Chapter 8
Conclusions, Discussion and Future Work

8.1 Summary of Contributions

In this monograph, we detailed the development of our Pocket Data Mining (PDM) framework. PDM is set to serve the next generation of applications in predictive analytics targeting users of smart handheld devices. We can summarize the presented contribution in this book in the following:

- We presented in Chapter 2, a detailed literature review of developments and research projects related to PDM.
- In the core chapter (Chapter 3), PDM has been presented giving details of the system architecture and a typical data stream mining process utilizing PDM. The chapter also detailed a thorough experimental study evidencing the feasibility of the framework.
- Chapter 4 is a point of reference to practitioners on how PDM can be ported from the desktop environment to run on-board Android smartphones.
- Chapters 5 and 6 present an important extension to PDM to deal with the concept drift issue, which is tightly coupled with change in the user’s context.
- Chapter 7 highlights the great potential of PDM in a variety of applications in different domains.

With all research projects, there is always room for improvement and extensions. In the following section, current and future work is investigated.

8.2 Ongoing and Future Work

This section provides the reader to pointers to current and future development related to the PDM framework.
8.2.1 Rating System for AMs

The current implementation of the MADM agent assumes that the local AMs are of good quality, and thus in the case of classification of unlabeled data instances, it is assumed that the weights are calculated correctly and truly reflect the AMs classification accuracy. This assumption may be true for the AMs we developed in-house, which we used for the evaluation, but third party implementations may not be trusted. For this reason, a rating system about AMs is currently being developed based on historical consultations of AMs by the MADM. For example, if the MADM remembers the classifications and weights obtained from AMs visited and the true classification of the previously unknown instances is revealed, then the MADM could implement its own rating system and rate how reliable an AM’s weight was in the past. If an AM is rated as unreliable, then the MADM may even further lower its weight. However, it is essential that this rating system is also able to loosen given ratings, as the AM’s performance might well change if there is a concept drift in the data stream. In order to detect such concept drifts, it is necessary that AMs that have a bad rating are still taken into consideration, even if it is with a low impact due to bad ratings.

We have just outlined a possible rating system for classification AMs. However, rating systems for other less generic PDM agents such as the GDF and LFA agents outlined in the applications of PDM in Chapter 7 remains an open area to be explored.

8.2.2 Intelligent Schedule for MADMs

In its current implementation, the MADM visits all available AMs, however, this may be impracticable if the number of AMs is very large. Currently, a mechanism is being developed for MADMs according to which the MADM can decide when to stop consulting further AMs. A possible stopping criteria could be that a certain time has elapsed or the classification result is reliable enough. Also the rating system outlined above can be used to determine an order in which AMs are visited. If there are time constraints the MADM may prioritise more reliable AMs.

8.2.3 PDM Beyond Data Stream Classification

PDM is a new niche of distributed data mining. The current implementation of PDM focuses on classification techniques, however, there exist many more data mining technologies tailored for data streams and mobile devices. For example, there are stream mining techniques that classify unlabelled data streams [40, 75] which could be introduced into PDM.