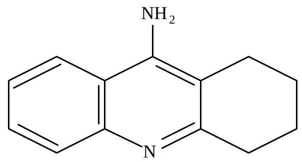
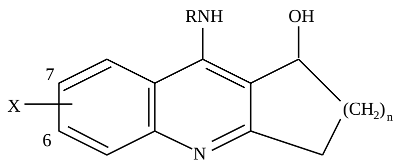
Learning rules for comparing drug activities

The majority of pharmaceutical R&D is based on finding slightly improved variants of patented active drugs (292 out of 348 US drugs introduced between 1981 and 1988 were of this kind). This involves laboratories of chemists synthesising and testing hundreds of compounds. The ability to automatically discover the chemical properties which affect the activity of drugs could provide a great reduction in pharmaceutical R&D costs. The average cost of developing a single new drug is $230 million.

Alzheimer's disease has long been recognised as one of the most disabling conditions affecting the aged. It is the major cause of dementia among elderly people. The drug Tacrine has drawn considerable attention because of a report of its clinical efficacy for this problem. This is structure of Tacrine



Large-scale clinical trials however show that the primary disadvantage of Tacrine is its high toxicity. This has led to investigations into possible variants of the basic Tacrine structure. Variations reported used the following structure



New compounds are created by substituting chemicals for "R" and "X". The substitution for "X" can be in position 6 or 7. The literature reports on experiments with a number of drugs. The experiments were concerned with testing various properties of the drugs. A good drug should have the following properties:

- Low toxicity.
- High acetocholinesterase (!) inhibition.
- Good reversal of scopolamine induced deficiency.
- Inhibit amine re-uptake.

Data from the chemical literature [3] was converted into a representation suitable for ILP. More details on the ILP experiments are available in [1]. The literature does not indicate any particular pattern followed in creating these new drugs. The plan is to learn rules that tell us what makes a drug have a desirable property (for example, low toxicity). Hopefully, these rules will offer better guidance for the synthesis of a new drug.

Rules are to be learnt for each property using the following information:

- The positive and negative examples are pairwise comparisons of drugs. There are different sets of examples for each property.
- Physical and chemical properties of chemicals that can be substituted for "R" or "X" are described by unary predicates in the background knowledge. These properties include hydrophobicity, hydrophilicity, charge, size, polarity, whether a chemical is aliphatic or aromatic, whether it is a hydrogen donator or acceptor etc.
- Sizes, hydrophobicities, polarities etc., are represented by constants (for example, polar0, polar1 etc.).
- Relations between the constants (less_than(polar0,polar1)) is also provided as background knowledge.
- **x_subst(Drug,Position,Subs)**. The "X" substitution for Drug in position Position is Subs.
- The "R" substitution is split into parts corresponding to bonds between different chemical structures.
- Sometimes the "R" substitution involves the presence of 1 or more ring structures. These are benzyl derivatives that arise from substitutions within the basic benzene ring structure. The number of such substitutions in the basic benzene structure, the position of the substitutions an the actual substituent are provided.

The data files are stored in a compressed TAR file, which will detar into several directories. The files themselves are as used in the original *Golem* [2] experiments. That is, background knowledge files have a ".b" suffix, positive example files have a ".f" suffix, and negative example files have a ".n" suffix.

References

[1] King, R.D., Srinivasan, A. and Sternberg, M.J.E. (1995). Relating chemical activity to structure: an examination of ILP successes. *New Gen. Comput.* (to appear).

[2] Muggleton S.H. and Feng C. (1990). Efficient induction of logic programs. *Proceedings of the First Conference on Algorithmic Learning Theory*, Tokyo.

[3] Shutske G.M., Pierrat F.A., Kapples K.J., Cornfeldt M.L., Szewczak M.R., Huger F.P., Bores G.M., Haroutunian V. and Davis K.L. (1989). 9-Amino-1,2,3,4-tetrahydroacridin-1-ols: Synthesis and Evaluation as Potential Alzheimer's Disease Therapeutics. *J. Med. Chem.*, 32:1805–1813.