**Introduction of FTRL – Proximal Online Learning Algorithm**

FTRL is an algorithm proposed in “[Ad Click Prediction: a View from the Trenches](https://static.googleusercontent.com/media/research.google.com/en/pubs/archive/41159.pdf)” paper by Google Research team. This paper addresses some challenges in real-world CTR system such as:

* Memory Saving on output model size.
* Online Learning on large and sparse data.

Normally, sample in FTRL has billion features, but only hundreds of them are non-zero. FTRL uses a Generalized Linear Model (e.g. Logistic Regression in implementation) and train its weights on non-zero features by streaming sample from disk or network. This allows Online Learning on large scaled possible with less demand computation. We can easily add regularization in the implementation to control the complexity of output model and reduce model size as well. Using Logistic Regression in implementation, on high level, training and inference steps in FTRL is similar as Logistic Regression.

* 1. Given sample with feature vector x, model will predict output by:

with is Logit function.

* 1. Calculate Loss Function w.r.t given sample:

,

with y is observed label (0 or 1) and p is prediction in above step.

* 1. Then Gradient of Loss function is used to update model weight to new value:

With

However, FRTL used some tricks in step 3 to:

* Replace very small weight by zero.
* Gradient descent update to use very small memory consumption.

For more mathematical details, please see section “3. ONLINE LEARNING AND SPARSITY” in paper.

In current implementation, we do not One-Hot-Encode input data. Just keep input data after Hashing then model can learn without any problem. This is also another nice property of FRTL in engineering view.

For example, input data after Hashing is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | site\_id (si) | site\_domain (sd) | site\_cat (sc) | Click |
|  | 1 | 4 | 7 | 1 |
|  | 1 | 4 | 8 | 1 |
|  | 2 | 5 | 9 | 0 |
|  | 2 | 6 | 7 | 0 |