Modeling Scientific Problem Solving by DOP

Rens Bod (rens@science.uva.nl)
Institute for Logic, Language and Computation
University of Amsterdam, and
School of Computing, University of Leeds
Plantage Muidergracht 24, 1018 TV Amsterdam, NL

Abstract

This paper deals with the problem of derivational redundancy in science, i.e. the problem that there can be extremely many different explanatory derivations for a phenomenon, while students and experts tend to come up with only one derivation. Given the remarkable agreement among humans in deriving phenomena, we need to have a story of how to select from the space of possible derivations of a phenomenon the derivation that humans come up with. In this paper we argue that the problem of derivational redundancy can be solved by a notion of "shortest derivation", by which we mean the derivation that can be constructed by the fewest (and therefore largest) partial derivations of previously derived phenomena that function as "exemplars". We show how a model of exemplar-based reasoning, known as DOP, can be used to select the shortest derivation. We evaluate DOP on a corpus of phenomena from classical and fluid mechanics that were derived by fourth-year physics students, showing that the shortest derivation closely corresponds to the derivations that humans construct.

Keywords: Problem Solving; Exemplar-Based Reasoning; Derivational Redundancy; Case-Based Reasoning; Data-Oriented Parsing; Philosophy of Science; Physics.

1 Introduction

This paper deals with the problem of derivational redundancy, i.e. the problem that there can be extremely many different explanatory derivations for a phenomenon while students and experts tend to come up with only one and the same derivation for a phenomenon. Given this remarkable agreement among students, we need to have a story of why humans focus on one derivation and not on others. In this paper we shall argue that the problem of derivational redundancy can be solved by a notion of "shortest derivation". By the shortest derivation of a phenomenon we mean the derivation that can be constructed by the fewest (and therefore largest) partial derivations of previously derived phenomena that function as "exemplars".

The idea that natural phenomena can be explained by modeling them on exemplars is usually attributed to Thomas Kuhn in his account on “normal science” (Kuhn 1970). Kuhn urges that exemplars are "concrete problem solutions that students encounter from the start of their scientific education" (ibid. p. 187) and that "scientists solve puzzles by modeling them on previous puzzle-solutions" (ibid. p. 189). Instead of explaining a phenomenon from scratch, Kuhn contends that scientists try to match the new phenomenon to one or more previous phenomena-plus-explanations.

In similar vein, Philip Kitcher argues that new phenomena are derived by using the same patterns of derivations ("argument patterns") as used in previously explained phenomena: "Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again" (Kitcher 1989, p. 432). Different from Kuhn, Kitcher proposes a rather concrete account of explanation, known as the "unificationist view", which he still links to Kuhn's view by interpreting exemplars as derivations (ibid. pp. 437-8). Yet, we will argue in section 3 that Kitcher’s account does not solve the problem of derivational redundancy.

Thomas Nickles relates Kuhn's view to Case-Based Reasoning (Nickles 2003, p. 161). Case-Based Reasoning (CBR) is an artificial intelligence technique that stands in contrast to rule-based problem solving. Instead of solving each new problem from scratch, CBR stores previous problem-solutions in memory as cases. When CBR begins to solve a new problem, it retrieves from memory a case whose problem is similar to the problem being solved. It then adapts the example's solution and thereby solves the problem. CBR has been instantiated in many different ways and has been used in various applications such as reasoning, learning, perception and understanding (cf. Carbonell 1986; Falkenhainer et al. 1989; Kolodner 1993; Veloso and Carbonell 1993; VanLehn 1998). However, none of these instantiations specifically addresses the problem of massive derivational redundancy.

An instantiation of CBR that does address the problem of derivational redundancy, albeit in a different domain, is Data-Oriented Parsing (DOP). DOP is a natural language processing technique that provides an alternative to rule-based language processing. It analyzes new sentences by modeling them on analysis-trees of previous sentences (Bod 1998; Scha et al. 1999; Collins and Duffy 2002). DOP operates by decomposing the given trees into “subtrees” and recomposing those pieces to build new trees. When a sentence has more than one possible analysis or interpretation -- which is the typical case in natural language -- DOP selects the analysis-tree that is constructed by the “shortest derivation”, which is the tree consisting of the fewest (and therefore largest) subtrees from previous trees (Bod 2000). DOP has been highly successful in solving syntactic and semantic redundancy (“ambiguity”) in natural language understanding (see Manning and Schütze 1999; Bod et al. 2003). In Scha et al. (1999) it is shown how DOP can be defined as an instantiation of CBR.

In the current paper we argue that DOP can also be used for solving derivational redundancy in physics. The DOP approach may be particularly suitable to tackle the redundancy problem because of the analogy between explanatory derivations in physics and tree structures in linguistics and logic. If we can convert explanatory derivations into trees, we can directly apply the DOP approach to the redundancy problem. That is, when a phenomenon has more than one derivation tree, DOP proposes to select the tree that can be constructed by the fewest subtrees from trees of previously derived phenomena.

In order to do so, we will first show in section 2 how derivations in physics can be interpreted as trees, and how explanations of new phenomena can be constructed by combining subtrees from previously explained phenomena. In section 3 we give an in-depth exploration of the problem of derivational redundancy and argue that DOP's notion of shortest derivation can solve this problem. The resulting DOP model, which we may term "data-oriented physics", is evaluated in section 4 on a corpus of phenomena from classical and fluid mechanics that were derived by fourth-year physics students. It turns out that there is a very close correspondence between the derivations constructed by humans and DOP's notion of shortest derivation.

2 Extending DOP to Scientific Explanation

What do derivational explanations in physics look like? Let's start with a simple textbook example. Consider the following derivation of the Earth's mass from the Moon's orbit in the textbook by Alonso and Finn (1996, p. 247):

> Suppose that a satellite of mass m describes, with a period P, a circular orbit of radius r around a planet of mass M. The force of attraction between the planet and the satellite is $F = GMm/r^2$. This force must be equal to m times the centripetal acceleration $v^2/r = 4\pi^2 r/P^2$ of the satellite. Thus,
>
> $$4\pi^2 mr/P^2 = GMm/r^2$$
>
> Canceling the common factor m and solving for M gives
>
> $$M = 4\pi^2 r^3/GP^2.$$

By substituting the data for the Moon, $r = 3.84\ 10^8$ m and $P = 2.36\ 10^6$ s, Alonso and Finn compute the mass of the Earth: $M = 5.98\ 10^{24}$ kg. In doing so, Alonso and Finn abstract from many features of the actual Earth-Moon system, such as the gravitational forces of the Sun and other planets, the magnetic fields, the solar wind, etc. Albeit heavily idealized, the derivation provides a concrete problem solution on which various other (idealized) phenomena can be modeled. In fact, Alonso and Finn reuse parts of this derivation to solve problems such as the velocity of a satellite and the escape velocity from the Earth.

In order to create a DOP model for derivational explanation, we first need to represent derivations of the type above by tree structures. Analogous to proof trees in formal logic, a tree structure of a natural phenomenon indicates how a mathematical description of the phenomenon is compositionally derived from laws, antecedent conditions and other knowledge. Figure 1 shows how the derivation for the Earth's mass above may be turned into a tree. This tree represents the various derivation steps (insofar as they are mentioned in the example above) from general laws to an equation of the mass M. We will refer to such a tree as a *derivation tree*. A derivation tree is a labeled tree in which each node is annotated with a formula; the boxes are only meant as convenient representations of these labels. The formulas at the top of each "vee" (i.e. each pair of binary branches) in the tree can be viewed as premises, and the formula at the bottom of each "vee" can be viewed as a conclusion which is arrived at by simple term substitution. The last derivation step in the tree is not formed by a "vee" but consists in a unary branch which solves the directly preceding formula for a certain variable (in the tree above, for the mass M). Thus, in general, a unary branch refers to a mathematical derivation step that solves an equation for a variable, while a binary branch refers to a physical derivation step which introduces and combines physical laws or conditions (or other knowledge such as phenomenological corrections and coefficients).

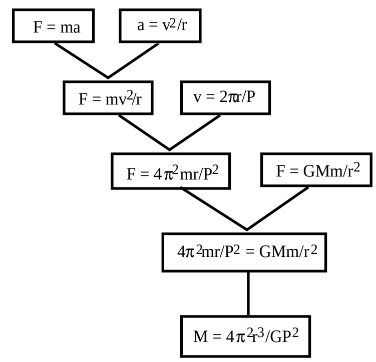

Figure 1: Derivation tree for the Earth's mass

Note that a derivation tree corresponds to the notion of deductive-nomological (D-N) explanation of Hempel and Oppenheim (1948). In the D-N account, a phenomenon is explained by deducing it from general laws and antecedent conditions. D-N explanations usually focus on the initial premises (laws and conditions) and the final conclusion (the phenomenon). But they can just as well represent the intermediate steps as derivation trees do. For every derivation tree there is a corresponding D-N explanation and vice versa. Although the D-N account is known to suffer from various shortcomings and is nowadays superseded by other approaches (cf. Friedman 1974; Kitcher 1989), most derivations in textbooks basically follow this scheme. In this paper, we will focus on fully fleshed-out derivations that describe the various steps from premises to conclusion, because that's the kind of derivations humans construct (see section 4). Our main reason for representing derivations by trees is of course that we can then apply the DOP approach to the problem of derivational redundancy, as we shall see in the next section.

But before we can do this, we will need to demonstrate how DOP builds new explanations out of previous ones (since our solution to the redundancy problem is defined in terms of subtrees of previous trees). Consider the following subtree in figure 2 which is obtained from the derivation tree in figure 1 by leaving out the last derivation step (i.e. the solution for the mass M).

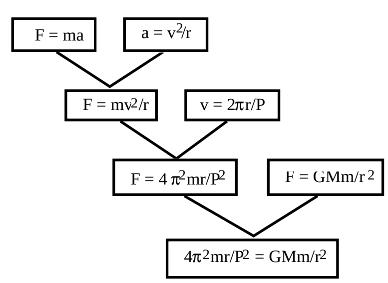

Figure 2: A subtree from the tree in figure 1

This subtree can be applied to various other situations. For instance, in deriving the regularity known as Kepler's third law (which states that r^3/P^2 is constant for all planets orbiting around the Sun, or satellites around the Earth if you wish) the subtree in figure 2 needs only to be extended with a mathematical derivation step that solves the last equation for r^3/P^2, as represented in figure 3.

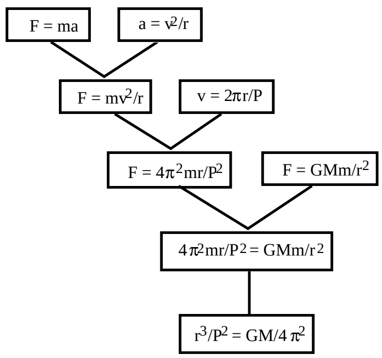

Figure 3: Derivation tree for Kepler's third law

In a similar way we can also derive the distance of a geostationary satellite, namely by solving the subtree in figure 2 for r and taking P as the rotation period of the Earth.

It is of course not typically the case that derivations involve only one subtree. In deriving the velocity of a satellite at a certain distance from a planet, we cannot directly use the large subtree in figure 2. Instead, analogous to DOP models for natural language, we *decompose* the tree in figure 1 into two smaller subtrees and *recompose* them by term substitution (represented by the operation "∘") and finally solve for the velocity v in figure 4.

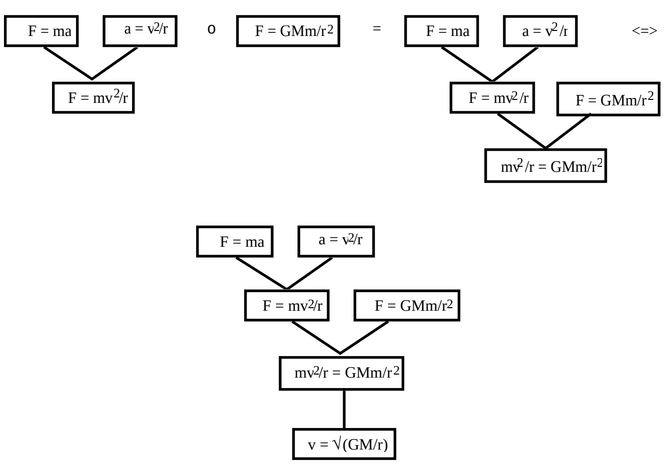

Figure 4: Deriving a phenomenon by combining subtrees

Figure 4 shows that we can create new derivation trees by combining subtrees from previous derivation trees. Note that subtrees can be of arbitrary size: from single equations to combinations of laws, up to entire derivations.

The notion of term substitution, though widely used in rewriting systems, may need some further specification. The term-substitution operation "∘" is a partial function on pairs of labeled trees; its range is the set of labeled trees. The combination of tree t and tree u, written as $t \circ u$, is defined iff the equation at the root node of u can be substituted in the equation at the root node of t (i.e. iff the lefthandside of the equation at the root node of u literally appears in the equation at the root node of t). If $t \circ u$ is defined, it yields a tree that expands the root nodes of copies of t and u to a new root node where the righthandside of the equation at the root node of u is substituted in the equation at the root node of t. Note that the substitution operation can be iteratively applied to a sequence of trees, with the convention that ∘ is left-associative.

We now have the basic ingredients for a DOP model of derivational explanation, which we may term "data-oriented physics". This DOP model employs (1) a *corpus of derivation trees* representing exemplars and (2) a *matching procedure* that combines subtrees from the corpus into new derivation trees. This brings us to the following definition for an explanation of a phenomenon with respect to a corpus.

Definition 1 Given a corpus C of derivation trees T_1, T_2,..,T_n representing exemplars and a term substitution operation ∘, an explanation of a phenomenon P with respect to C is a derivation tree T such that (1) there are subtrees t_1, t_2,..,t_k in T_1, T_2,..,T_n for which $t_1 \circ t_2 \circ \; ... \circ t_k = T$, (2) the root node of T is mathematically equivalent to P and (3) the leaf nodes of T are either laws or antecedent conditions or equations that cannot be derived from higher-level equations.

In our examples above, the mathematical derivation steps all occur at the end of a derivation (figures 1, 3 and 4). But they may of course just as well occur in the course of a derivation between two subtrees. We will come back to this in section 4, where we discuss an evaluation of our DOP model.

3 The Redundancy Problem

Now that we have extended the DOP model to derivational explanation, we can go into the main problem of this paper, and show how DOP may solve it. This is the problem that there can be many, often extremely many, different derivations for the same phenomenon, even if they are subsumed under the same general laws and even if they do not contain spurious laws that are non-explanatory or irrelevant. In the worst case, the number of derivation trees grows exponentially with the number of terms in the description of the phenomenon. In other words, derivational explanation is *massively redundant*.

In order to show this, we will first enlarge our tiny corpus used in section 2 (which consisted only of the derivation in figure 1) with another derivation from Alonso and Finn's textbook. This derivation again provides an exemplary problem solution for the Earth's mass but this time by computing it from the acceleration of an object at the Earth's surface (Alonso and Finn 1996, p. 246). This second exemplar can be represented by the derivation tree in figure 6.

By substituting the values for g (the acceleration at the Earth's surface), R (the Earth's radius) and G (the gravitational constant) in figure 6, Alonso and Finn obtain roughly the same value for the Earth's mass as in the previous derivation in figure 1. They argue that this agreement is "a proof of the consistency of the theory" (ibid. p. 247). (Note that the derivation is again idealized:

no centrifugal force is taken into account, let alone influences from the Sun or other planets.) Thus the simple problem of the Earth's mass is derivationally redundant in that it can be solved in at least two different ways. And *both* derivations are used in Alonso and Finn's textbook as exemplars for deriving other phenomena.

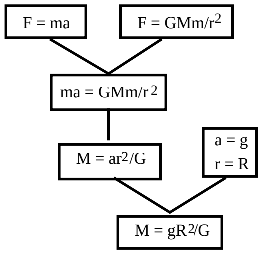

Figure 6: A additional exemplar in the corpus

When we add the tree in figure 6 to our corpus, we can model Kepler's regularity also on this exemplar, resulting in an alternative derivation tree which is constructed by using a large subtree from figure 6 in combination with two small subtrees from the exemplar in figure 1, the result of which is shown in figure 7. (And it easy to see that there are many more trees: by combining subtrees from the exemplars in figures 1 and 6 in different ways, we get an explosion of different derivation trees for Kepler's regularity.)

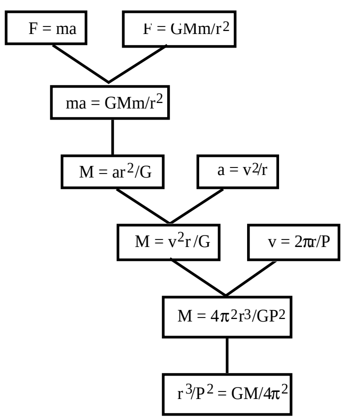

Figure 7: An alternative tree for Kepler's regularity

There is nothing wrong with this alternative derivation tree: there are no spurious non-explanatory laws that are irrelevant (as would be e.g. Hooke's or Boyle's law). The main difference is that the derivation in figure 7 is modeled on a *different* exemplar than the derivation in figure 3. In fact, the alternative derivation in figure 7 is even insightful as it refers to the conceptual equivalence between terrestrial and celestial mechanics in Newtonian dynamics. The fact that Kepler's regularity can be derived from figure 6 suggests that if we bring a satellite down to the Earth's surface it still follows the same regularity.

Yet, it turns out that no physics student comes up with the derivation tree in figure 7. Why? Apart from the fact that the derivation tree in figure 3 is smaller, the tree in figure 3 is more "derivationally similar" to an exemplar in the corpus. That is, the tree in figure 3 can be constructed by just one large subtree from the corpus – i.e. from the exemplar in figure 1 – whereas the tree in figure 7 needs at least 3 subtrees to be constructed from the corpus – one from the exemplar in figure 6 and two from figure 1. Of course, for another phenomenon it may be the exemplar in figure 6 rather than in figure 1 which can derive the phenomenon in one go. For example, to derive a formula for the gravitational acceleration at the Earth’s surface we can use one large subtree from figure 6 and not from figure 1. Thus different problems may be modeled on different exemplars.

Thus in solving the problem of derivational redundancy, it seems that we need to determine on which exemplar a phenomenon can best be modeled. Note that Kitcher’s account of explanation does not help us here. According to Kitcher (1989, p. 432), “Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again”. But his “unificationist” account does not tell us whether humans model a phenomenon like the gravitational acceleration at a planet’s surface on the exemplar in figure 1 or on the exemplar in figure 6. This is what should perhaps be called “Kuhn’s problem”, i.e. the problem of how we know on which exemplar a phenomenon can be modeled. DOP’s answer is: the exemplar from which the largest possible subtree can be reused. This finally brings us to our notion of “shortest derivation” and to a solution of the problem of derivational redundancy.

Let’s get more concrete. We have seen that there can be different derivation trees for a phenomenon. A distinctive feature between different derivation trees is that *some trees are more similar to exemplars than others*. The larger the partial match between a derivation tree and an exemplar, in terms of their largest common subtree, the more "derivationally similar" they are. Since students learn physics not just by memorizing laws, but also by studying exemplary problem solutions, they try to derive a phenomenon by maximizing derivational similarity with previously derived phenomena, or equivalently, by *minimizing derivation length* where the length of a derivation is defined as the number of corpus-subtrees it consists of. We will refer to the derivation of minimal length as the "shortest derivation". Since subtrees in DOP can be of arbitrary size, *the shortest derivation corresponds to the derivation tree which consists of largest partial match(es) with previous derivation trees in the corpus*. This brings us to the following definition of the "best derivation tree" for a phenomenon derived by DOP:

Definition 2 Let $L(d)$ be the length of derivation d in terms of its number of subtrees, that is, if $d = t_1 \circ ... \circ t_k$ then $L(d) = k$. Let d_T be a derivation which results in tree T. Then the best tree, T_{best}, derived by DOP is the tree which is produced by a derivation of minimal length:

$$T_{best} = \operatorname*{argmin}_{T} L(d_T)$$

It is important to understand the difference between a tree produced by the smallest number of subtrees and an absolute smallest tree. While the tree in figure 3 is produced by the shortest possible combination of corpus-subtrees, it does not correspond to the smallest possible tree, i.e. the tree with the smallest possible number of nodes (or labels). There exists a smaller tree that simply applies all laws at once to arrive at the formula for Kepler's regularity. However, it turns out that no student constructs such minimal derivations, and we therefore believe that our notion of shortest derivation consisting of the smallest number of (corpus-)*subtrees* is more appropriate than a notion of shortest derivation defined as the smallest number of *nodes*. Only in case our notion of shortest derivation does not lead to a unique result, i.e. if a phenomenon can be derived by the same smallest number of subtrees, it seems reasonable to choose the tree with the fewest nodes from among the shortest derivations, reflecting a preference for economy if DOP does not break ties.

4 Evaluating DOP

How can we evaluate our DOP model? Since we propose DOP to be a cognitive model of human problem solving, it seems appropriate to evaluate the model against derivations constructed by humans. To this end, we developed a *test corpus* of manually solved problems by fourth-year physics students and a *training corpus* of exemplary problem solutions taken from textbooks. Next, we developed an implementation of DOP which computed T_{best} for each test problem by means of the subtrees from the training corpus. The performance of DOP on the test problems was compared with the derivations provided by the students.

Method and procedure

We paid 19 third-year physics students to solve 6 elementary problems from classical mechanics and 5 elementary problems from fluid mechanics. The students had previously followed courses in classical mechanics and fluid mechanics. The 11 problems given to them consisted in deriving a phenomenon from laws, initial conditions and, in the case of fluid mechanics, empirical coefficients. The students were given no other instructions than that they should solve the problems by paper and pencil in class. The first two and the last two of the problems are given below.

> *Problem nr. 1*
> Show that the period of the Earth's rotation for which an object at the equator would become weightless is given by $P = 2\pi\sqrt{(R/g)}$ where R is the Earth's radius and g is the gravitational acceleration at the Earth's surface.
>
> *Problem nr. 2*
> Show that the theoretical velocity which an object attains in free fall from height h is given by $v = \sqrt{(2gh)}$ where g is the gravitational acceleration at the Earth's surface.
>
> *Problem nr. 10*
> When water flows through a right-angled V-notch, show that the discharge is given by $Q = KH^{5/2}$ in which K is a constant and H is the height of the surface of the water above the bottom of the notch.
>
> *Problem nr. 11*
> Show that the theoretical rate of flow through a rectangular notch is given by $Q = (2/3)B\sqrt{(2g)}H^{3/2}$ where B is the width of the notch and H is the height of the water level above the bottom of the notch.

After the students had solved the problems on paper, they were given a short, fifteen-minutes tutorial on the concept of derivation tree, especially on the difference between binary branches in a tree (used for combining laws, conditions etc.), and unary branches (used for mathematical derivation steps). The students were told that the exact order of combining laws in a tree was not important as long as these laws could be properly combined by term substitution to solve the problem. What *was* important was to represent in the tree the derivation steps they had used to get from laws to the phenomenon, so we told them. Thus we did not distinguish between trees whose only difference was the order of the applied laws, as we found out in a pilot experiment that this was neither done by the students. We will see below that even with this abstraction there were still many different derivations. After this brief tutorial, the students were asked to draw derivation trees for their problem solutions.

There was a high agreement among the derivation trees constructed by the students: on average 91.4% (SD=1.2) of the derivation trees per problem matched (modulo law order). Only the most frequent derivation tree for each problem was put in the test corpus. It is important to mention that the students had no difficulties with drawing trees for their problem solutions, and there were no questions during this task. This suggests that derivation trees are suitable structures for representing problem solutions by humans.

Next, the students were asked to draw derivation trees for 9 exemplary problem solutions that are used as exemplars in the textbooks by Alonso and Finn (1996, chapter 11) and Douglas and Matthews (1996, chapter 7). The three example problems in figures 1, 5 and 6 were among these exemplars. It should be said, however, that none of the students derived $F=mv^2/r$ from $F=ma$ and $a=v^2/r$, as we did in figure 1. Instead, all students used the equation for centripetal force $F=mv^2/r$ directly as a law. The agreement among the derivation trees for the exemplary solutions was very high: 97.7% (SD=0.3). The most frequent tree for each exemplary solution was put in the training corpus.

All test problems could be solved out of subtrees from the training corpus, but this fact was not told to the students: they first had to solve the test problems after which they were asked to draw trees for the exemplary problem solutions from the textbooks. Each student accomplished the task in less than 2.5 hours (including the tutorial).

The goal for DOP was to construct T_{best} for each of the 11 problems from the test corpus by means of the subtrees from the training corpus of 9 exemplars. To accomplish this, we implemented DOP by using *TK Solver* as a backbone (release 5.0, Universal Technical Systems Inc.). *TK* solves an equation given a list of other equations -- provided that there is a solution. To make *TK* suitable for DOP, we programmed a shell around *TK* (written in *C*) which first converted each derivation tree from the training corpus into all its subtrees and next extracted the equations from the subtree-roots. Each equation was indexed to remember the subtree it was extracted from. This resulted in a list L of 148 equations. For each test problem, *TK* derived a set of solutions given the list L. All problems received more than 60 different solutions, even *after* abstracting from the order of the equations used in the solution, which gives an idea of the derivational redundancy if we do not have any mechanism to break ties.

From *TK*'s output, our program selected the shortest solution for each problem that used the fewest equations.

The equations of the shortest solution were converted back to their corresponding subtrees, which were combined into the derivation tree T_{best}. Note that in this way T_{best} consisted of the largest partial matches with trees in the training corpus. In case T_{best} was not unique the program chose the tree with fewest nodes among the best trees. A major advantage of *TK* is that it hides the algebra, which is good as this was also asked from the students and which corresponds to our use of unary branches in trees.

Results

The best trees computed by DOP were compared with the derivation trees constructed by the students in the test corpus. Abstracting from the order of the laws in the trees, the accuracy of DOP was 91%. That is, for 10 out of 11 phenomena, the derivation trees predicted by DOP matched (modulo law order) the derivation trees assigned by (the vast majority of) the students.

To put our 91% accuracy into perspective, we also computed the accuracy by choosing a *random* derivation tree, T_{random}, from among *all* possible trees that were constructed by DOP (i.e. trees that did not necessarily correspond to the shortest derivation but that could be constructed from *TK*'s output). In this case, the accuracy was only 9% (again, modulo law order). Although our test set consists of only 11 trees, the difference between 91% accuracy obtained by T_{best} and the 9% accuracy obtained by T_{random} -- which is however still a "correct" derivation -- suggests that T_{best} mimics human problem solving more closely than T_{random}. We also computed the accuracy by choosing the *absolute* smallest tree containing fewest nodes among all proposed trees in *TK*'s output. This resulted in 18% accuracy (2 out of 11).

These results suggest that if we want to predict the derivations that humans construct for phenomena, we should not search for the smallest tree in terms of nodes, let alone take a random tree, but search for the tree which consists of the fewest subtrees from previous derivations, as generated by our notion of shortest derivation. Further research, with larger corpora and more complex problems, will be needed to support these results.

One may claim that our results are not very surprising since they reconfirm Kuhn's insight that scientists explain phenomena by modeling them on previously explained phenomena with only minimal recourse to additional derivation steps (Kuhn 1970, p. 189). However, Kuhn does not provide an exact procedure that tells us on which exemplar a new phenomenon can be modeled. DOP does provide such a procedure, albeit indirectly by solving the problem of derivational redundancy. In doing so, DOP also suggests a precise notion of similarity between a phenomenon and a set of exemplars: the most similar exemplar is the one from which the largest subtree can be reused to derive the phenomenon. Moreover, even when we know on which exemplar a phenomenon can be modeled, there may *still* be several different derivations for the phenomenon, also if we abstract from the order of law application. Thus we need an additional notion to break ties, as given by DOP's shortest derivation.

5 Conclusion

This paper proposed a solution to the problem of derivational redundancy in physics. We showed that the DOP model provides a way to reduce the combinatorial explosion of different derivations of a phenomenon by selecting the shortest derivation consisting of the fewest partial derivations of previously derived phenomena. A preliminary investigation with a corpus of phenomena from classical and fluid mechanics showed that DOP accurately predicts the derivations humans come up with. To the best of our knowledge, this paper provides the first concrete proposal to tackle the problem of massive derivational redundancy in scientific explanation.

Acknowledgments

This paper benefitted from discussions with many people. Special thanks go to Margaret Morrison whose extensive comments were particularly helpful.

References

Alonso, M. & E. Finn (1996). *Physics*, Addison-Wesley.

Bod, R. (1998). *Beyond Grammar*. CSLI Publications, Stanford.

Bod, R. (2000). Parsing with the Shortest Derivation. *Proceedings COLING'2000*, Saarbruecken, Germany, 2000.

Bod, R., R. Scha & K. Sima'an (2003). *Data-Oriented Parsing*. University of Chicago Press.

Carbonell, J. (1986). Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition. In Michalski et al. (eds.), *Machine Learning*, Vol. II, Morgan Kaufmann.

Collins M. & N. Duffy (2002). New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete Structures, and the Voted Perceptron. *Proceedings ACL'2002*, Philadelphia, PA.

Douglas, J. & R. Matthews (1996). *Fluid Mechanics*. Vol. 1, 3rd edition, Longman.

Falkenhainer, B., K. Forbus and D. Gentner (1989). The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligenc*, 41, 1-63.

Friedman, M. (1974), Explanation and scientific understanding, *Journal of Philosophy*, 71, 5-19.

Hempel, C. & P. Oppenheim (1948). Studies in the Logic of Explanation, *Philosophy of Science*, 15, 135-175.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher and W. Salmon (eds.), *Scientific Explanation*, University of Minnesota Press.

Kolodner, Janet (1993), *Case-Based Reasoning*, Morgan Kaufmann.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*, 2nd edition, University of Chicago Press.

Manning, C. & H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press.

Nickles, T. (2003). Normal science: from logic to case-based and model-based reasoning, In Thomas Nickles (ed.), *Thomas Kuhn*, Cambridge University Press.

Scha, R., R. Bod & K. Sima'an (1999). A Memory-Based Model of Syntactic Analysis: Data-Oriented Parsing. *Journal of Experimental and Theoretical Artificial Intelligence*, 11(3), 409-440.

VanLehn, K. (1998). Analogy Events: How Examples are Used During Problem Solving, *Cognitive Science*, 22(3), 347-388.

Veloso, M. & J. Carbonell (1993). Derivational Analogy in PRODIGY: Automating Case Acquisition, Storage, and Utilization. *Machine Learning*, 10(3), 249-278.