *VC ordered*. Changing this to $(C_0 \cup ... \cup C_k \cup N) = \Omega$ bypasses that problem, but proves to be too permissive. Since both $(H_0 \cup H_1 \cup H_2 \cup N_h)$ and $(G_0 \cup G_1 \cup G_2 \cup N_g) = \Omega$ in my counterexample, this modification makes it possible to show that the corollary to proposition 3(c) is false. As he states, this corollary is crucial to answering Goodman's riddle. So adding a condition that Π must contain as a member the same catch all as the partition that *VC orders* it can block this counterexample. If an infinite partition of hypotheses can strongly VC order an inductive problem, a similar counterexample to the first can be raised. So that must be ruled out. If these modifications are made, I believe the conclusion follows.

But the modifications come at a cost. Steel seems to suggest that the move to an infinite partition should not make problems for the convergence claim, but it turns out that it does. Limiting ourselves to finite partitions might also strike one as *ad hoc*. For instance, Steel's result will hold if we allow for thirty trillion color changes, but once we allow them to be infinite, the modification tells us all bets are off. So I'm uncertain if the modifications presented here will be seen as acceptable, or rather if another way of modification should be sought after.

References

Steel, D. [2008] "Mind Changes and Testability: How Formal and Statistical Learning Theory Converge in the New Riddle of Induction." *Formal Epistemology Workshop, 2008*: http://socrates.berkeley.edu/~fitelson/few/steel.pdf

Appendix:



Red-Blue to Rue-Bled Hypothesis Grouping Translation Schema

H_0 = {all are red} = {...111...}
H_1 = {n: red$_n$-blue}= {n: ...111$_n$000...; n > 98)
H_2 = {n, m: red$_n$-blue$_m$-red} = {n, m: ....111$_n$0...$_m$111...; m > n > 98}
$N_h = \{H_0, H_1, H_2\}^C$

G_0 = {all are rue} = {...111$_{99}$000...}
G_1 = {n: rue$_n$-bled} = {n: ...111$_{99}$0...$_n$111... ; n > 99, or ...111... if n = 99}
G_2 = {n, m: rue$_n$-bled$_m$-rue} = {n, m: ...111$_{99}$0...$_n$1...$_m$000...; m > n > 99, or ...111$_m$000... if n = 99}
$N_g = \{G_0, G_1, G_2\}^C$