



9

More ▾ Next Blog»

Create Blog Sign In

Pragmatic Programming Techniques

Saturday, January 14, 2012

Machine Learning: Ensemble Methods

Ensemble Method is a popular approach in Machine Learning based on the idea of combining multiple models. For example, by mixing different machine learning algorithms (e.g. SVM, Logistic regression, Bayesian network), ensemble method can automatically pick the best algorithmic model that fits the data the best. On the other hand, by mixing different parameter set of the same algorithmic model (e.g. Random forest, Boosting tree), it can pick the best set of parameters of the same algorithmic model.

Bagging

Bagging is based on the idea of learning multiple models using different sets of sample data, basically by random sampling (with replacement) from the same training set. After the models are learned, we use a voting scheme to predict future data. In case of the classification problem, we use the majority class voted by all models. In case of the regression problem, we take the average value of estimated output from all models.

The model doesn't have to be equally weighted. We can use a weighted average of individual models to come up with the final model. $m = w_1.m_1 + w_2.m_2 + w_3.m_3 + \dots$

To obtain the weights $w_1, w_2 \dots$ we can use machine learning to figure out. In the case, we first use the training data set to train individual models. After that, we use the validation data set to train the weights. Concretely, we feed each data point from the validation set to each model to come up with the final prediction. Based on how the ensembled prediction deviates from the actual outcome, we can learn the optimal set of weights.

Boosting

Boosting extend the idea of bagging by putting more emphasis on the training data that is wrongly predicted. In boosting, weight sampling is used. Initially each training data is equally weighted but as each iteration goes, the data that is wrongly classified will have its weight increased. On the other hand, the model also has its weighted according to how good this model is predicting the data in its round.

1. Each training data carry a sample weight (initially are the same)
2. For each iteration, take sample (with replacement) based on weights
3. Compute the error rate e
4. For each sample that is wrongly predicted, adjust its sample weight by $e/(1-e)$
5. Weight the trained model according to its error. For example, $\log(e/(1-e))$
6. Stop when e is small enough
7. Ensemble the overall model according to model weight.

Gradient Boosting Method is one of the most powerful and popular boosting methods. It is based on incrementally add a function that fits the residuals.

Found function F to fit $y \sim F(x)$ using a gradient descent approach.

Start with some random guessing function F . Compute the loss based on the residual.
loss = $L(y - F(x))$ where L is some loss function

Partial differentiate loss w.r.t. function F , and compute the value at every data point. Use another machine learning model to learn another function $g(x)$ that predicts the partial differentiation value. Then update $F(x) \leftarrow F(x) + a.g(x)$ where a is the learning rate.

Other ways of Sampling

Instead of sampling the training data, we can also sample on the attributes (e.g. we can randomly pick a subset of the input variables) Random Forest is an example of this approach.

About Me



Ricky Ho

I am a software architect and consultant passionate in Distributed and parallel computing, Machine learning and Data mining, SaaS and Cloud computing.

[View my complete profile](#)

Popular Posts

MongoDb Architecture

NOSQL has become a very heated topic for large web-scale deployment where scalability and semi-structured data driven the DB requirement tow...

Designing algorithms for Map Reduce

Since the emerging of Hadoop implementation, I have been trying to morph existing algorithms from various areas into the map/reduce model. ...

NOSQL Patterns

Over the last couple years, we see an emerging data storage mechanism for storing large scale of data. These storage solution differs quite...

Couchbase Architecture

After receiving a lot of good feedback and comment on my last blog on MongoDB, I was encouraged to do another deep dive on another popular ...

Predictive Analytics: Overview and Data visualization

I plan to start a series of blog post on predictive analytics as there is an increasing demand on applying machine learning technique to ana...

Predictive Analytics: Generalized Linear Regression

In the previous 2 posts, we have covered how to visualize input data to explore strong signals as well as how to prepare input data to a fo...

BigTable Model with Cassandra and HBase

Recently in a number of "scalability discussion meeting", I've seen the following pattern coming up repeatedly ... To make you...



87

Blog Archive

- 2013 (8)
- ▼ 2012 (18)
 - November (1)
 - October (1)


Sliding window

Machine Learning is based on an important assumption that the future is repeating the same pattern as the history. As we all know, as time goes by, this assumption becomes more and more invalid. In other words, we should put more weight in recent history than long term history.

Ensemble method also gives us an elegant way to decay the weight of old data. In this case, we maintain a sliding window of models (say the last 7 days). We learn a model daily and expire the oldest model we compute 7 days ago.


M = M1 + M2.a + M3.a^2 + ... + M7.a^6
M = M / (1 + a + a^2 + ... + a^6)

Posted by [Ricky Ho](#) at 7:45 AM

 +9 Recommend this on Google

Labels: [data mining](#), [ensemble method](#), [machine learning](#)

1 comment:

 **Unknown said...**

Hi Rick !
I'm looking for a research topic for my MD, I'm interested in Ensemble methods in general but I'm more into industry and somewhat away from academia.
How could I find a reasonably hot topic related to ensemble methods ?
I've been suggested to explore the latest three years' papers in all DMAI/ML related conferences, but that in itself is somewhat of a daunting task.
Can you provide any advice ?
Thanks in advance,
Daniel

March 24, 2012 at 8:39 AM

[Post a Comment](#)

[Newer Post](#)

[Home](#)

[Older Post](#)

Subscribe to: [Post Comments \(Atom\)](#)

- ▶ [September](#) (1)
 - ▶ [August](#) (2)
 - ▶ [July](#) (1)
 - ▶ [June](#) (3)
 - ▶ [May](#) (3)
 - ▶ [April](#) (3)
 - ▶ [March](#) (1)
 - ▶ [February](#) (1)
 - ▼ [January](#) (1)
- [Machine Learning: Ensemble Methods](#)
- ▶ [2011](#) (6)
 - ▶ [2010](#) (18)
 - ▶ [2009](#) (31)
 - ▶ [2008](#) (22)
 - ▶ [2007](#) (11)

Search This Blog

Search

- Labels**
- [machine learning](#) [data mining](#) [map reduce](#) [Architecture](#) [Design](#) [Cloud computing](#) [algorithm](#) [NOSQL](#) [Hadoop](#) [scalability](#) [Distributed system](#) [parallel processing](#) [big data](#) [predictive analytics](#) [Design patterns](#) [SOA](#) [performance](#) [REST](#) [ensemble method](#) [recommendation engine](#)
- Pages**
- [Home](#)