



Doctoral thesis review

The doctoral thesis of Alan Eckhardt deals with the issue of preference learning. This is a hot topic in web mining applied to building recommender systems. The thesis is structured into 8 chapters and two appendices. The main methodological contributions are in chapters 3 and 4. The basic idea is to perform preference learning in two steps, at first, the preferences are learned for each attribute describing an object separately, then the aggregated preference value is computed. The thesis shows various ways how to perform both steps. The proposed method can be used for a recommender system in an arbitrary e-shop. The method is thoroughly tested and compared with other techniques; the framework used for testing is described in chapter 6 and the results of testing are given in chapter 7 and in appendix A.

Although the thesis consists of 154 pages, some subchapters are too brief and do not describe sufficiently what is in the headings (subchapter 2.3.1 on content based models, subchapter 2.4.8 on ILP). There are also some missing or unclear points:

- chapter 6 introduces a framework for running experiments with preference learning. The required data format (as shown in 6.3) contains userID, objectID, rating, and the attributes describing the objects, but chapter 7, that reports the experiments, doesn't mention the question of different users,
- chapter 7 describes several experiments using machine learning algorithms implemented in Weka (SVM, MLP, JRip and PART), but different experiments have been carried out with different subsets of these algorithms. Subchapter 7.9.3 presents figures for SVM, MLP, sometimes also for PART, subchapter 7.10 uses SVM, JRip, PART. No explanation is given why

There are also some mistakes in the thesis (C4.5 is a tree learning, not rule learning algorithm and its implementation in Weka is J48, formula at page 20 is weighted sum not average).

I'd like to know answers to following questions during the defense:

1. Subchapter 2.4.6 describes only some general purpose data mining tools (e.g. Weka or Clementine), are there any suitable web mining systems that can be used/customized for the preference learning task?
2. The selection of algorithms from Weka seems to be arbitrary, was there any special reason why to choose MLP, SVM, JRip and PART and to omit e.g. Bayesian or nearest neighbor methods?

Despite of my critical comments given above, I'd like to express my opinion that the thesis of Alan Eckhardt contains his original, previously published research results and thus fulfill all requirements to award him the doctoral degree (and title PhD).

Prague, June 25

Prof. Ing. Petr Berka, CSc.