Parts of a research paper
CSCI 4380/6380

Below are sections that often appear in scientific research papers. In actual papers, there will be considerable variation, but the below items can serve as a template.

When you write, you should have a particular audience in mind. In this case, you should assume that the audience is a researcher familiar with the basics of machine learning and data mining. What this means is that you do not need to explain every detail of the algorithm or technique used. However, if you discuss decisions you made in setting up your experiments, then it might make sense to give a brief overview of the techniques used.

1) Abstract (1-2 paragraphs)

The purpose of an abstract is to summarize the entirety of the paper in a few short paragraphs. The problem and the results should both be described. The reader should be able to tell, just by reading the abstract, whether the rest of the paper contains material the reader is interested in.

2) Introduction (1-2 pages)

The introduction should

1) provide a motivation for the project as well as background information for it;
2) describe the specific objectives of the project and indicate how they are to be met; and
3) outline the remainder of the paper

The specific problem to be solved or research question to be answered is usually specified in the introduction. Be as clear as possible. Without clarity in the introduction, it's difficult to assign meaning to anything that comes later in the paper. That is, the reader won't be able to judge the rest of the work, because he or she won't know what the work is supposed to achieve.

3) Literature review (2-3 pages)

This should detail other research that is related to the current work. If other researchers have attacked the same problem or a similar problem, or if other researchers have used the same data set, then their work and results should be briefly described. Proper citations should be made.

You should identify a handful of relevant conference or journal papers and describe

- what the authors were working on;
- their results;
- how their work is relevant to your current work, and also how the current work is different (that is, why is the current work needed?).

Ideally, you should identify many papers (e.g., 10, 15) that are relevant. You need not understand every detail of them, but you should be able to talk reasonably about what the researchers were doing and how it relates to your work. You might be able to focus on a smaller number (e.g., 5) and describe them in more detail than the others. For some, only a few sentences are needed.

If you are working in teams, it might be efficient to divide the literature review into pieces, each member working on a specific piece.

4) Methods (3-4)

This section should present the data sets and machine learning strategies used, and describe the experiments that were performed. The source of the data as well as any interesting characteristics of it should be described. If multiple data sets are combined or if the data is transformed or pre-processed, then the processes should be described.

Regarding the machine learning strategies, their configuration should be described in sufficient detail to allow another researcher to reproduce the results. The computers used to perform the experiments should also be described. It is common to give a brief summary of how each ML scheme works.

Some questions to think about:

1) Where does the data used come from? Who generated it and for what purpose?
2) Are there missing values? Does it have other interesting characteristics?
3) What techniques were used to transform it (and why)? Are these described in sufficient detail that others could achieve the same results?
4) Why were the particular ML frameworks used chosen? Are they particularly suited to the problem?
5) When running the experiments, were any problems encountered?

Results and Analysis (3-4)

In this section, you present the results of your experiments, being as clear as possible, and you provide an analysis and discussion of them. It is recommended that you present the factual results of your work in a single location as a coherent unit and then, separately, provide an analysis of them. Doing this makes it easier for the reader to identify the key points of your work.

Do not simply record the output of whatever ML framework you used. You should attempt to interpret the results. Are they significant? Is it possible that the data, or the way it was partitioned, or the ML schemes used, or their configuration, resulted in results that are misleading?

Conclusion and future work (<1 page)

Here, you can recapitulate, perhaps using only a few sentences, the main results of the paper. You can also discuss ways in which the current work can be extended.