Laboratorium „Rachunek Prawdopodobieństwa i Statystyka” — zadanie zaliczeniowe 2010/11

10 stycznia 2011

Wstęp

Manager uczelnianego bufetu musi na co dzień rozwiązywać ciekawy problem optymalizacyjny. Z jednej strony niższe koszty przygotowania posiłków pozwalają na większe marże i potencjalnie większy zysk, z drugiej mogą spowodować, że mniej osób będzie korzystać z usług bufet (nie każdy chce jeść na obiad kaszę ze skwarkami), co oznacza mniejszy zysk. W tym zadaniu opisujemy bardzo prosty model takiej sytuacji. Twoim zadaniem będzie zarządzanie bufetem.[1]

Model

Ceny, koszty i zyski

Przyjmujemy, że zestaw obiadowy ma stałą cenę C zł. Koszt przygotowania zestawu (w tym koszt produktów) może się zmieniać z dnia na dzień. Koszt przygotowania zestawu w dniu i będziemy oznaczać przez x_i, $0 \leq x_i \leq C$. Jeśli w dniu i-tym zestaw kupi n_i osób, to zysk z dnia i-tego jest równy $n_i(C - x_i)$.

Klienci

Bufet ma N klientów. Każdego z klientów można opisać trójką liczb rzeczywistych: (α, p^-, p^+). Liczby te interpretujemy następująco:

- jeśli koszt przygotowania obiadu był $< \alpha$, to klient kupi zestaw obiadowy tylko jeśli nie ma innego wyjścia, t.j. z prawdopodobieństwem p^-,
- jeśli koszt przygotowania obiadu był $\geq \alpha$, to klient kupi zestaw obiadowy, o ile nie będzie miał zaplanowanego obiadu w innym miejscu, t.j. z prawdopodobieństwem p^+.

Przyjmujemy oczywiście, że $p^+ \geq p^-$.

Liczby (α, p^-, p^+) odpowiadające różnym klientom nie są znane managerowi, może on jednak starać się je ocenić obserwując decyzje klientów.

Problem

Twoim zadaniem jest opracowanie strategii wyboru kosztu przygotowania obiadu tak, aby zmaksymalizować łączny zysk w okresie $1 \leq M \leq 1\,000$ dni. Przyjmujemy następujące założenia:

- Cena zestawu obiadowego wynosi 10 zł.

[1] Opisany tutaj bufet nie ma nic wspólnego z żadnym istniejącym bufetem. W szczególności nie twierdzimy nawet, że osoby zarządzające bufetami uczelnianymi zastanawiają się nad tego rodzaju problemami.

- Istnieje nie więcej niż $1 \leq k \leq 5$ grup/rodzajów klientów różniących się parametrami (α, p^-, p^+), klienci w tej samej grupie mają te same parametry. Nie znasz podziału na grupy, ale znasz liczbę grup k.
- Klientów jest $N = 240$ i wszystkie grupy są równoliczne, tzn. rozmiaru N/k.
- Dla każdej z grup parametry (α, p^-, p^+) ustalane są następująco: α jest wybierana losowo z rozkładem jednostajnym z przedziału $[0, C]$, p^- z przedziału $[0, 0.5]$, p^+ z przedziału $[0.5, 1]$.
- Po każdym dniu otrzymujesz dla każdego klienta informację o tym, czy kupił on zestaw obiadowy.

Sposób oceny rozwiązania

Zadanie ma charakter otwarty i nie jest istnieje „poprawne rozwiązanie". Chcielibyśmy jednak, aby rozwiązujący:

1. zaproponował i zaprogramował w języku R mechanizm testowania i różnych strategii,
2. zaproponował kilka różnych nietrywialnych strategii i zbadał ich zachowanie, w szczególności zastanowił się nad sytuacjami, w których strategie te działają najlepiej/najgorzej (w zależności od M, k, wartości wylosowanych parametrów, itp),
3. opisał wyniki swoich eksperymentów w postaci raportu, w szczególności zilustrował je graficznie; osobom poszukującym inspiracji polecamy stronę http://addictedtor.free.fr/graphiques/, czyli „R Graph Gallery", gdzie oprócz ilustracji różnych rodzajów wykresów można znaleźć przykłady generującego je kodu.

Mile widziane są również wszelkiego rodzaju rozważania teoretyczne, jak również uogólnienia modelu, np.:

- nieznane k, nieznane liczności grup, ew. każdy klient z własnymi parametrami,
- bardziej skomplikowany model klienta,
- ....

Wskazówki i sugestie

Aby ułatwić rozpoczęcie pracy nad projektem, zamieszczamy poniżej listę przykładowych pytań, nad którymi być może warto się zastanowić:

- Jak rozwiązać problem w przypadku, gdy znamy wartości parametrów wszystkich grup?
- Jak rozwiązać problem, gdy interesuje nas maksymalizacja zysku w bardzo długim okresie, np. $M = 10\,000$?
- Jak, mając dane z długiego okresu, ocenić wartości parametrów?
- Jak rozwiązać problem, gdy interesuje nas maksymalizacja zysku w bardzo krótkim okresie, np. $M = 1, 2, 3, \ldots$?
- Kiedy twój algorytm działa najlepiej/najgorzej?
- Czy można go jakoś poprawić tak, aby w tej najgorszej sytuacji działał lepiej?