Command line completion: learning and decision making using the imprecise Dirichlet model

Erik Quaeghebeur
SYSTeMS research group
Universiteit Gent
Technologiepark 914, 9052 Zwijnaarde
Belgium
Email: Erik.Quaeghebeur@UGent.be

Gert de Cooman
SYSTeMS research group
Universiteit Gent
Technologiepark 914, 9052 Zwijnaarde
Belgium
Email: Gert.deCooman@UGent.be

1 Introduction

Our purpose is to to illustrate how models using *imprecise probabilities* [1] can be used in a practical application such as command line completion.

The central model is the imprecise Dirichlet model (IDM) [2]. It is a model for representing probabilistic information about multinomial processes, combined with an updating scheme. The updating scheme allows for learning from observations. An IDM, together with utility specifications for completion actions, can be used in decision making.

Our research, using the IDM for learning in Markov models, is incorporated in the current illustration to allow the completion method to take preceding commands into account.

2 Concepts and models

Command line completion is a feature that operating systems (OS's) provide for users working in text-mode. The user invokes the feature by pressing a predefined key after having typed part of a command. The user is then either presented with a list of n possible completions i, or, if there is only one, this completion is given on the command line.

Our objective is to add supplementary functionality to command line completion when multiple completions are possible. By modelling (making assumptions about) the behavior of the user, the OS can order the completion actions.

We assume that the user chooses an i according to an unknown, but fixed, multinomial distribution $\boldsymbol{c}$ over all possible completions. A component c_i gives the chance that the user chooses i. The set of all possible distributions $\boldsymbol{c}$ is C_n.

Practically, we only have partial, probabilistic, information about $\boldsymbol{c}$. The *imprecise Dirichlet model*, used by the OS to represent this probabilistic information, consists of a convex set of Dirichlet distributions over C_n. The expectation of $\boldsymbol{c}$ under this model consists of a subset M_t of C_n.

Every time the user enters a command, the OS updates the IDM. This results in a new IDM with a more precise expectation M_{t+1}. We take M_0 to be the whole interior of C_n.

We can make the assumptions about the user's behavior more complex by allowing the multinomial distribution $\boldsymbol{c}$ over all possible completions to depend on the command typed on the preceding command line. This behavior corresponds to a *Markov model* with an unknown transition matrix T. Each row of T will correspond to a multinomial distribution, conditional on the preceding command.

The OS then uses a modified version of the IDM, consisting of one set of Dirichlet distributions per row, to represent the probabilistic information available about T. This IDM can be updated after observing two successive commands.

3 Actions, utility and decision making

Completion actions can consist of actions a_i (presenting completion i on the command line) and a default action (giving a specially ordered list of n completions).

Whatever action the OS takes, there is only one completion that the user is looking for. Each action the OS takes thus has a certain *utility* for the user. For example, action a_i gives a utility of $+1$ if i is the correct completion and -1 if it isn't.

The OS now has a model both for what completion the user is looking for, the IDM, and for the utility of his actions. By combining these two models, the OS can calculate the *expected utility* for each of the completion actions.

The expected utility for the a_i's can then be used for *decision making*, i.e., a partial ordering of these actions can be constructed. If this ordering has one maximal element, the OS decides to present the corresponding completion. If there are multiple maximal elements, the default action is taken: showing the partial ordering to the user, such that the maximal elements are the most conspicuous ones.

References

[1] Peter Walley, *Statistical reasoning with imprecise probabilities*, Chapman and Hall, London, 1991.

[2] Peter Walley, "Inferences from multinomial data: learning about a bag of marbles," *Journal of the Royal Statistical Society* B, vol. 58, no. 1, pp. 3–57, 1996.