LEARNING WITH A QUALITATIVE DOMAIN THEORY BY MEANS OF PLAUSIBLE EXPLANATIONS

Gerhard Widmer
Department of Medical Cybernetics and Artificial Intelligence,
University of Vienna
and
Austrian Research Institute for Artificial Intelligence

Abstract

This chapter describes an approach to learning on the basis of a qualitative domain theory. The theory consists of a mixture of strict rules and general dependency statements. The domain theory supports *plausible explanations* of training instances. These explanations are used to create initial concepts via a kind of 'plausible EBG', and also to guide subsequent empirical generalization of learned concepts. The method has been implemented in a system that learns to solve complex problems in the domain of tonal music. This chapter presents the application domain, describes the learning method (with special emphasis on the plausible inference strategies used), presents empirical results, and shows how this approach naturally leads to a framework for multistrategy learning.

1 INTRODUCTION

In recent years, there has been a growing awareness among researchers in machine learning that, in their pure forms, both inductive learning and explanation–based generalization suffer from severe limitations that restrict their applicability in many domains. The subsequent research has concentrated mainly on combinations of empirical and explanation–based learning (Michalski and Kodratoff, 1990). Newer approaches to this problem now try to flexibly integrate several learning strategies so that a system can dynamically apply these strategies in response to the specific requirements of the learning task (see, e.g., Widmer, 1989; Tecuci and Kodratoff, 1990; Tecuci and Michalski, 1991). That also necessitates investigations into the possibilities of guiding or constraining learning by reasoning methods other than pure deductive or inductive inference. This chapter will be devoted to such matters.

There are many motivations for studying these issues. For one thing, psychological evidence suggests that, in the absence of precise knowledge, people employ various forms of *plausible reasoning* to arrive at explanations or predictions (Collins and Michalski, 1989), and that such weak forms of inference can considerably constrain the set of hypotheses a person is willing to make. For instance, [1] a person, when asked whether she thinks that Taiwan grows rice, might remember that growing rice has something to do with the amount of water available, and hence with the amount of rainfall in the area – a very general and abstract piece of knowledge. Now if the person knows that there is a lot of rainfall in China and that China does grow rice and that Taiwan also has high rainfall, then she might conclude that, yes, Taiwan does probably grow rice. (This, incidentally, is an example of *determination–based analogy* (Russell, 1987)). This illustrates how imprecise background knowledge (about the relevance of rainfall to growing rice) can make a similarity–based judgement more plausible. To continue the example, suppose that, in addition, the person also knows that the ability of a country to grow rice is roughly *positively proportionally* related to the amount of rainfall in the country, that is, the more rainfall there is, the better the country's chances to grow rice (ignoring other factors). Given this knowledge, the person might be willing to conclude that Taiwan grows rice even in the *absence* of the similar instance China, because high rainfall may plausibly be associated with high possibility of growing rice. Both of these examples show how imprecise background knowledge (no knowledge about the exact shape of the function connecting rainfall to growing rice) can be used to produce plausible inferences, given only few examples. The connection to *learning* should be obvious: there, plausible inference

[1] The following example has been adapted from (Collins and Michalski, 1989).

can be used in an analogous way to judge the relative plausibility of hypothesized generalizations.

In addition to such psychological considerations, there are also practical reasons for studying learning with qualitative background knowledge and plausible reasoning. In many domains, complete *a priori* knowledge is simply not available, and hence, pure explanation–based learning is out of the question. And even when a lot is known about a domain, the knowledge is often not precise enough to be cast in the form of a strict deductive domain theory. Rather, it may be very abstract knowledge about the structure of the domain and about general dependencies between various parameters. A system wishing to use such knowledge for learning must employ novel reasoning techniques and integrate these into the learning process.

This chapter presents a model of knowledge–intensive learning that was motivated by such considerations. The basis for learning in this model is a *qualitative* domain theory that consists of a mixture of strict rules and general dependency statements. An implemented system will be presented that realizes the model and learns to solve complex problems (harmonization) in the domain of tonal music. The target concepts are rules specifying necessary conditions for harmonization decisions. The main learning mechanism is a kind of *Plausible Explanation–Based Learning* (DeJong, 1989) where the system tries to construct a plausible explanation of the training instance, using its qualitative domain theory, and then generalizes the explanation to arrive at a general concept. The following sections will describe how this is done, and will also show how these plausible explanations can be used to guide subsequent empirical generalization of the learned concepts. The overall effect of the method is that the available background knowledge is maximally exploited to guide the learning process, even though it is incomplete and too abstract for classical explanation–based generalization (EBG) (Mitchell et al., 1986).

Generally, the author's research in knowledge–based learning has been inspired by the idea that for a learner, trying to learn entails trying to relate, by some reasoning mechanisms, the incoming information to the knowledge it already possesses (Tecuci, 1992 – CHAPTER IN THIS BOOK; Michalski, 1992 – CHAPTER IN THIS BOOK). The work presented here shows how methods of plausible inference can be used to verify that the learner's background knowledge at least weakly implies the new information (notwithstanding the abductive element in the reasoning process – see section 3.1.2). In contrast to classical EBG, however, learning is not logically redundant in this model. The model is another instantiation of a framework that was already proposed in (Widmer, 1989). There, it was argued that the EBG learning model offers a natural basis for many forms of multistrategy learning if we only generalize our notion of what an *explanation* is. If explanations can include non–deductive types of inference and if they can refer to information from outside the current training instance, then multistrategy learning behaviour can be achieved naturally within a simple and uniform framework. A similar course of action is being pursued by Tecuci (1992; CHAPTER IN THIS BOOK).

In the presentation that follows, examples from the particular domain of application will be used to illustrate various features of the learning method. This will also give the reader a feeling for the complexity of the task. Readers not familiar with musical issues should not worry, however. It is not necessary to understand the musical details; it is the structure of the examples and explanations that matters. Also, it should be understood that there is nothing music−specific in the learning method itself; the method is applicable to any domain that can be modelled as a qualitative dependency hierarchy. The author hopes that the generality of the method will become clear throughout the presentation.

2 LEARNING PROBLEM AND DOMAIN THEORY

First, a short description of the particular learning task is needed to set the stage for the following presentation. The system is to learn rules for solving a class of problems in tonal music, namely, harmonizing given melodies by attaching harmonies/chord symbols to the notes of a melody in a musically meaningful way. This is the kind of problem a guitar player, say, is confronted with when s/he is asked to accompany a singer and knows only the melody of the song. Examples of harmonized melodies appear later in this chapter.

This domain is a good example of problem areas where there is no precise theory that could be used to prove the correctness of training instances. However, one can easily come up with a lot of general, abstract intuitions that identify potentially relevant domain features and relationships. In the case of harmonization, one possible way to 'explain' specific harmonizations is to use general knowledge about how people listen to and what they expect from harmonized music. This is the approach that has been chosen for the current project: the *domain theory* is a general *qualitative model* describing in abstract terms how people perceive ('hear') simple tonal music. The model is meant to be a psychologically and musicologically plausible hypothesis about musical listening; it was conceived independently of the particular learning task. Readers interested in the music−theoretic aspects of the model are referred to (Widmer, 1992) for a detailed description.

More precisely, the domain theory is an *abstraction hierarchy* that relates certain audible effects of musical situations to more abstract perceivable effects. Its structure resembles that of an EBG−type domain theory. However, it is qualitative in that internal variables in the model can only take qualitative values from the domain {extremely_low ... extremely_high} and, more importantly, most of the relationships between parameters are described in a qualitative way only. Specifically, the domain theory contains statements of the following form:

- *Partial monotonic dependencies:* [2] A statement q+(A,B) means that parameters A and B are positively monotonically related, or in other words, "if A increases or has a high value, B will also, *all other things being equal*". q+(A,B) does *not* mean that A is the only factor on which B depends, nor does it mean that A must always be present when B is. q− is interpreted analogously for inverse proportionality. An example:

  q+ (relative_chord_distance(Chord1, Chord2, D), contrast(Chord1, Chord2, C)).

  ("Given a sequence of two chords (Chord1, Chord2), the listener may experience a feeling of contrast C between the chords which is positively proportionally related to the harmonic distance D (along the circle of fifths) between the chords")

- *Additional proportionality relations*: Statements of the form addq+(A,B) and addq−(A,B) are to be interpreted like q+ and q−, respectively, except that they specify only *additional influences*; that is, they are relevant only if there is already some reason to believe that B holds or has a particular value. An example:

  addq+ (metrical_strength(Chord2, S), contrast(Chord1, Chord2, C)).

  ("If there is some perceived degree of contrast C, it may be felt the more strongly the stronger Chord2's metrical position S is")

- The domain theory does also contain some *strict deductive rules* (as in standard EBG domain theories); an example:

  ```
  relative_consonance( Chord, Note, extremely_high) :-
            chord_contains_note( Chord, Note).
  ```

  ("The relative consonance between a note and a simultaneously played chord is extremely high if the chord contains the note (in its basic triad)")

In principle, the concept of partial monotonic dependencies is very general; any monotonic function could be hidden behind such an abstract specification of dependency. For instance, for numeric variables X and Y, functions like $Y = \log(X)$, $Y = \exp(X)$, $Y = 5X^3 + 4X + 17.4$ would all satisfy q+(X,Y). For the current project, however, the dependencies are assumed to describe roughly *linear* relationships. This simplifying assumption seems justified for the types of parameters occuring in the present domain of application. Also, since internal variables in the qualitative model have very restricted discrete domains − they range over only five qualitative values: äextremely_low, low, medium,

---

[2] The notation 'q+' was borrowed from Forbus' qualitative proportionality relations (Forbus, 1984). These relations are also related to Michalski's M−descriptors (Michalski, 1983) and the directed dependencies in (Collins and Michalski, 1989).

high, extremely_highü, which are themselves very 'fuzzy' – assuming functional dependencies with very complex shapes would seem to be missing the point. The main role of the qualitative dependencies in plausible reasoning is to support a weak notion of relative plausibility of arguments, i.e., to make certain inferences more plausible than others. All this should be kept in mind when the heuristics for finding plausible explanations are discussed in section 3.1.1.

3 THE LEARNING METHOD

A teacher is assumed to provide pre–classified training examples, where an example is a specific chord in a harmonized piece, along with a particular note in the melody. The *representation* of training instances is just a list of notes and chords along with their basic attributes. The *goal concepts* are good(Chord,Note) and bad(Chord,Note) – i.e., the goal is to learn sets of conditions under which a certain chord will be a good or bad harmonization for a given note. The learned concepts are represented in the form of rules, and the terms 'concept' and 'rule' will be used interchangeably hereafter to refer to the result of learning. The learning scenario is *incremental*, so examples are presented one by one. Two phases can then be distinguished in the process of learning a rule. In phase I, a new, preliminary rule is learned from generalization of a single training instance: the system searches for a plausible explanation of the correctness of the training instance and then compiles the explanation into a general rule. This process could be called 'plausible EBG'. A rule learned in this way can be incrementally generalized later on when new, similar situations are encountered (phase II). Both the decision whether to generalize an existing rule or create a new one, given a new training instance, and decisions as to how to proceed in incremental generalization are based on information provided by the plausible explanation underlying the rule in question. Fig.1 gives a sketch of the learning process. This chapter will be more concerned with the 'plausible EBG' part. Section 3.1 describes how plausible explanations are constructed and how initial concepts are derived from them. Section 3.2 then gives a rough account of how incremental empirical generalization can be made more effective by using information from plausible explanations.

3.1 Phase I: Single instance generalization via 'Plausible EBG'

3.1.1 Constructing plausible explanations

Given a training instance and its classification, the system tries to find an *explanation* of the instance with the help of its domain theory. Such an explanation will be in the form of a justification tree much like in traditional EBG, but will have a different semantics. As the domain theory contains knowledge items of various degrees of strength (strict

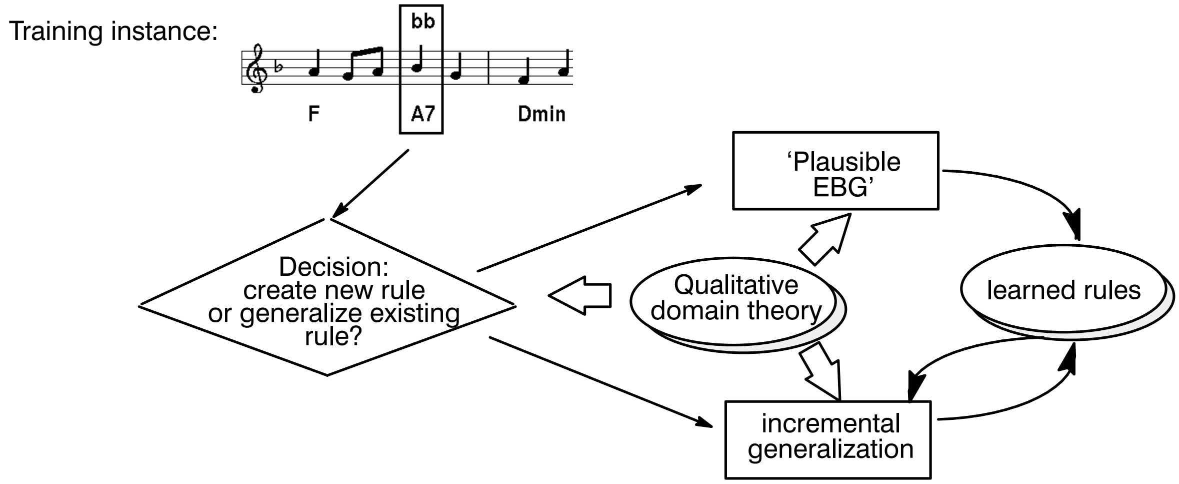


Figure 1: Overall structure of the learning process

rules and qualitative dependencies of various sorts), explanation trees will consist of a mixture of explanatory argument types. Stronger arguments are preferred, so features that are defined by strict rules in the domain theory are explained by standard deductive reasoning. When a feature is defined through directed dependencies, however, only a weaker kind of explanation is possible. Given a particular value y of some feature Y to explain, and knowing that q+(X,Y), one possible way to 'explain' why Y=y is to show that X has a value that is roughly proportional (or reasonably close to proportional) to the value y of Y. [3] Such an 'explanation' can at best be plausible. Additional problems arise because the domain theory has several layers. This means that an explanation will contain chains of such plausible arguments, where the intermediate features are not observable and can only be hypothesized. It is clear, then, that such plausible explanations are by no means unique, and also that they cannot be interpreted as logical proofs in the sense of EBG explanations.

Constructing a plausible explanation is a heuristic search process; the goal is to find the most plausible explanation. Fig.2 sketches a typical situation in this search. The explanation that the system is looking for is constrained from two sides: it is constrained from the *top* (in this case, for instance, the goodness of the example is known to be high) and from the *bottom*, by the features of the training instance itself. It is in the middle, so to speak, where decisions have to be made concerning which parameters to pair with which, and which arguments to include in the explanation and which ones to omit. For example, for some branch of the explanation the system might need an explanation for a certain value y of a parameter Y, and there might be a known dependency of type q+ between

[3] In a very loose sense, one might say that such dependencies are interpreted as 'causal' links. They are assumed to express a directed influence from X to Y, but not necessarily the other way round.

another parameter X and Y. So certain values x of X might be used to *explain* parameter Y – at least partially, since there may be still other factors that are known to influence Y. The possible domains of X and Y are known (see Fig.2); x and y may or may not be known (some values are known by inheritance from above, some are known by inference from features of the training instance). In such a situation, the following kinds of decisions have to be made:

- If x and/or y are not known, which values should be hypothesized for them?
- If several combinations of x and y are possible, which one is the most plausible by itself and makes the explanation as a whole more plausible?
- Should the explanation 'Y=y because X=x' be included at all? Is it consistent with other branches and arguments?

This situation will occur in many places in the process of constructing an explanation. The system makes use of a number of heuristics and constraints in order to make decisions in such a situation:

- *Linearity of qualitative dependencies* (heuristic):
  It is assumed that directed qualitative dependencies describe roughly linear relationships. As a consequence, given two parameters X and Y, related via q+(X,Y), an explanation that assigns values to X and Y that are roughly in the same range of their respective domains will be more plausible than one that pairs a low value for X with a high value for Y, say.

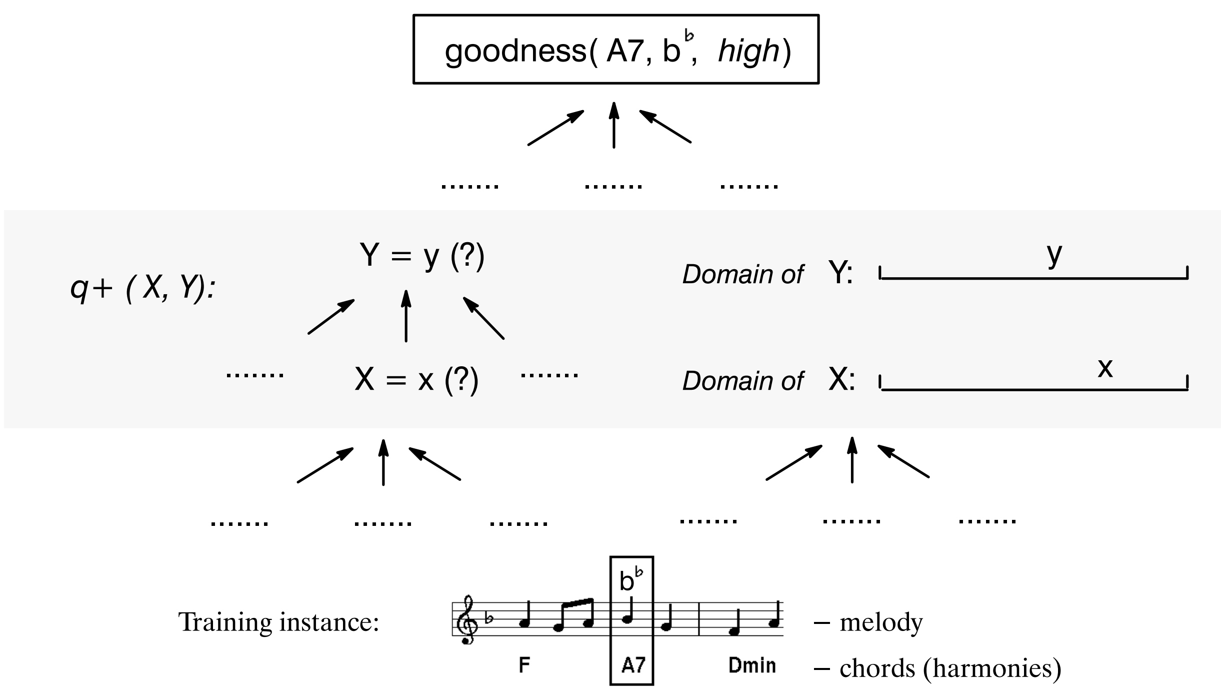


Figure 2: Search for most plausible explanation

- *Special importance of extremal values* (heuristic):
  The above hypothesis is assumed to be particularly true for *extremal* parameter values. That is, it is assumed to be highly unlikely that a value of, say, extremely_high for some parameter Y can be caused by a moderate or low value of another parameter X.

- *Consistency of assignments* (constraint):
  The fact that the qualitative dependencies denote *monotonic* relationships excludes certain combinations of inconsistent parameter mappings. For instance, knowing that q+(X,Y) and that there are no other factors on which Y depends, one cannot use X = moderate to explain Y = extremely_high in one place and X = high to explain Y = high in another. This would conflict with the monotonicity of the relationship between X and Y.

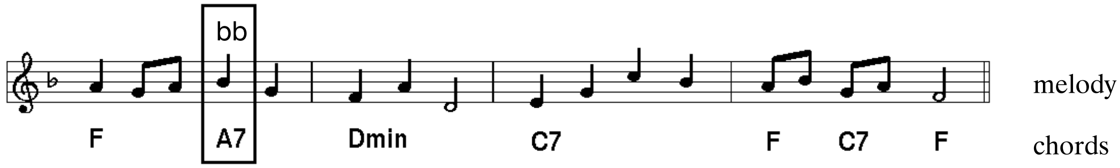
- *Local coherence – agreement of arguments* (constraint):
  Finally, in the case of multiple influences Xi on a parameter Y, only those are included in the explanation of Y that can be made to agree on the value y of Y. The system prefers explanations with many supporting arguments to those with fewer ones.

Underlying these heuristics are two basic assumptions, namely, that multiple influences on a parameter Y are more or less independent in their effect on Y, and, what is more, that multiple influences obey a kind of linear additivity; in particular, negative influences may neutralize positive ones. These assumptions may seem rather strong, but turned out to be adequate for the current application domain.

The heuristics and the system's preference for multiple support of plausible arguments are combined in an *evaluation function* that rates competing (sub–)explanations. It computes a crude estimate of plausibility which is expressed in qualitative terms (see Fig.3). To summarize, the estimated plausibility of an explanation is a function of (1) the degree of "fit", given the known dependencies, between parameters, (2) the number of supporting subexplanations, and (3) the respective plausibilities of these subexplanations.

To find the most plausible explanation of an instance, the system performs a kind of best–first search. The explanation is constructed in a mixed top–down/bottom–up manner: already known instantiations of parameters (e.g. the known value high for goodness in Fig.2) are propagated downward through the domain theory, and partial explanations are then constructed bottom–up and combined into explanations of higher–level features. Evaluation of competing partial explanations is based on the above–mentioned evaluation function.

Fig.3 presents a major portion of an explanation created by the system. The specific instance explained is the chord–note pair <A7,b$^\flat$> (indicated by a box in Fig.3). This tree structure explains why it can plausibly be assumed that listeners will hear the A7




```
                                   type of explanatory link
                                   qualitative measure of plausibility
good(A7,bb)  <– DED (highly_plaus) ––
  [ . . . . . ]                                               qualitative value of
                                                              domain feature
                                                              (from qualitative model)
  goodness(A7,bb,extremely_high)  <– QPLUS (highly_plaus) ––
     interestingness(A7,bb,extremely_high)  <– QPLUS (mod_plaus) ––
       contrast(F,A7,moderate)  <– QPLUS (highly_plaus) ––
          previous_chord(A7,F)  <– TRUE (certain)
          relative_chord_distance(F,A7,moderate)  <– QPLUS (highly_plaus) ––
             distance_on_circle_of_fifths(F,A7,4)  <– TRUE (certain)
     interestingness(A7,bb,extremely_high)  <– QPLUS (highly_plaus) ––
       tension_buildup(A7,bb,extremely_high)  <– QPLUS (highly_plaus) ––
          previous_chord(A7,F)  <– TRUE (certain)
          local_tension_increase(F,A7,high)  <– DED (certain) ––
             intrinsic_chord_tension(F,low)  <– DED (certain) ––
                chord_type(F,triad)  <– TRUE (certain)
             intrinsic_chord_tension(A7,high)  <– DED (certain) ––
                chord_type(A7,7)  <– TRUE (certain)
       tension_buildup(A7,bb,extremely_high)  <– QMINUS (highly_plaus) ––
          relative_consonance(A7,bb,extremely_low)  <– DED (certain) ––
             not_chord_contains_note(A7,bb)  <– TRUE (certain)
             chord_scale(A7,scale(a,mixolydian))  <– TRUE (certain)
             not_scale_contains_note(scale(a,mixolydian),bb)  <– TRUE (certain)
       tension_buildup(A7,bb,extremely_high)  <– ADD.QPLUS (mod_plaus) ––
          salience(bb,moderate)  <– DED (certain) ––
             structural_salience(bb,moderate)  <– TRUE (certain)
             metrical_strength(bb,moderate)  <– TRUE (certain)
  [ . . . . . ]
```

Figure 3: Training instance and part of plausible explanation (operational leaves italicized)

chord as a good harmonization for note $b^{\flat}$ (b−flat). The reader is not expected to understand the musical details of the explanation; the figure is just meant to convey a feeling for the structure and complexity of such explanations. Note that each branch of the explanation is labeled according to the type of argument (DEDuctive, based on q+ or q−, etc.). Each argument is also explicitly annotated with the rough degree of plausibility computed during the search. The annotations are used in the second learning phase − the incremental generalization of learned rules (see section 3.2).

The explanation is then generalized, much like in traditional EBG, by propagating the general goal concept (in this case good(Chord,Note)) through the explanation tree, and the generalized leaves are conjoined to form a new rule (Fig.4). The effect of generalization here is mainly the replacement of specific objects (notes, chords, musical intervals, etc.) by universally quantified variables. Qualitative and quantitative values of domain parameters are generalized only if they are explained by strict rules in the domain theory; arguments based on qualitative dependencies are too uncertain to warrant analytical generalization. They will be generalized incrementally if subsequent examples indicate a need for it (see section 3.2). In this way, the single−instance generalization step serves mainly to select and construct relevant attributes for a first hypothesis and to relate them to the goal concept through a hierarchical explanation structure, which can later provide further guidance in empirical generalization.

```
RULE1:   good(Chord,Note) :-   chord_root(Chord,Root),
                               chord_mode(Chord,major),
                               chord_type(Chord,7),
                               global_key(key(KRoot,KMode)),
                               previous_chord(Chord,PrevChord),
                               distance_on_circle_of_fifths(PrevChord,Chord,4),
                               chord_type(PrevChord,triad),
                               chord_type(Chord,7),
                               not_chord_contains_note(Chord,Note),
                               chord_scale(Chord,scale(SRoot,SMode)),
                               not_scale_contains_note(scale(SRoot,SMode),Note),
                               structural_salience(Note,moderate)
                               metrical_strength(Note,moderate),
                               plausible_local_key(Note,key(LRoot,LMode)),
                               dominant_7_chord(key(LRoot,LMode),Chord).
```

Figure 4: Rule learned by generalizing and compiling explanation

3.1.2 Plausible explanation as constrained abduction

The notion of 'plausible explanation' deserves to be examined in a bit more detail. Although, on the surface, the structure of a plausible explanation very strongly resembles the structure of a 'classical' EBG−type explanation, there are important differences.

First, of course, plausible explanations are not logical proofs. Indeed, as DeJong (1989) has already noted, plausible inferences *per se* are weak; that is, evaluation of the theory in a 'forward' direction would produce many nonsensical statements that are simply not consistent with the 'real world'. It is the actual training instances that make certain inferences more plausible than others. To quote DeJong (1989, p.4), "The existence of the training example itself adds credibility to the faithfulness of the plausible explanation and, therefore, to the new generalized concept." Thus, actual observations (instances) play a much more important role in plausible explanation−based learning than in traditional EBG. Rosenbloom and Aasman (1990) also present a lucid discussion on this topic.

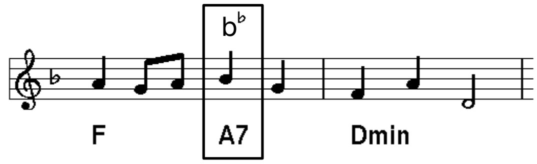


Figure 5: Training instance to be explained: good(A7, $b^{\flat}$)

Generally, constructing plausible explanations involves non−deductive types of inference. For one thing, since general dependency statements usually allow different combinations of actual parameters, it is a matter of heuristics (see above) to choose between these. And second, constructing a plausible explanation often entails hypothesizing relationships that are not directly observable, and that lends a certain abductive quality to plausible reasoning.

The following example, taken from the current application, illustrates this effect. Among other things, the domain theory contains the following two statements of dependency:

q+ (relative_consonance(Chord, Note, C), harmonic_stability(Chord, Note, HS)).
  /* *The harmonic stability HS of a Chord depends on the relative consonance C between the Chord and the Note that it accompanies* */
q+ (stability_in_local_key(Chord, Note, S), harmonic_stability(Chord, Note, HS)).
  /* *The harmonic stability HS of a Chord also depends on the functional stability S of the Chord in the local key (tonality) that is implied in the current context* */

The abstract features relative_consonance and stability_in_local_key are defined by some strict rules in the domain theory, the relevant ones for the example being

relative_consonance(Chord, Note, extremely_low) :− dissonant(Chord, Note).
stability_in_local_key(Chord, Note, extremely_high) :− local_key(Note, Key),
        tonic_chord(Chord, Key).
stability_in_local_key(Chord, Note, high) :− local_key(Note, Key),
        dominant_chord(Chord, Key).

Now assume that the example shown in Fig.5 has been classified as good by the teacher; the system's goal is to explain good(A7, $b^{\flat}$), that is, the A7 chord is a good harmonization for the note $b^{\flat}$ in the melody. In order to explain this, the system must show, among other things, that there is at least a relatively high degree of harmonic stability (HS) in the current situation. Given the two dependencies listed above, this reduces to establishing relative_consonance(A7, $b^{\flat}$, C) and stability_in_local_key(A7, $b^{\flat}$, S) and checking whether the obtained values for C and S do indeed plausibly imply a relatively high value for HS. relative_consonance(A7, $b^{\flat}$, C) is established with the help of the first of the above rules, with the result C = extremely_low (because A7 and $b^{\flat}$ are extremely dissonant). Given the positive proportionality between C and HS that is postulated by the domain theory, this contradicts the assumption that HS is relatively high (which is the current explanation goal). So, the system must at least show that stability_in_local_key(A7, b^b, S) has a high value, which again reduces to finding out what the local key of the musical passage in the vicinity of b^b is, and showing that the A7 chord is either the tonic_chord or the dominant_chord in this key.

Now, local_key is a non−deterministic predicate; there are usually several keys that can plausibly be perceived in a musical situation, so local_key returns, upon backtracking, a set of plausible keys, sorted in the order of decreasing plausibility: Key $\in$ (F_major, B^b_major, E^b_major, G_minor, D_minor). Of these, only D_minor would attribute high stability to the A7 chord (because A7 is the dominant chord of D_minor), so, in order to be able to complete its explanation, the system assumes that the local_key is D_minor and asserts the following explanation branch:

> harmonic_stability(A7, b^b, high) *because*
>   stability_in_local_key(A7, b^b, high) *because*
>   local_key(b^b, D_minor) *and*
>   dominant_chord(A7, D_minor)

The fact relative_consonance(A7,b^b,extremely_low), contradicting the explanation goal, is excluded from this specific explanation.

To reiterate: given just the knowledge in the domain theory, it would have been more plausible to assume that local_key is F_major in the example. However, given a specific training instance that is known to be good and, by implication, to display relatively high harmonic_stability, the system chooses the less plausible (but still possible) assumption that the local key is D_minor, because that allows it to explain the instance. [4] Many plausible arguments have such a distinctly abductive flavor.

Finally, note that the system also faces the problem of deciding which factors to include in an explanation and which ones to exclude. The present program has a built−in bias in favor of descriptions that include as many influences as possible, as long as they appear consistent with the explanation goal and the instance. In general, whether maximally detailed or maximally simple explanations should be considered more plausible and/or more useful depends very much on the characteristics of the application domain and also on the learning task.

---

[4] This situation might also be the starting point for a theory revision episode, where the rules for establishing local_key might be revised so that in similar situations in the future, D_minor would be determined to be the most plausible local key.

3.2 Phase II: Incremental generalization

As noted above, the domain theory was conceived as a general qualitative theory of musical listening for a restricted type of tonal music. It mentions a multitude of factors that might potentially influence a listener's perception of a musical situation. Also, the system prefers detailed explanations to simpler ones. As a consequence, the explanations and the rules learned by plausible EBG tend to be too detailed and specific. That is where the need for incremental empirical generalization arises: the system should not create a new rule for every new instance that is not covered by an existing rule. Rather, if an existing rule 'almost' matches the new instance, it should be generalized to accommodate the instance.

The straightforward way to do this would be to somehow measure the 'distance' between each of the rules and the new instance (by counting matches and mismatches) and generalize a rule if it is 'close enough' to warrant generalization. However, a much higher degree of effectiveness and context−sensitivity of the generalization process can be achieved if empirical generalization is based not only on the *rule* to be generalized, but also on the *explanation* that led to that rule. The plausible explanation can be used to *bias* empirical generalization. Similar observations were already made by Danyluk (1987; see also − DANYLUK−CHAPTER IN THIS BOOK), but only for strictly deductive explanations. Plausible explanations provide more differentiated information that can be exploited − they rely on a richer set of types of explanatory links, and they include explicit plausibility information (see Fig.6).

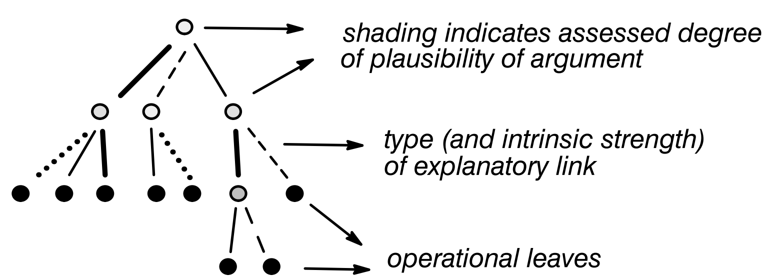


Figure 6: Structure of a plausible explanation

Let X denote a condition in a rule (which corresponds to a leaf in the explanation underlying the rule) that is not satisfied by the current instance. The following criteria are then used to decide whether (and if so, how) to generalize the rule:

- *Possibility of safe generalization:* If X can be generalized to apply to the new instance in such a way that the argument in the original explanation that depends on X still holds, the generalization will be *safe* (at least with respect to the plausible explanation). This criterion also provides a strong bias on the *type* of generalization in cases where there are multiple ways to generalize X – only those will be considered that preserve the validity of the argument that depends on X.

- *Strength of explanatory link:* The system considers the *type* of explanatory link of X; some types are intrinsically more important than others. For instance, arguments based on addq+ or addq– relationships are by definition less salient than those based on q+ or q–, and may thus more safely be dropped.
- *Assessed plausibility of argument:* The system looks at the plausibility with which X was thought to hold in the original explanation. An argument that was not very credible to begin with can more safely be dropped or generalized.
- *Strength of remaining arguments:* If X were to be dropped, how strong would the hypothesis depending on X (some ancestor of X in the explanation tree) still be? That is, how many arguments supporting it remain, and how strong are they? Obviously, if there are strong arguments left that support the original hypothesis, the overall integrity of the explanation is not compromised too much by dropping X.

Information from these heuristics is combined to yield one approximate value indicating how 'likely' it is that generalization of the explanation (and the rule derived from it) is justified. In summary, the explanations serve a dual purpose: first, they provide a measure of *'deep similarity'* – matches and mismatches between instances and rules are rated according to the role they play in an explanation structure; this is a better measure of similarity than just simple counting of syntactic matches. And second, they can provide bias on the type of generalization that seems most plausible. The interested reader can find an example of the heuristics at work in (Widmer, 1991).

4 AN EXPERIMENT

The following experiment was meant to illustrate that this explanation–sensitive approach to incremental generalization can considerably improve the learning performance, both in terms of the number of training instances needed and the generality of the concepts learned. The informed incremental generalization algorithm of section 3.2 (*algorithm 1*) was compared to a simpler incremental generalizer (*algorithm 2*) that did not use the underlying explanations in empirical generalization. Both algorithms used the plausible explanations to derive an initial generalization from the first instance. When considering incremental generalization, however, algorithm 2 based its decisions whether to generalize a particular rule on a general threshold (ratio of matching vs. nonmatching conditions in a rule).

The experiment consisted in first selecting a *target problem* (a piece that the system should be able to harmonize after learning) and several *training pieces*, from which training examples (specific pairs of chords and notes) were then presented to the learners until they could solve the target problem. Target problem and training pieces – all of them beginnings of well–known Piano Sonatas by W.A.Mozart – are shown in Figs. 7 and 8.

Beginning of Sonata # 5 in G major (K.283):

Figure 7: Target problem (piece to be harmonized)

Beginning of Sonata # 16 in C major (K.545):

Beginning of Rondo in Sonata # 16 (K.545):

Beginning of Sonata # 11 in A major (K.331):

Figure 8: Training pieces from which examples were selected

Fig. 9 displays the solution to the target problem that was found by both algorithms after the learning session. (Incidentally, this is more or less the harmonization Mozart himself chose.) The main results of the experiment are summarized in Table 1. The data indicate that paying attention to the underlying explanations in incremental generalization can considerably improve the learning performance: the set of rules learned by algorithm 1 is both more concise and more general, and, what is more, it was learned from a considerably smaller number of training instances.

Figure 9: Solution found after learning

	Algorithm 1 (guided by expl.)	Algorithm 2
Complexity of learned rule base:		
# of rules learned:	9	12
total size of rule base (# literals)	103	129
avg. # literals / rule	11.4	10.8
Cost of learning:		
# instances required	20	32

Table 1: Results of comparative learning experiment

5 DISCUSSION, RELATED WORK, AND FUTURE DIRECTIONS

To summarize briefly, this chapter has presented a method for learning rules from examples with the help of a qualitative domain theory that consists mainly of qualitative dependency relations. The theory is too weak to permit standard explanation−based learning. Plausible inference strategies are used to guide the learning process, both via a kind of 'plausible EBG' and by biasing incremental empirical generalization. Section 3.1 described the inference techniques used to construct plausible explanations and discussed their non−deductive nature. Section 3.2 showed how underlying explanations can be used to guide incremental generalization of learned concepts.

An additional important role of the domain theory that should be pointed out here is the *dynamic construction* of an appropriate hypothesis language, depending on the context. The language in which instances are represented and the language for explanations and rules are not the same. In section 3, it was briefly mentioned that training instances are represented simply as lists of notes and chords along with their basic attributes (such as duration, pitch, chord type, etc.). Explanations and rules, on the other hand, refer to higher−level concepts and various relations (see Figs.3 and 4). These higher−level concepts are introduced by the domain theory and enter into learned concepts because of the particular level of operationality defined in the system. The method thus demonstrates the utilization of a qualitative domain theory for *constructive generalization* (Michalski, 1983).

A general problem with plausible explanations as described here is that they are rather weak, being based as they are on abstract background knowledge and on single examples. This uncertainty effect multiplies if explanations contain chains of plausible arguments. That is why, in the current system, analytical generalization is applied very cautiously (see section 3.1.1). It is in the incremental generalization phase that the appropriate degree of generality is determined empirically, by analyzing new instances against the background of the explanation of the original concept. An alternative approach would be to strengthen the explanations by checking the validity of plausible inferences against several examples at once. That would mean the loss of full incrementality, but would produce plausible explanations that have more empirical support. Bergadano & Giordana's (1988) ML–SMART framework might, in fact, be extended to form the basis for such a 'multi–instance plausible explanation system'.

Learning with a qualitative domain theory has also been investigated by DeJong (1989; 1990). He presented a method for learning in continuous domains, where the domain is modelled in *Qualitative Process Theory* terms (Forbus, 1984). The main advances of the method described here over DeJong's approach are a) the use of the domain theory for constructive (instead of just selective) generalization, b) the definition of explicit criteria for assessing the relative plausibility of competing explanations, which makes possible a heuristic search for the 'most plausible' explanation, and c) the methods for exploiting plausible explanations to bias incremental generalization of learned concepts. In case of conflict, DeJong's system simply discards the old explanation and looks for a more consistent one. (This is partly due to the fact that his system generates explanations in a simple–to–complex order.)

The author's own work on a predecessor of the current system resulted in a learning algorithm (Widmer, 1989) that flexibly integrated deductive, analogical (determination–based) and inductive arguments in an explanation–based generalizer and thus exhibited multistrategy learning behaviour, depending on the knowledge available. Tecuci and Michalski (1991; see also TECUCI–CHAPTER IN THIS BOOK) have developed a similar approach based on learning from plausible justifications. They integrate different inference types (deduction, determination–based analogy, abduction, and empirical generalization) in a plausible explanation system. The work described in the present chapter adds a new, complex inference type – hypothesizing consistent explanations on the basis of directed qualitative dependencies – to this collection. On the other hand, empirical generalization is done outside of the explanation process, in a separate learning phase. So one immediate goal for further research is the integration of other types of reasoning (analogy and empirical similarity arguments, various forms of abduction, etc.) directly into the explanation process. That would allow explanations to refer to information from outside the current training instance and would naturally lead to generalization effects during explanation.

A problem that needs to be addressed is the specialization of overly general concepts. In the current application, the richness of the domain theory and the fact that the heuristics for including plausible arguments in an explanation (section 3.1.1) are very much on the permissive side, render specialization virtually unnecessary: Initial concepts constructed by plausible EBG are very specific, and finding the 'correct' degree of generality is then a matter of stepwise (careful) generalization. However, overgeneralization may become a problem in other applications. Investigations are under way to find out whether plausible reasoning can also be used to help in the problem of effectively specializing concepts in the face of conflicting evidence.

The improvement of the qualitative domain theory in response to specific experiences is another interesting topic for further research. Currently, the learning problem is restricted to concept learning, with the underlying domain theory remaining unchanged, being used only to guide the system in acquiring rules. More research is needed both on the automatic *refinement* of the domain theory (i.e., filling in missing details, making abstract relationships more precise) by induction from examples, and the *revision* of the theory in response to incorrect generalizations or abductive explanation needs (see, e.g., the local_key example in section 3.1.2).

In conclusion, the author hopes that the work presented in this chapter is another step in the direction of flexible multistrategy learning, both from a theoretical and from a practical point of view. The chapter has shown how qualitative background knowledge can support powerful explanation methods, and how these methods, when integrated in an EBG−like 'Explain, Generalize, and Compile' schema, lead to very effective learning. The idea of extending the notion of 'explanation' to include weaker kinds of inference is now gaining more and more popularity. In principle, a wide variety of types of knowledge and an equally wide variety of inference types, including abduction and determination− or similarity−based analogy, can be used in plausible reasoning. If all these can be integrated directly into the plausible explanation process, flexible multistrategy learning behaviour will naturally emerge.

On the practical side, it seems worthwhile to emphasize once again the idea of using qualitative models as powerful tools for describing common, abstract domain knowledge. There are many domains where qualitative models are much easier to obtain than precise domain theories. For such domains, the approach outlined here appears very promising. The complexity of the musical application − the entire domain theory comprises about 1000 lines of Prolog code − could be taken as an indication that real−world problems are within the reach of this approach.

ACKNOWLEDGMENTS

The author would like to thank Prof. Robert Trappl for his support and Christian Holzbaur, Werner Horn, and Igor Mozetic for helpful comments. Special thanks are due to the editors of this book for their clarifying and challenging comments. This research was sponsored in part by the Austrian *Fonds zur Foerderung der wissenschaftlichen Forschung* under grant P8756−TEC. Financial support for the Austrian Research Institute for Artificial Intelligence is provided by the Austrian Federal Ministry for Science and Research.

References

Bergadano, F. and Giordana, A., "A Knowledge Intensive Approach to Concept Induction," in *Proceedings of the Fifth International Conference on Machine Learning*, Ann Arbor, MI, pp. 305−317, 1988.

Collins, A. and Michalski, R.S., "The Logic of Plausible Reasoning: A Core Theory," *Cognitive Science,* Vol. 13, No. 1, pp. 1−49, 1989.

Danyluk, A., "The Use of Explanations for Similarity−Based Learning," in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI−87),* Milano, Italy, Milano, Italy, pp. 274−276, 1987.

−−−−−−− , "GEMINI: An Integration of Analytical and Empirical Learning," in *Machine Learning: A Multistrategy Approach, Vol. IV,* R.S.Michalski and G.Tecuci (Eds.), Morgan Kaufmann, San Mateo, CA, 1992.

DeJong, G., "Explanation−Based Learning with Plausible Inferencing," in *Proceedings of the 4th European Working Session on Learning (EWSL−89),* Montpellier, France, pp. 1−10, 1989.

−−−−−−− , "Explanation−Based Control: An Approach to Reactive Planning in Continuous Domains," in *Proceedings of the DARPA Workshop on Innovative Approaches to Planning, Scheduling, and Control,* San Diego, CA, pp. 325−336, 1990.

Forbus, K.D., "Qualitative Process Theory," *Artificial Intelligence,* Vol. 24, No. 1−3, pp. 85−169, 1984.

Michalski, R.S., "A Theory and Methodology of Inductive Learning," in *Machine Learning: An Artificial Intelligence Approach, Vol. I*, R.S.Michalski, J.G.Carbonell, and T.M.Mitchell (Eds.), Tioga, Palo Alto, CA, 1983.

−−−−−−− , "Inferential Learning Theory as a Conceptual Basis for Multistrategy Learning," in *Machine Learning: A Multistrategy Approach, Vol. IV,* R.S.Michalski and G.Tecuci (Eds.), Morgan Kaufmann, San Mateo, CA, 1992.

Michalski, R.S. and Kodratoff, Y., "Research in Machine Learning: Recent Progress, Classification of Methods, and Future Directions," in *Machine Learning: An Artificial Intelligence Approach, Vol. III*, Y. Kodratoff and R.S. Michalski (Eds.), Morgan Kaufmann, San Mateo, CA, 1990.

Mitchell, T., Keller, R., and Kedar−Cabelli, S., "Explanation−Based Generalization: A Unifying View," *Machine Learning,* Vol. 1, No. 1, pp. 47−80, 1986.

Rosenbloom, P. and Aasman, J., “Knowledge−Level and Inductive Uses of Chunking (EBL),” in *Proceedings of the 8th National Conference on Artificial Intelligence (AAAI−90),* Boston, MA, pp. 821−827, 1990.

Russell, S., *Analogical and Inductive Reasoning*, Ph.D. thesis, Report STAN−CS−87−1150, Stanford University, Stanford, CA, 1987.

Tecuci, G., “Understanding−based Multistrategy Learning,” in *Machine Learning: A Multistrategy Approach, Vol. IV,* R.S.Michalski and G.Tecuci (Eds.), Morgan Kaufmann, San Mateo, CA, 1992.

Tecuci, G. and Michalski, R.S., “A Method for Multistrategy Task−adaptive Learning Based on Plausible Justifications,” in *Proceedings of the 8th International Workshop on Machine Learning,* Evanston, Ill., pp. 549−553, 1991.

Tecuci, G. and Kodratoff, Y., “Apprenticeship Learning in Imperfect Domain Theories,” in *Machine Learning: An Artificial Intelligence Approach, Vol. III*, Y. Kodratoff and R.S. Michalski (Eds.), Morgan Kaufmann, San Mateo, CA, 1990.

Widmer, G., “A Tight Integration of Deductive and Inductive Learning,” in *Proceedings of the 6th International Workshop on Machine Learning*, Ithaca, N.Y., pp. 11−13, 1989.

−−−−−−−− , “Using Plausible Explanations to Bias Empirical Generalization in Weak Theory Domains,” in *Proceedings of the 5th European Working Session on Learning (EWSL−91),* Porto, Portugal, pp. 33−43, 1991.

−−−−−−−− , “Qualitative Perception Modelling and Intelligent Musical Learning,” *Computer Music Journal,* Vol. 16, No. 2, 1992.