Summary - Dendral: a case study of the first expert system for scientific hypothesis formation by Robert K. Lindsay et al.

25 September 2019 - Paul Meierling (6.844)

Overall summary of the paper
Dendral was one of the first heuristic programming tools that could actually be used in a field of empirical science (chemistry). It used mass spectrometry readings to infer, which potential molecules could create those readings. In theory this would greatly alleviate the workload from scientist. It worked by combining a rule-based system with a rule-generation system as well as a plan-generate-test procedure. The novel aspect of this procedure was the ability to limit the search space of the generated potential solutions by using expert knowledge about the subject (knowledge engineering).

Architecture of the Dendral system
The Dendral system did consist of two major subsystems: Heuristic Dendral and Meta Dendral.
Heuristic Dendral was a performance system that was given a set of specific rules enabling it to generate a list of potential molecules based on mass spectrometry data. The novel approach of Heuristic Dendral in comparison to existing generate-test systems was that by using expert knowledge the system was able to greatly reduce the potential search/test space. The so called plan-generate test approach worked as follows: the generator (GENOA) created a list of all potential molecules based on chemical rules. However, since the number would be too large for bigger molecules a planner module was able to limit those solutions by using expert knowledge like for example a user provided list of atoms to include or avoid (GOODLIST and BADLIST). After pruning the search-tree the tester (PREDICTOR) was then able to determine whether which of the generated molecules were feasible.
Meta Dendral was a learning system that when given a set of possible structures and their respective mass spectrometry readings was able to generate rules for their relationship. These rules could then be used by Heuristic Dendral to better test the output of the generated solutions

Results and learnings
It took 15 years of research to create the functional Dendral program, which was then trying to be commercialised by the Stanford marketing department. However, Dendral did not achieve commercial success and was mostly unknown to any chemist or laboratories who could benefit from it. While several reasons contributed to the lack of success the key problem was that Dendral worked well with a small amount of potential atoms (8-10) per molecule. However, for these numbers the scientist could also easily determine the right molecule themselves and did not need an AI system. Nevertheless, the system made critical discoveries in the field of AI and develop key learnings such as: The efficiency of generating potential solutions is key. Spending a small amount of time at the beginning to later reduce the complexity is worth it. Planning is essential for complex problems.

Discussion
The importance of the Dendral system to development of modern AI is hard is easy to see. In my opinion the interaction of Meta Dendral, with the Heuristic system, while not the key part of the paper, is actually the most interesting. Current systems (e.g. CVNNs) in a way also create rules based on input data and labeled output data. However, since the generated rules are very opaque (e.g. trained weights in a CVN) it is hard to combine them with expert knowledge. Maybe a combination of a system that is able to interpret rules, which are generated by another system as well as rules written down by experts will enable us to develop smarter systems in the future.