

elect 式 value-tracker

e.tmailbank@gmail.com

1 目標を絞った現段階での目標

ある目標 に対する収益予測 API を組んで見る

- 例えば...
つくば市のコンビニ A の売上予想 ... 必要なデータ例

内容	気温	湿度	仕入れ量	交通量	株価	人口遷移
重み	5	4	2	5.5	0.1	0.2

1.1 データの入手

はじめのデータは既存のデータ提供サービスから入手する。例えば,json などの形式で入手した後、それをデータベースに登録する。最終的には利用者から提供されたデータ（試行に用いられた）データなどの利用も視野に入りたい。
実験初期のデータは国の出している財務関係のデータなどを利用したいと考えている。

1.2 検索

関連性のあるデータを計算に入れることにするが、その関連性のあるデータの選択については現在検討中である。（利用者に選択してもらうスタイルを取る、何らかの学習でサジェストするなど）

1.3 データの出力

データの出力は利用者の入れたデータに関連して、利用者の求めたデータを返す。

2 天竺先生にお願いしたいこと

- データベースアプリケーションの選び方、使い方、運営方法、最適化など

例えば Hbase を使う場合には、どのようにデータを格納することが効率的なのか。また、機械学習側にデータを渡す際にどのように出力できるのか。

- 機械学習をするに当たっての基礎知識として身につけるべきもののリスト
どんなものを勉強すべきなのか。どこから手をつけて行けば良いのか
- 他の担当との進捗に合わせることになるが、その使われたデータベース (ex.Hbase) から機械学習ツール (ex.TensorFlow) にデータを持ち込んでいく部分についてのアドバイス。
- その他、理解で行き詰まった部分でのアドバイス

3 自分たちで進めるもの

- 示された方法の学習

使うべきツールが示されれば、それについては 基本的には 自分達で学習します。

- この演習全体のスケジュール管理

グループリーダーが統括します。

4 具体的な内容を詰めてみる。

4.1 データベース

4.1.1 データベースにほしいもの

より高精度な計算のためあらゆるデータが欲しい (※すべてのデータを計算するかどうかは別)

4.1.2 データをどうやって集めるのか

前項で述べたとおり、初期段階では例えば weather.api などからの公開されているデータを、それからは各利用者から与えられるデータを集める。

4.1.3 データベースを管理するに当たって

Hbase などの導入を考えている。

4.1.4 処理するデータの量

取捨選択して毎 s～毎日に合計で GB 単位で流し込んでみたい。

4.2 機械学習

4.2.1 機械学習とは

初学者ながら、TensorFlow という機械学習のツールを用いることを考えている。他に
ある素晴らしいツールが目につけば、それを用いる。

4.2.2 処理するデータの大きさは

GB 単位になると予想している。

4.2.3 与えられるデータは

主に数値データを扱う。

5 全体を通して担当教員に求めたいもの

大規模データの機械学習、データベース処理に関する知識。
また、それらをつなげるためのメソッド。
更にそれに関するロードマップの制作。
疑問点ができ次第 質問させていただきたいと思っています。

5.1 現在質問したい内容

- 大規模なデータを毎秒単位で収集しなければならない可能性がでてくるが、どうやってコンピュータで処理すべきなのか（メモリ等の問題）

- 大容量のデータを処理できるデータベースにはどのようなものがふさわしいか (Hbase でこと足りるのか)
- データベースと機械学習をつなげるパイプをどのように設計すればいいのか
- 機械学習全般の知識 (学習は主には自分たちでついていく予定です)

6 担当教員に求めないもの

プログラミング言語の教育技術 (使う言語は自分達で学習)
全体のスケジュール管理 (グループリーダーが統括する予定です)

7 最終目標

ある目標 → 社会で通用する収益予測 API の作成
公開できる〇〇予測 API にする。