1)kmeansについて:

kmeansはあるデータ群のデータを分類する方法である。本来ユーザが決めたk個のクラスタに分類することに使われているが、そうすると主観性が高く、分類されたデータは本当に違いのあるデータかどうかは確かでないため、kmeans+は考案された。kmeans+とは、kはユーザで決められた数でなく、アルゴリズムによって導き出された数である。ここでは比較的に簡単な方法で（ランダムで生成した）初期のクラスタ中心を増やし、減らすことにした。これは正確性、欲しい情報により変更できる物とされている。

このプログラムによるとiris.txtの一番目と二番目のデータに対して分析すると、2つのクラスタに分類できる。しかし、実際は3種類の花から取ったデータなのでそこに違いが見える。だからといって、プログラムの結果が違うとも言えない。それはその3種類に分類したことも一つの分類法による結果に過ぎないからである。

　つまり、kmeansは現実を反映する道具というより、もう一つの観点を提供することができる道具であることが言える。それを踏まえて、次の使い方が考えられる。これから成長する企業を見つけたいとすると、いま成長した企業の昔のデータ（売上、固定資産、現金等々）をクラスタ中心として設定し、今の目立たない企業のデータに対して分析を行うと、一緒のクラスタに入れば、もっと研究する価値がある目立たない企業を発見できるかもしれない。また、貴社のアプリ、「あすかぶ！」ではユーザは株が上がるか下がるかの予想をする機能があるのを利用し、各ユーザの正解率と回答回数に対して分析を行うと、各クラスタにランクをつけ、正解率高く、回答回数が多い順に、相当する票数が多いこととかにできるのではないか。それで何が変わるかは予想できないのだが、一つの試みとして考えられる。

(P/S:経済関連の用語は詳しくありませんので、分かりにくいところがあれば申し訳ありません。)

2)Particle Filterについて

Particle Filterは簡単に言うと変化しているデータの中にある個体のデータを特定することができるアルゴリズムのこと。具体的な計算は調べたがプログラムを組むまでのかかる時間が長くなりそうで、個人的に新しい領域でもあるので、実装までの時間が足りないため、未完成のプログラムを提出することにした。Particle Filterの機能の一つのキーワードは「確率」。ある個体がどの確率でどの値になるのと、次のステップ（データの変化）でその値はあるのかを重ねて学習して行くと、個体を特定できる。つまり、それを達成したあと、次のステップでどの確率でどの方向に変化していくことはその個体の変化の予想も可能になるはず。そこで、次のような使いかたが考えられる。金融商品などの価格に対して分析を行うと次のステップの価格の確率分布が得られると考えられる。金融商品以外の使い道だと、また「あすかぶ！」の例でいうと、各ユーザの回答を四つに分類し（「上がる、正解」、「下がる、正解」、「上がる、外れ」、「下がる、外れ」）、時間系列に並べて分析すると、次の正解率の確率分布が求められるかもしれない。