Version Space Algebras and Category Theory

Bailey Wickham

March 4, 2022

Abstract

We give a category theoretic definition of version space algebras and their operations.

1 Introduction

Version Space Algebras (VSAs) are useful as a framework for machine learning in Computer Science, however their associated definitions are limited by this applied context. Many of the constructions can be generalized to be category theoretic.

2 Version Spaces

[] defines a Version Space $VS_{H,D}$ as such:

Definition 2.1 (Version Space)**.** Let the hypothesis space $H = \{f | f : A \to B\}$, where A and B are sets. Let the examples $D = \{(i, o) | i \in A \text{ and } o \in B\}$ Then our version space $VS_{H,D} = \{f | f(i) = o \text{ for all } (i, o) \in D\}$

A version space algebra provides us with a type of "ambient space", H. H provides a context for which the functions exists. The examples, D, are pairs of elements in the domain and range of these functions. Our version space consists of the functions in the ambient space that are consistent with the examples.

In practice the ambient space may be something like "words in our document", or "integers representing row and column numbers", however then only need be functions between (small) sets. Version spaces are used to define all possible actions of a specific type, then constrained to the examples observed in the current document. This is used to generate predictions, where the examples we have used in our document are used to predict future actions.

[] introduces the idea of Version Space Algebras (VSAs) which are version spaces generated by operations on other version spaces. A version space which is defined explicitly, or not defined in terms of operations on other version spaces, is called an *atomic version space*. From a set theory perspective there is no difference between a version space algebra and a version space, there is only a difference in how they are constructed.

There are four main operations defined on a version space.

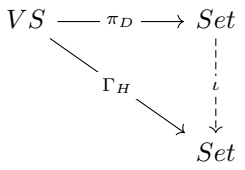
Definition 2.2 (Version Space Union)**.** Let $H_1 = \{f | f : A \to B\}$ and $H_1 = \{f | f : A \to B\}$ be hypothesis spaces with the same domain and range. Let D be a set of examples. Then $VS_{H_1,D} \bigcup VS_{H_2,D} = VS_{H_1 \bigcup H_2,D}$.

3 Category Theory

In the second section we defined a version space and a set of operations on it. In this section we will redefine version spaces in categorical terms.

Definition 3.1 (Version Space (Category))**.** Let A and B be small sets. Then a hypothesis space $H \hookrightarrow Hom_{Set}(A, B)$. Let the examples, D, be a set of pairs $\{(a, b) : a \in A, b \in B\}$ (a, b) that project via maps π_A, π_B into A and B respectively.

A version space is the set $VS_{H,D} = \{f : D \to \Gamma_f \text{ is injective}\}$, where Γ_f is the graph functor: $\Gamma_f = \{(a,b)|b = f(a)\} \subseteq A \times B$. Therefore in VSA we have a commutative diagram where π_D is a projection of $VS_{H,D}$ into D in Set. If VS is a version set, ι should be injective.



This construction exactly mirrors the set theoretic construction given by [] et al. Union and Intersection are defined almost identically to the set theory version. [] these are derived directly from ZFC? maybe?

Definition 3.2 (Union (Category theory))**.** $VS_{H_1,D} \bigcup VS_{H_2,D} = F_D(H_1 \bigcup H_2)$

Definition 3.3 (Category of VSAs)**.** Let **VSA** be the category of version spaces. We define the category of all version spaces.

Objects in our category are of the form $VS_{H,D}$, over any H and D. A morphism assigns $VS_{H_1,D_1} \to VS_{H_2,D_2}$ if and only if $H_1 \subseteq H_2$.

Define $\Gamma_H = \{\Gamma_f : f \in H\}$ $\Gamma_f = \{(a,b)|b = f(a)\} \subseteq A \times B$. We define our examples, D, as $D \subseteq \{\Gamma_f | \forall f \in H\}$

Proof. Associativity of morphisms follows from associativity of the inclusion map. $H \subseteq H$, so there exists an identity morphism for every object. □