

Algorytm indukcji klasyfikatora za pomocą EA z automatycznym przełączaniem ukierunkowań

Anna Manerowska, Michał Kozakiewicz

2.12.2009

1 Wstęp

Jako projekt na przedmiot MEUM (Metody Ewolucyjne Uczenia Maszyn) dostaliśmy za zadanie stworzyć klasyfikator rozwijany za pomocą algorytmu ewolucyjnego, którego ukierunkowanie zmieniałoby się podczas uruchomienia. EA zostanie użyty do zmian drzew decyzyjnych. Zmienne ukierunkowanie rozwoju drzew uzyskamy przez modyfikację funkcji celu, oceniającej zarówno efektywność klasyfikacji jak również strukturę drzewa.

2 Opis problemu klasyfikacji

Zadaniem klasyfikacji będzie zakwalifikowanie instrumentu giełdowego do jednej z trzech grup: [kup, sprzedaj, nie ruszaj].

- kup - algorytm uważa, że cena danego instrumentu wzrośnie
- sprzedaj - algorytm uważa, że cena danego instrumentu zmaleje
- nie ruszaj - algorytm nie potrafi zdecydować

3 Dane

Klasyfikację będziemy przeprowadzać na cenach archiwalnych wybranych instrumentów. W zależności od analizowanego okresu czasu, wymagany będzie inny typ danych:

- Intra-day trading, High Frequency trading - czyli kupno i sprzedaż instrumentów wielokrotnie w ciągu jednego dnia, nie pozostawianie otwartych pozycji na noc. Wymaga danych tick-to-tick czyli każdej zmiany ceny instrumentu w ciągu dnia; Dane tick-to-tick są z reguły dostępne za opłatą, zajmują też olbrzymią ilość miejsca;
- Daily trading - czyli transakcje na przestrzeni dni/tygodni; Nie potrzebujemy danych tick-to-tick, wystarczy parametry takie jak:
 - cena otwarcia
 - cena zamknięcia
 - cena maksymalna/minimalna z danego dnia
 - ilość transakcji (volume)

Dane o tej częstotliwości można zdobyć np z Yahoo Finance[12]: dla okresów o dowolnej, mającej znaczenie dla algorytmu, długości.

- Long term trading - zakłada długotrwałe inwestycje, tą formą nie będziemy się zajmować

Analiza rynków znacząco różni się w zależności od częstotliwości transakcji, którą rozpatrujemy[6]. Jeżeli nie uda się zdobyć wystarczającej ilości danych intra-day, zajmiemy się klasyfikacją danych dziennych.

Wektorem wejściowym będzie ciąg cen (oraz innych dostępnych parametrów) z pewnego okresu czasu (zależnie od wybranej częstotliwości), również dla instrumentów innych niż analizowany.

4 Narzędzia realizacji

Podczas pracy nad algorytmem, prototyp powstanie w środowisku R, natomiast wersja ostateczna wykonana będzie w Javie, przy użyciu następujących bibliotek:

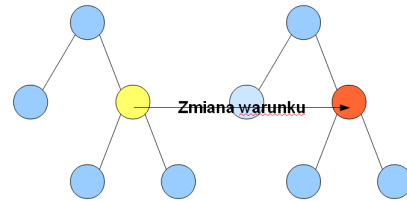
- WEKA [10] - biblioteka do eksploracji danych, zawiera implementację różnych algorytmów budowy drzew decyzyjnych (m.in. ID3)
- Watchmaker [11] - biblioteka do metod ewolucyjnych

Wybór języka kompilowanego, jest uzasadniony dużymi wymaganiami obliczeniowymi metod ewolucyjnych; Javy, ze względu na dostępność bibliotek i łatwość programowania.

5 Opis algorytmu

Prototyp algorytmu wygląda następująco:

- zaczynamy od populacji bazowej drzew wygenerowanych algorytmem ID3, których struktura i parametry są zmienione losowo



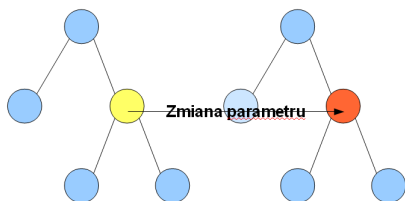
Rysunek 1: Mutacja warunku

- osobniki dokonują klasyfikacji
- oceniamy sprawność osobników na podstawie pewnej funkcji celu F
- wybieramy x osobników do mutacji, metodą ruletkową
- z pośród nowej populacji (powiększonej o x) usuwamy x osobników o najniższej sprawności
- proces powtarzamy przez g generacji

Operator mutacji dla drzew decyzyjnych może zmieniać wartości parametrów (Rys. 2) oraz warunki podziału (Rys. 1) węzła drzewa.

Funkcja celu F będzie zmieniać się w trakcie działania algorytmu. Jako podstawę sprawności wygenerowanego osobnika uznamy ilość poprawnych klasyfikacji. Dodatkowymi warunkami, aktywowanymi po pewnej (jeszcze nie ustalonej) liczbie generacji, bądź przy spełnieniu określonych wymagań (np. osobnik klasyfikuje poprawnie częściej niż 50%), będą:

- ocena potencjalnej wielkości zysku, obliczona na podstawie dalszego wzrostu bądź spadku ceny instrumentu



Rysunek 2: Mutacja parametru

- ocena struktury drzewa decyzyjnego, np kara za wysokość
- możliwość oceny na podstawie historii jeżeli osobnik przeżył do następnych generacji

6 Oczekiwania

Podstawowym pytaniem jest, czy algorytm tego typu może wygenerować zysk. Nawet jeżeli to się nie uda, chcielibyśmy zaobserwować czy algorytm wykryje korelację między różnymi instrumentami (np. akcjami firm działających w jednej branży [9]). Jakie parametry będą używane do klasyfikacji, np czy ilość transakcji będzie miała znaczenie? Czy model cen wygenerowany przez algorytm będzie miał cechy wspólne z innymi, znanymi modelami rynków[8]? Interesującym jest również jak kara za wysokość drzewa wpłynie na zachowanie algorytmu - dążenie do upraszczania drzew decyzyjnych może prowadzić np do zwiększenia odporności na zakłócenia czy zwiększania zdolności algorytmu do generalizacji.

7 Możliwe rozszerzenia

Możliwe są rozszerzenia do algorytmu bazowego (o różnej złożoności) np:

- informacja o pewności z jaką dokonana była klasyfikacja
- informacja o przewidywanej długości trendu instrumentu (np dla [kup] chcielibyśmy wiedzieć jak bardzo cena może jeszcze urosnąć)
- propozycja zleceń na innych instrumentach zmniejszających ryzyko danej inwestycji (Ang. Hedge)
- zastosowanie algorytmu w czasie rzeczywistym na aktualnych cenach obowiązujących na giełdzie
- zastosowanie predefiniowanych (wybranych przez użytkownika) węzłów drzew, zwiększających determinizm algorytmu, parametry których mogły by być rozwijane drogą ewolucji

8 Podsumowanie

Algorytm ewolucyjny jest nieintuicyjny podczas gdy umysł ludzki poszukuje rozwiązań na utartych ścieżkach, ciężko mu wybić się ze schematu. Najczęściej sprawdza jedynie proste symetryczne konstrukcje. Algorytm ewolucyjny jest wolny od tych ograniczeń. Swobodnie sprawdza wszystkie poprawne rozwiązania, poszukując najlepszego. Istotną zaletą ewolucyjnego podejścia do modelowania jest również znaczne ograniczenie wiedzy a priori. W ekonometrii to analityk decyduje o formie funkcyjnej modelu, sugerując się testami statystycznymi, wiedzą ekonomiczną na

temat modelowanego zjawiska oraz własnym doświadczeniem. W podejściu ewolucyjnym to ewolucja steruje wyborem najbardziej odpowiedniej funkcji. Proces modelowania jest więc w pełni sterowany danymi, bez wiedzy z zewnątrz. Wykonanie projektu pozwoli nam, w jakimś stopniu na ocenę wartości algorytmu ewolucyjnego w modelowaniu ekonomicznym.

Literatura

- [1] D. E. Goldberg: Algorytmy genetyczne i ich zastosowania. Warszawa: WNT, 1998.
- [2] Paweł Cichosz: Systemy uczące się, WNT, Warszawa 2000
- [3] Arabas J., Wykłady z algorytmów ewolucyjnych, WNT, Warszawa 2001
- [4] Kwaśnicka H.: Obliczenia ewolucyjne w sztucznej inteligencji, Oficyna Wydawnicza PWr., Wrocław, 1999
- [5] Michalewicz Z.: Algorytmy genetyczne + struktury danych = programy ewolucyjne, WNT, Warszawa, 1996
- [6] blog: <http://tr8dr.wordpress.com/>
- [7] blog: <http://www.puppetmastertrading.com/blog/>
- [8] blog: <http://www.maxdama.com/>
- [9] Carol Alexander: Market Models: A Guide to Financial Data Analysis, Wiley, 2001
- [10] <http://www.cs.waikato.ac.nz/ml/weka/>
- [11] <http://watchmaker.uncommons.org/>
- [12] <http://finance.yahoo.com/>