Ensemble Learning Models for CTR Prediction

Sampoorna Biswas
Department of Computer Science
The University of British Columbia
sbiswas@cs.ubc.ca

Sam Creed
Department of Computer Science
The University of British Columbia
samcreed@cs.ubc.ca

Abstract

The abstract paragraph should be indented 1/2 inch (3 picas) on both left and right-hand margins. Use 10 point type, with a vertical spacing of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

In this section we will

- state the problem being addressed (CTR prediction)
- define CTR prediction
- explain what a kaggle competition is and how it is relevant to this project
- motivate why this is an important problem
- highlight challenges of limiting approaches
- summarize contribution of work/why and what we want to do

2 Related Work

In this section we will

- identify existing research in this area (at least three sources)
- discuss how our approach differs from existing research or is incomplete (perhaps does not use ensemble methods, or not for this particular application)
-

3 Ensemble Learning

In this section we will give an overview of the different learning stubs used in our ensemble learning algorithm.

3.1 Ensemble Learning

We should also perhaps explain what ensemble learning is.

3.1.1 Stub 1: SVM?

Description.

3.1.2 Stub 2: Naive Bayes?

Description.

3.1.3 Stub 3: Random Forest?

Description.

3.1.4 Stub 4: Other?

Description.

3.1.5 Google Fast Solution Approach

Description. Discuss the fast python solution, maybe use as a sub solver, and compare our implementation against it.

3.2 Dataset and Feature Selection

Description. For kaggle competition, discuss train/test data given.

Discuss analysis of features and how we attempt to limit or ignore certain features in the learning process. Maybe something else to compare against with time/quality performance.

4 Experimental Results

In this section we should talk about

- the kaggle competition and the train and test datasets
- comparison in quality of ensemble, fast solution, stubs, and using feature selection
-

5 Discussion and Future Work

State main conclusion obtained from the course project. List one strength and one weakness of our contribution. State what we would do with more time.

References

References follow the acknowledgments. Use unnumbered third level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to 'small' (9-point) when listing the references. **Remember that this year you can use a ninth page as long as it contains** *only* **cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D. S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609-616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer-Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.