

Editorial: The Terminology of Machine Learning

Science is a communal endeavor, and communication is essential to its effective operation. Journals play an important role in this process, and the primary goal of *Machine Learning* is to improve communication between researchers in our emerging field. One way to maximize this communication is to encourage clear and consistent terminology, and this is especially important in new fields like machine learning. An author's choice of words can have a major impact on his paper's ability to communicate, and this makes terminology an important concern for both the authors and the editors of any journal.

The fields of artificial intelligence and cognitive science have a long history of terminological disputes. Researchers in both fields have been criticized (both from within and without) for using semantically loaded terms. The pages of journals and proceedings abound with statements that an AI system *understands* natural language, that it *reasons* about some problem, or even that it *learns* new concepts or *discovers* scientific laws.

However, we should not forget that AI and cognitive science are unique fields. Unlike physics and chemistry, they study processes that can be observed through introspection (at least to some extent) by all humans. As a result, our everyday languages already contain terms for most of the mechanisms that we are interested in explaining. Thus, it is natural for us to describe our systems as “understanding” or “learning”, even when we realize that our current models of these processes only shadow the human versions. In contrast, “hard” scientists are free to invent entirely new names for concepts, or to use existing terms (such as *charm* and *color*) in new senses without fear of confusion.

But despite the criticisms, I would argue that AI and related fields are in little danger from their use of anthropomorphic terms. Most researchers are well aware of the limitations of their systems, and realize they have far to travel before approaching the human's capacity to reason, understand, or learn. In the meantime, these terms remind us of our ultimate goals, and keep too many from being sidetracked into well-defined (but intrinsically dull) problems. In fact, I would encourage machine learning researchers to continue talking about learning, discovery, and analogy, though they should certainly go beyond these general terms to present the details of their methods.

However, words can be used (and thus misused) in many ways, and I think that AI in general, and machine learning in particular, are in danger from a quite different terminological trend. Although words often serve to clarify thought, they can also

lead thought to *crystalize* into rigid patterns, and this can be devastating to developing fields that are sorely in need of new ideas and approaches. Let us consider two examples of this trend from the field of machine learning.

Considerable research has focused on the task of learning from examples, in which one formulates some rule or description to cover a set of training instances. In this context, the term *generalization* has often been used to describe the learning process. Unfortunately, this word has been used in two rather different senses, and authors have made few attempts to clarify which sense they intend.¹ Let us summarize the two senses:

- (1) *Generalization* is any method for generating some rule or concept description from a set of instances; this describes a relation between the inputs and outputs of a learning system.
- (2) *Generalization* is any method for moving from more specific descriptions to more general descriptions; this describes a process for systematically searching the space of descriptions.

The first meaning refers to *any* method for generating descriptions from instances, regardless of whether that method requires one or many instances, whether it operates incrementally or nonincrementally, or whether it moves from specific to general descriptions or vice versa. All learning systems carry out *generalization* in this sense of the term.

The second sense is more limited, referring only to learning methods that systematically search the space of descriptions in a specific-to-general manner (usually in response to new positive instances that were not matched by an earlier description). These methods can be distinguished from alternative approaches that search the space in a general-to-specific fashion; the latter have been called *discrimination* or *specialization* methods. In contrast, the first meaning of *generalization* subsumes specific-to-general methods, general-to-specific methods, bi-directional methods (such as version spaces), and “lateral transfer” methods such as analogy.

The danger in such dual meanings is that one can easily confuse the two senses. In this case, one is tempted to believe that *all* methods for generating a rule or description from data (sense 1) must start with specific descriptions and move to more general ones (sense 2). This thesis is false, but I suspect that many researchers effectively hold to it nevertheless.

Such a confusion can occur if one fails to distinguish between the space of instances and the space of rules or concept descriptions. The first sense above describes

¹ The term *generalization* has also been used as a noun, to indicate any general rule or description; this usage is most consistent with sense (1) in the text. In experimental psychology, the term has yet another meaning. *Generalization* is the tendency, given the behavior that response R is produced under stimulus S_1 , to produce the response R given another stimulus S_2 that is similar to S_1 . Note that this is a *performance* definition, in that it involves no changes in behavior over time.

the act of transforming instances into rules, while the second describes the act of transforming rules into more general rules. The confusion is encouraged by the fact that most specific-to-general learning methods base their initial rules or descriptions on a single positive instance.

Thus, the first step in a specific-to-general method's search through the space of descriptions *looks* like it is moving from an instance to a general rule. However, this is a degenerate case, and the difference between sense (1) and (2) becomes apparent as soon as the second positive instance is processed. At this point, it becomes clear that rules or descriptions are being modified rather than instances. Unfortunately, it is tempting to focus on the first step of this process, in which the differences between the two senses of *generalization* are obscured.

The fact that most work on learning from examples has employed specific-to-general methods has further encouraged the confusion between the two senses of *generalization*. As a result of this confusion, researchers may well turn to specific-to-general methods, even when a general-to-specific approach may be more appropriate for their purposes. A failure to distinguish between two senses of a term can seriously limit one's options.²

The recently coined term *similarity-based* is even more confusing. Although the problem here is slightly different, the result is the same. This word has been used to encompass all empirical learning methods that rely on large collections of data, but the term itself suggests this set of methods is limited to schemes that find *similarities* between positive instances of a concept. In fact, much of the work on learning from examples has used such techniques, and this is probably how the term originated.³

However, other learning methods rely on exactly the *opposite* strategy, noting *differences* that exist between positive and negative instances. Still other (model-driven) empirical methods generate rules or descriptions using knowledge of the domain, and use instances (data) only to *test* these descriptions. Thus, researchers who use the term *similarity-based* to cover all empirical learning methods may blind themselves (and others) to alternative approaches that might be more useful for their application.

So much for the problem, but what is the solution? In general, I would encourage machine learning researchers to think carefully about the terms they use, and to

² This particular confusion should only affect work on empirical learning methods. The more recently developed explanation-based approaches often formulate a rule or description from a single instance, and do not systematically explore the space of descriptions. Thus, only the first sense of *generalization* applies to explanation-based methods. Although one *can* view these methods as selectively dropping conditions and replacing constants with variables, one can just as easily view them as adding conditions and replacing variables with constant terms.

³ One might argue that *similarity-based* describes the relation between inputs and outputs of empirical learning systems, rather than the learning process itself. However, this sense of term is misleading as well. Although all such systems find aspects held in common among the positive instances, they also find *differences* between positive and negative instances.

clarify them when they are ambiguous. In particular, I would strongly encourage authors to abandon the term *similarity-based*, since it is misleading and since the more neutral term *empirical* already has a long history in the field.

The situation with *generalization* is more complex, since both senses of this word have been used for over a decade in machine learning. Personally, I would prefer to see the term abandoned entirely, replacing each meaning with a less ambiguous expression. Elsewhere, I have proposed the term *characterization* to describe the process of formulating a general description, and the term *specific-to-general* seems quite adequate to replace the second sense of *generalization*.

Of course, such decisions must ultimately reside with the authors themselves, and *Machine Learning* will not try to enforce any particular terminology on its contributors (though we may well ask them to clarify their terms). However, the fate of these particular terms is much less important than a general concern with clear writing and clear thinking. I trust the reader agrees that these are important goals, and I sincerely hope they will become central aspects of the machine learning community.

Pat Langley
University of California, Irvine
Langley@CIP.UCI.EDU