An Integrated View of Structural Model Driven Scientific Discovery

Adrian Gordon (ADRIAN.GORDON@UNN.AC.UK)
Department of Computing
University of Northumbria at Newcastle
Newcastle-Upon-Tyne NE18ST UK

Much of chemistry has been concerned with determining the fundamental chemical entities which exist in nature, the structural relationships between them, the properties of these chemical entities, and the mechanisms for combining and breaking them apart.

Previous work in computational scientific discovery has addressed how each of these types of chemical knowledge can be used to elucidate the others. For example, the STAHL/DALTON systems (Langley, Simon, Bradshaw & Zytkow, 1987) use knowledge about the process of chemical reaction to determine the components of a substance (STAHL), and its molecular composition (DALTON). Our own work has looked at how knowledge of structure has been used by scientists to explain the properties of substances. We have shown how a search space of structural models can be defined which accounts for many of the historical discovery episodes in the development of solution chemistry (Gordon, Edwards, Sleeman & Kodratoff, 1994). We have also shown how the HUME discovery system (Gordon, 1995) can use individual models from this search space to guide the discovery of empirical laws which can then explain the behaviour of solutions (explaining their freezing points, for example).

In our earlier work, we assumed that a search space of structural models existed in a domain *a-priori*. Subsequently, we showed how such a search space could be generated systematically, by the application of a set of model generation operators. These operators could be applied to one structural model to generate a set of new models. These model generation operators have much in common with the knowledge that is output by systems such as DALTON. However, whereas DALTON generated instances of model construction operators based on a single process (the *chemical reaction* process) our analysis of solution chemistry history has shown how many other processes were employed by scientists to account for the behaviour of solutions. These include *hydration*, *polymerisation*, and *ionisation*. As with the chemical reaction process, each of these other processes has its own particular characteristics. Like chemical reaction, hydration tends to involve small whole number ratios of the number of molecules involved in the hydrate. As well as requiring the conservation of mass, ionisation also requires the conservation of electric charge. Ionisation also requires the presence of an electric current as a precondition for its application.

Figure 1 below shows how we intend to generalise our model of structural-model driven scientific discovery to allow the integration of knowledge of structural chemistry from different sources. Figure 1 shows how a generalisation of the DALTON system (termed DALTON_X) could empirically discover instances of model generation operators. Using appropriate heuristics, DALTON_X could discover instances of hydration, polymerisation and ionisation operators, as well as those for chemical reaction. These operators could be systematically applied to generate a search space of hypothetical structural models, and each of these structural models could be used to guide the discovery of physical laws. The output from the integrated system would be a general theory of (for example) solution chemistry, consisting of a set of empirical laws and confirmed structural models, which could together explain the phenomenon under investigation.

Figure 1 seems to imply that a set of model generation operators can be directly applied to a new domain. This is not the case. Two of the most significant advances in the history of solution chemistry came about when scientists *relaxed the preconditions for operator application*. It was Rüdorff who first proposed that the hydration process could occur *inside a solution*. Similarly, Arrhenius first proposed that the ionisation process did not require the presence of an external current to "tear apart" molecules into their constituent ions.

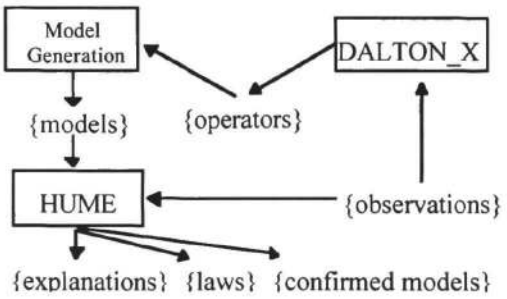

Figure 1. An Architecture for Integrated Model-Driven Discovery

References

Gordon, A. (1995). The HUME Model-Driven Discovery System. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society.*

Gordon, A., Edwards, P., Sleeman, D. & Kodratoff, Y. (1994). Scientific Discovery in a Space of Structural Models: An Example from the History of Solution Chemistry. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* (pp. 381-386).

Langley, P., Simon, H.A., Bradshaw, G.L., and Zytkow, J.M. (1987). Scientific Discovery: Computational Explorations of the Creative Processes. Cambridge, MA: MIT Press.