Hanbit eBook

Realtime 54

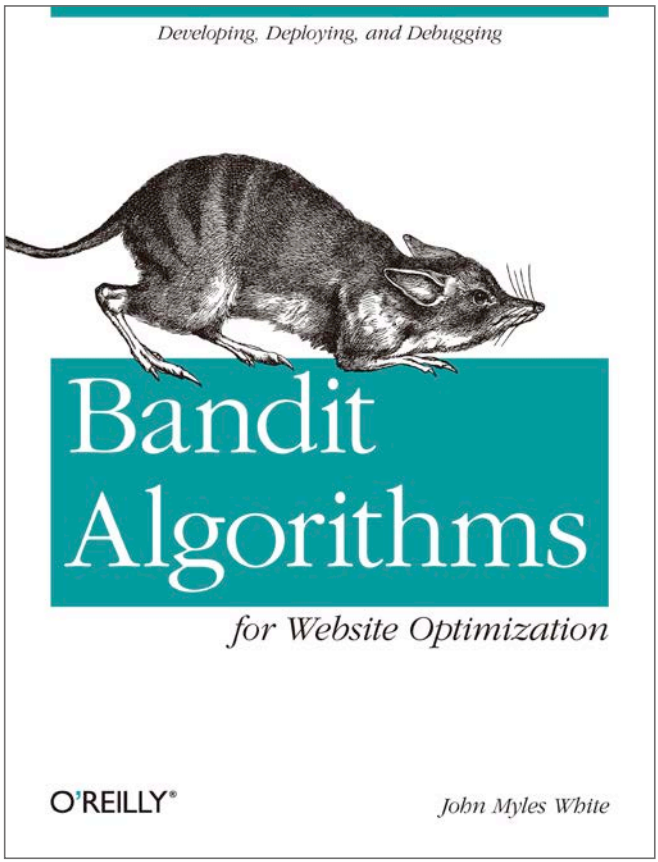
웹 사이트 최적화를 위한

밴디트 알고리즘

Bandit Algorithms for Website Optimization

존 마일즈 화이트 지음 / 김학민 옮김

O'REILLY® 한빛미디어 Hanbit Media, Inc.



이 도서는 O'REILLY의
Bandit Algorithms for Website Optimization의
번역서입니다.

웹 사이트 최적화를 위한 **밴디트 알고리즘**

초판발행 2014년 05월 02일

지은이 존 마일즈 화이트 / **옮긴이** 김학민 / **펴낸이** 김태헌
펴낸곳 한빛미디어(주) / **주소** 서울시 마포구 양화로 7길 83 한빛미디어(주) IT출판부
전화 02-325-5544 / **팩스** 02-336-7124
등록 1999년 6월 24일 제10-1779호
ISBN 978-89-6848-660-9 15000 / **정가** 9,900원

책임편집 배용석 / **기획·편집** 김창수
디자인 표지 여동일, 내지 스튜디오 [밈], 조판 최송실
영업 김형진, 김진불, 조유미 / **마케팅** 박상용, 서은옥, 김옥현

이 책의 저작권은 오라일리 사와 한빛미디어(주)에 있습니다.
한빛미디어 홈페이지 www.hanbit.co.kr / **이메일** ask@hanbit.co.kr

Published by HANBIT Media, Inc. Printed in Korea Copyright ⓒ 2014 HANBIT Media, Inc.
Authorized Korean translation of the English edition of Bandit Algorithms for Website Optimization, ISBN 9781449341336 ⓒ 2013 John Myles White. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.
이 책의 저작권은 오라일리 사와 한빛미디어(주)에 있습니다.
저작권법에 의해 보호를 받는 저작물이므로 무단 복제 및 무단 전재를 금합니다.

지금 하지 않으면 할 수 없는 일이 있습니다.
책으로 펴내고 싶은 아이디어나 원고를 메일(ebookwriter@hanbit.co.kr)로 보내주세요.

한빛미디어(주)는 여러분의 소중한 경험과 지식을 기다리고 있습니다.

저자 소개

지은이_ **존 마일즈 화이트**John Myles White

프린스턴 대학원의 심리학 박사 학위 예정자다. 패턴 인식, 의사 결정, 행동 방법론과 '기능적 자기공명영상fMRI'을 이용한 경제 행동을 연구하였다. 특히, 가치 평가의 예외성에 관심이 많다. 학술 활동 이외에도 오픈 소스 소프트웨어 접근 방식을 데이터 분석에 접목한 데이터 과학 운동에 아주 많이 관여하고 있다. 또한, ProjectTemplate과 log4r 등 인기 있는 R 패키지의 리드 메인테이너Lead Maintainer다.

역자 소개

옮긴이_ **김학민**

2001년 3월부터 지금까지 삼성SDS에 근무 중이다. 초반에는 개발자와 Engineer의 삶을 살았고, 몇 년 전부터는 IT Architect로서 활동 중이다. 주로 대규모 웹 시스템의 기술 아키텍처를 설계하는 업무를 하고 있고, 이 때문에 Java, DBMS, OS, Backup 등 매우 다양한 IT 분야에 대한 지식을 항상 습득하려 노력 중이다.

회사에 다니는 도중 연세대학교 공학대학원 컴퓨터공학과에 입학해서 2013년 8월에 석사 학위를 취득했다. 대학원에서 인공 지능 수업을 들으면서 기계 학습에 많은 관심을 두게 되었다. 기계 학습과 관계가 있는 밴디트 알고리즘을 처음 접했을 때 우리나라에 이 알고리즘을 소개하는 첫 책을 번역하고 싶다는 욕심으로 번역을 시작했다.

저자 서문

이 책의 코드 찾기

이 책은 알고리즘에 관한 책이다. 그렇다고 알고리즘의 이론을 다루는 책은 아니다. 새로운 아이디어들을 실제로 실험하며 배우기를 좋아하는 사람에게 초점을 맞춘 알고리즘에 대한 간략한 입문서다.

실험하며 배우기를 좋아하는 여러분을 위해 여러분이 좋아하는 프로그래밍 언어용 인터프리터를 곁에 두고 읽을 수 있게 구성하였다. 책에서 서술된 모든 알고리즘은 파이썬을 사용하여 설명하였다. 본문에서 링크 형태로 제공하는 외부 온라인 자료 중 일부는 밴디트 알고리즘을 구현하는 데 이상적인 새로운 프로그래밍 언어, 줄리아Julia [01]로 구현한 코드다. 파이썬과 줄리아 코드 외에도 자바스크립트와 같은 다른 언어들로 구현한 코드도 있다.

파이썬은 모든 프로그래머들에게 꽤 괜찮은 공용어로 생각하여 이 책에서는 파이썬을 사용하였다. 파이썬이 여러분 스타일이 아니라 해도 그리 어려운 언어가 아니므로, 책에 나온 파이썬 코드는 여러분이 선호하는 프로그래밍 언어로 쉽게 전환할 수 있을 것이다.

파이썬이나 줄리아로 코딩을 하는 걸 좋아한다면 이 책에 나온 코드는 GitHub (https://github.com/johnmyleswhite/BanditsBook)에서 찾을 수 있다. 코드의 오류를 발견하거나 다른 언어로 구현한 것을 제출하고 싶다면, Pull Request(제

01 (역자주) 줄리아는 매트랩(MATLAB)과 같은 테크니컬 컴퓨팅(Technical Computing)을 위한 고수준, 고사양의 동적 프로그래밍 언어이다. 줄리아는 정교한 컴파일러, 분산 병렬 실행, 수치적 정확성을 제공한다. 그리고 선형대수, 난수 생성, 신호 처리, 스트링 처리와 같은 수많은 수학 함수 라이브러리를 제공한다. 자세한 내용은 http://julialang.org를 참조하라.

출 요청)[02]를 하기 바란다.

전문 용어 다루기. 용어 사전

이 책의 목적은 '멀티암드 밴디트 문제Multiarmed Bandit Problem'에 대한 이론적 연구를 소개하거나 이 문제를 풀기 위해 독창적인 알고리즘을 활용하도록 여러분을 준비시키려는 것이 아니다. 다만, 여러분이 이 책을 읽고 나서는 멀티암드 밴디트 문제에 관한 문헌을 이해할 수 있도록 기존 연구 결과에 대한 충분한 이해를 가지기를 바랄 뿐이다. 여러분의 이해를 돕기 위해 꽤 많은 전문 용어가 나온다. 이 전문 용어는 처음엔 조금 이상하게 느껴질 수 있지만, 멀티암드 밴디트 문제에 대한 학술 문헌에서 보편적으로 사용하는 용어들이다. 이 책을 읽으면서 전문 용어의 뜻이 가물가물하다면 아래에 있는 전문 용어 목록을 참고하면 된다. 당장에라도 이 전문 용어를 쭉 훑어볼 수 있겠지만, 아직은 여러분이 이 단어들을 이해할 것이라고 기대하지 않는다. 책에서 사용하는 용어가 혼란스러울 때 참고할만한 자료가 여기 있다는 것만 알아두기 바란다.

보상 Reward

성공success의 양적 측정(값). 비즈니스 측면에서는 궁극적인 보상이 이익이지만, 책에서는 광고 클릭률 또는 신규 사용자 등록률과 같이 간단한 메트릭metric를 보상으로 다룬다. 중요한 것은 (A) 명확한 '양적 측정 등급quantitative scale'이 있는가와 (B) 보상을 적게 하는 것보다는 많이 하는 것이 좋은가다.

---

02 (역자주) Git 자체에 있는 기능은 아니고 Github에서 제공하는 기능으로 수정된(또는 새로운) 내용을 Pull Request로 보내면 Repository(저장소)의 커미터가 내용을 확인한 뒤 승인하면 보낸 내용이 원본 저장소에 머지(merge)된다(물론 거절당할 수도 있다).

암 Arm[03]

어떤 옵션option이 우리에게 유용한가? 어떤 행동을 우리가 취할 수 있는가? 책에서는 사용 가능한 옵션을 '암Aarm'으로 지칭할 것이다. 멀티암드 밴디트 문제에 관한 역사의 일부를 듣고 나면 이렇게 명명한 이유를 쉽게 이해할 수 있을 것이다.

밴디트 Bandit

'밴디트Bandit'는 암들의 집합이다. 유용한 옵션을 많이 가지고 있을 때 이 옵션의 집합을 '멀티암드 밴디트Multiarmed Bandit'라 한다. 멀티암드 밴디트는, 여러분이 취할 수 있는 행동이 많을 때와 이 행동을 수행한 후 얻게 될 보상에 대해 불완전한 정보를 가지고 있을 때, 결정을 내리는 법을 추론하는 데 사용할 수 있는 수학적 모델이다. 이 책에서 제시된 이 알고리즘은 언제 어느 암을 당길지 결정하는 문제를 해결하기 위한 방법이다. 잡아당길 암을 선택하는 문제를 '멀티암드 밴디트 문제'라 부른다.

플레이/실험 Play/Trial

밴디트를 다룰 때 각 암들을 여러 차례 당기게 된다고 가정한다. 암을 당겨야만 하는 각 기회를 '플레이Play'라 하거나 좀 더 자주 '실험Trial'이라 할 것이다. '플레이'는 '암'에 영감을 준 도박의 개념을 불러일으키는 데 도움이 되나 '실험'이 보편적으로 더 많이 사용된다.

시야 Horizon

게임이 끝나기 전에 시도할 수 있는 실험을 얼마나 많이 가지고 있는가? 이렇게 남은 실험의 횟수를 '시야Horizon'이라 한다. 각 암을 플레이할 기회를 많이 가진다는 것은 잘못되더라도 여전히 이를 복구할 수 있는 시간을 가진다는 뜻이므로 더 큰

---

03 (역자주) '멀티암드 밴디트 문제(Multiarmed Bandit Problem)'은 도박장에 여러 대 놓여 있는 슬롯머신 중 어떤 것을 선택할 때 상금을 가장 많이 획득할 것인지에 대한 확률 문제이다. 이 문제에서 '밴디트(Bandit)'는 슬롯머신이고, '암(arm)'은 슬롯머신에 달려있는 레버를 의미한다. 하지만 멀티암드 밴디트 문제 자체가 학술 용어로 사용되고 있어 밴디트와 암을 의역하는 것보다 원어 자체로 사용하는 것이 더 이해하기 쉽다 판단하여 이 책에서는 원어를 그대로 사용한다.

위험을 감수할 수 있다. 따라서 시야가 길 때 사용하는 전략과 시야가 짧은 때 사용하는 전략은 많은 경우 다를 것이다.

활용 Exploitation

멀티암드 밴디트 문제를 풀기 위한 알고리즘에서 '활용'이란 이전 플레이들을 기반으로 한 가장 높은 추정치를 가지고 암을 사용하는 경우이다.

탐색 Exploration

멀티암드 밴디트 문제를 풀기 위한 알고리즘에서 '탐색Exploration'이란 이전 플레이들을 기반으로 한 가장 높은 추정치를 가지지 않는 암을 사용하는 경우이다. 다시 말해, 탐색은 활용이 일어나지 않을 때마다 발생한다.

탐색/활용 딜레마 Explore/Exploit Dilemma

모든 학습 시스템은 탐색에 대한 자극과 활용에 대한 자극 사이에서 타협을 해야 한다는 것을 의미한다. 이 딜레마에 대한 정확한 솔루션은 없으나, 책에서 기술한 알고리즘들은 탐색과 활용의 상반된 목표를 풀기 위한 유용한 전략들을 제공한다.

어닐링 Annealing

멀티암드 밴디트 문제를 풀기 위한 알고리즘에서, '어닐링Annealing'이란 시간이 지남에 따라 탐색을 덜 하는 경우를 말한다.

온도 Temperature

멀티암드 밴디트 문제를 풀기 위한 소프트맥스Softmax 알고리즘에서 탐색의 양을 증가하기 위해 조정하는 파라미터parameter. 시간이 지남에 따라 '온도' 파라미터를 줄일 경우 알고리즘은 어닐링하게 된다.

스트리밍 알고리즘 Streaming Algorithms

한 번에 데이터 한 조각을 처리하는 경우 해당 알고리즘을 '스트리밍 알고리즘Streaming Algorithm'이라 한다. 처리를 위해 모든 데이터에 접속할 필요가 있는 '배치

처리 알고리즘batch processing algorithm'과는 반대 개념이다.

온라인 학습 Online Learning

한 번에 데이터 한 조각만 처리하고 데이터의 각 조각들이 보인 후 분석의 잠정 결과를 제공할 경우 해당 알고리즘을 '온라인 학습 알고리즘'이라 한다.

능동 학습 Active Learning

가장 효과적인 학습을 위해 다음으로 보기를 원하는 데이터의 조각을 결정할 수 있는 경우 해당 알고리즘을 '능동 학습Active Learning' 알고리즘이라 한다. 대부분의 전통적 기계 학습 알고리즘들은 능동적이지 않다. 이들은 우리가 제공하는 데이터만 수동적으로 받아들이고 우리가 다음으로 수집해야 하는 데이터가 무엇인지 알려주지 않는다.

베르누이 공식 Bernoulli

베르누이 시스템은 확률 p에 대해 1을 출력하고, 확률 1-p에 대해 0을 출력한다.

감사의 글

이 책은 '탐색-활용 문제Explore-Exploit problem'에 대해 필자가 프린스턴 대학 심리학과 멤버들과 수년간 토론했던 결과의 산물이다. 그들에게 감사를 전하고 또한 3명의 테크니컬 리뷰어인 Conductrics의 매트 거쇼프Matt Gershoff, Esty의 로베르토Roberto, RTI의 팀 호퍼Tim Hopper에게도 감사를 전한다. 이들 모두는 이 책을 모두 읽고 수많은 개선 사항을 찾아주었다. 그들의 의견들은 이루 말할 수 없는 값어치가 있었다. 또한 책의 최종 출간에 이르는 동안 거의 에러 없이 진행해주어 너무나 감사할 따름이다. 마지막으로 이 책에서 기재된 밴디트 알고리즘의 코드 구현에 공헌해준 수많은 사람들에게 감사를 전한다. 출간되지 않은 책을 위한 보충 코드에 공헌한 Pull request를 받는 것은 작가로서도 가장 즐거운 경험이었다.

역자 서문

편집자로부터 원서를 받았을 때 처음 보는 알고리즘이라 무척 흥미가 생겼다. 책을 구해서 읽어보니 생각 외로 재미있어서 기회가 된다면 이 책 번역은 내가 하고 싶다고 편집자에게 부탁했었다. 다행히 편집자가 나를 무척 잘 봐준 관계로 번역할 수 있는 기회를 가지게 되었다.

아마존으로부터 시작된 웹 사이트 최적화를 위한 A/B 테스팅은 인터넷 기반 전자상거래에서는 이제 필수 불가결한 요소가 되었다. 조금이라도 신규 고객을 더 끌어들이려 하는 인터넷 쇼핑몰들은 어느 누구라도 모두 쓸 것이다. 하지만 2가지 경우를 놓고 비교하여 어느 쪽의 접속이 더 많은지를 단순 비교하는 A/B 테스팅은 장점만큼 단점이 있다. 이 책에서 제시하는 밴디트 알고리즘은 확률에 기반한 기계학습 기법을 사용하기 때문에, A/B 테스팅에 비해 복잡하지만 좀 더 세밀하고 정확한 분석을 할 수 있다. 이 때문에 이 책에서 소개된 밴디트 알고리즘을 적용한다면 경영자와 개발자 모두에게 이점을 줄 수 있다. 경영자는 밴디트 알고리즘 분석 결과를 토대로 다른 경쟁 업체보다 한 차원 더 높은 고차원의 마케팅을 할 수 있기 때문에 더 높은 매출 신장을 기대할 수 있을 것이다. 개발자는 확률 기반의 기계학습 기법을 적용한 프로그래밍을 배울 수 있기 때문에 향후 빅데이터 분석 프로그래밍에 대한 역량을 가지게 될 것이다.

번역 분량이 그리 많지는 않았지만, 그래도 주로 주말에 작업하느라 가족들에게 무척 미안했었다. 나를 믿고 응원해준 사랑하는 아내와 주말에 제대로 놀아주지 않는다고 매일 투정하며 아빠를 괴롭혔던 세상에서 가장 예쁜 딸에게 먼저 감사의 인사를 전하고 싶다. 또한, 첫 번역임에도 항상 용기를 심어주고 이런 좋은 기회를 주신 편집자께도 감사의 인사를 전한다.

대상 독자 및 예제 파일

초급 **초중급** 중급 중고급 고급

이 책에서 사용한 예제 파일은 https://github.com/johnmyleswhite/BanditsBook 에서 받을 수 있습니다.

한빛 eBook 리얼타임

한빛 eBook 리얼타임은 IT 개발자를 위한 eBook입니다.

요즘 IT 업계에는 하루가 멀다 하고 수많은 기술이 나타나고 사라져 갑니다. 인터넷을 아무리 뒤져도 조금이나마 정리된 정보를 찾는 것도 쉽지 않습니다. 또한 잘 정리되어 책으로 나오기까지는 오랜 시간이 걸립니다. 어떻게 하면 조금이라도 더 유용한 정보를 빠르게 얻을 수 있을까요? 어떻게 하면 남보다 조금 더 빨리 경험하고 습득한 지식을 공유하고 발전시켜 나갈 수 있을까요? 세상에는 수많은 종이책이 있습니다. 그리고 그 종이책을 그대로 옮긴 전자책도 많습니다. 전자책에는 전자책에 적합한 콘텐츠와 전자책의 특성을 살린 형식이 있다고 생각합니다.

한빛이 지금 생각하고 추구하는, 개발자를 위한 리얼타임 전자책은 이렇습니다.

1. **eBook Only - 빠르게 변화하는 IT 기술에 대해 핵심적인 정보를 신속하게 제공합니다.**

   500페이지 가까운 분량의 잘 정리된 도서(종이책)가 아니라, 핵심적인 내용을 빠르게 전달하기 위해 조금은 거칠지만 100페이지 내외의 전자책 전용으로 개발한 서비스입니다. 독자에게는 새로운 정보를 빨리 얻을 수 있는 기회가 되고, 자신이 먼저 경험한 지식과 정보를 책으로 펴내고 싶지만 너무 바빠서 엄두를 못 내는 선배, 전문가, 고수 분에게는 보다 쉽게 집필할 수 있는 기회가 될 수 있으리라 생각합니다. 또한 새로운 정보와 지식을 빠르게 전달하기 위해 O'Reilly의 전자책 번역 서비스도 하고 있습니다.

2. **무료로 업데이트되는, 전자책 전용 서비스입니다.**

   종이책으로는 기술의 변화 속도를 따라잡기가 쉽지 않습니다. 책이 일정 분량 이상으로 집필되고 정리되어 나오는 동안 기술은 이미 변해 있습니다. 전자책으로 출간된 이후에도 버전 업을 통해 중요한 기술적 변화가 있거나 저자(역자)와 독자가 소통하면서 보완하여 발전된 노하우가 정리되면 구매하신 분께 무료로 업데이트해 드립니다.

3. 독자의 편의를 위해 DRM-Free로 제공합니다.

구매한 전자책을 다양한 IT 기기에서 자유롭게 활용할 수 있도록 DRM-Free PDF 포맷으로 제공합니다. 이는 독자 여러분과 한빛이 생각하고 추구하는 전자책을 만들어 나가기 위해 독자 여러분이 언제 어디서 어떤 기기를 사용하더라도 편리하게 전자책을 볼 수 있도록 하기 위함입니다.

4. 전자책 환경을 고려한 최적의 형태와 디자인에 담고자 노력했습니다.

종이책을 그대로 옮겨 놓아 가독성이 떨어지고 읽기 힘든 전자책이 아니라, 전자책의 환경에 가능한 한 최적화하여 쾌적한 경험을 드리고자 합니다. 링크 등의 기능을 적극적으로 이용할 수 있음은 물론이고 글자 크기나 행간, 여백 등을 전자책에 가장 최적화된 형태로 새롭게 디자인하였습니다.

앞으로도 독자 여러분의 충고에 귀 기울이며 지속해서 발전시켜 나가도록 하겠습니다.

지금 보시는 전자책에 소유권한을 표시한 문구가 없거나 타인의 소유권한을 표시한 문구가 있다면 위법하게 사용하고 있을 가능성이 높습니다. 이 경우 저작권법에 의해 불이익을 받으실 수 있습니다.

다양한 기기에 사용할 수 있습니다. 또한 한빛미디어 사이트에서 구입하신 후에는 횟수에 관계없이 내려받으실 수 있습니다.

한빛미디어 전자책은 인쇄, 검색, 복사하여 붙이기가 가능합니다.

전자책은 오탈자 교정이나 내용의 수정·보완이 이뤄지면 업데이트 관련 공지를 이메일로 알려드리며, 구매하신 전자책의 수정본은 무료로 내려받으실 수 있습니다.

이런 특별한 권한은 한빛미디어 사이트에서 구입하신 독자에게만 제공되며, 다른 사람에게 양도나 이전은 허락되지 않습니다.

차례

1 | 두 가지 특성, 탐색과 활용

본론에 들어가기 앞서서 소규모 웹 비즈니스를 운영하며 대부분의 소득을 얻고 있는 웹 활용자인 '데보라 널'Deborah Knull, 이하 '데비 널' 또는 '데비'에 대한 짧은 이야기를 해보겠다. '데비 널'의 이야기에는 '밴디트 알고리즘'을 학습할 때 떠오르는 '탐색과 활용'[01]이라는 부르는 핵심 개념이 나온다. 이 개념을 확실하게 잡기 위해 '탐색하는 과학자'와 '활용하는 사업가'라는 두 유형의 사람과 이 개념을 관련 지을 것이다. 이들 두 유형의 사람의 특성이 더 나은 웹 사이트를 구축하기 위해서 이 두 유형의 사람들이 가지고 있는 욕망 사이에서 균형 잡는 방법을 찾는 것이 왜 필요한지 이해하는 데 도움이 되기를 바란다.

1.1 과학자와 사업가

어느 일요일 오전, 젊은 인터넷 쇼핑몰 운영자인 데비 널은 사이트 로고의 기본 색상을 변경하면 사용자들이 좀 더 편안하게 느낄 거라고 생각을 했다. 여기서 중요한 점은 고객들이 더 편안하게 느낀다면 사이트에서 판매하는 제품들을 더 많이 구매할 것이라 데비 널이 생각했다는 것이다.

그러나 데비 널은 새로운 색상이 어쩌면 사용자들을 어리둥절하게 하여 불편하게 느끼지 않을까 걱정했다. 만약 이렇게 된다면 판매 증진을 위해 생각한 아이디어가 실제적으로는 오히려 고객들이 제품을 더 적게 구매하게 만들게 된다. 그녀는 자신의 직감을 확신하지 못했기 때문에 두 명의 친구들에게 조언을 구했다. 한 명은 과학자인 '신시아Cynthia'고 다른 한 명은 사업가인 '밥Bob'이다.

---

01 (역자주) 스탠퍼드대 제임스 마치(James March) 교수가 1991년 발표한, 경영학 전 분야에서 응용되는 논문이다. 기업의 혁신 활동에서 활용(Exploitation)과 탐색(Exploration)을 상충 관계(Trade-off)로 인식하고, 기업의 혁신 활동을 양자 간의 자원 배분 문제로 파악한다. Exploration은 "탐험", "개발" 등으로 번역되기도 하지만, 이 책에서는 경제학에서 주로 사용하는 "활용"으로 번역한다.

1.1.1 과학자 '신시아'의 조언

과학자인 신시아는 데비가 제시한 로고 변경에 대해 호의적이었다. 새로운 무언가를 시도할 수 있다는 기회에 흥분한 신시아는 데비에게 로고 변경을 신중하게 테스트할 수 있는 방법에 대해 강의하기 시작했다. "단순히 로고 색상을 바꿔선 안 돼. 로고 색상을 바꾸는 건 이후에 발생하는 결과에 대한 원인이 된다고 봐야 해. 통제된 실험을 실행할 필요가 있어. 아이디어를 통제된 실험으로 테스트하지 않으면 결코 색상 변화가 실제적으로 매출에 도움이 주었는지 악영향을 미쳤는지 알 수가 없게 돼. 어쨌든 조금 있으면 크리스마스 시즌이잖아. 지금 로고를 변경한다면 지난 두 달에 비해 큰 폭의 매출 증가를 보게 될 거라 확신해. 하지만 꼭 새로운 로고가 긍정적이라고 말할 수도 없어. 새로운 색상의 로고가 실제로는 매출에 악영향을 미칠 수도 있어."

"크리스마스는 새로운 색상으로 로고를 바꾸는 나쁜 결정을 했음에도 불구하고 매출 상승을 볼 수 있는 수익성이 좋은 연중행사야. 아이디어에 대한 실질적인 이점을 알기를 원한다면 제대로 된 비교를 적절하게 할 필요가 있어. 그리고 내가 아는 유일한 비교 방법은 전통적인 무작위 실험법이야. 신규 방문객이 사이트에 올 때마다 너는 동전을 던지는 거지. 앞면이 나오면 방문객을 그룹 A에 넣고 기존 로고를 보여주는 거야. 뒷면이 나오면 그룹 B에 넣고 새로운 로고를 보여주고. 방문객이 보는 로고는 완전히 무작위로 선택되기 때문에, 기존 로고와 신규 로고에 대한 비교를 왜곡시킬 수 있는 어떤 인자라도 시간이 지나면 균형을 잡게 될 거야. 동전 던지기를 가지고 방문객에게 어떤 로고를 보여줄지 결정한다면, 해당 로고에 대한 효과는 크리스마스 시즌과 같은 것들의 영향 때문에 왜곡되지는 않을 거야."

데비도 단순히 로고 색상만 바꾸고 싶지는 않다고 했다. 과학자 신시아가 제안한 것처럼 데비도 사이트 로고 변경에 대한 비즈니스적 가치를 평가할 수 있는 통제된 실험이 필요하다고 보았다.

신시아가 제안한 A/B 테스팅 환경에서 그룹 A와 B의 사용자들은 같은 웹 사이트에 접속하지만 미묘하게 약간 다른 버전의 사이트를 보게 된다. 충분히 많은 사용자에게 양쪽 디자인을 노출한 후, 데비는 두 그룹 간 비교를 통해 제안된 로고 수정안이 사이트에 정말 도움이 되는지 해가 되는지 판단할 수 있다.

A/B 테스팅의 장점을 확신하고 있는 신시아는 더 큰 규모의 실험을 고려하기 시작했다. A/B 테스트를 하는 대신, 기존 검은 로고 외에 보라색과 연두색 등과 같은 매우 독특한 색상까지 포함하여 총 6개의 색상을 함께 비교하는 실험을 말이다. 겨우 몇 분도 지나지도 않았는데 신시아는 A/B 테스팅에서 A/B/C/D/E/F/G 테스팅으로 생각을 바꿨다.

이 아이디어를 가지고 제대로 된 실험을 수행하는 것은 과학자 신시아를 흥분케 했지만, 데비는 신시아가 제안한 몇몇 색상이 현재 로고와 비교하면 훨씬 더 안 좋게 보이는 것은 아닌지 걱정했다. 무엇을 해야 할지 확신할 수 없어 데비는 큰 다국적 은행에서 근무하는 밥에게 근심을 털어놓았다.

1.1.2 사업가 '밥'의 조언

밥은 사이트에서 새로운 로고 색상을 실험할 것이라는 데비의 아이디어를 들고는 이 실험은 유익하다는 점에는 동의했다. 하지만 신시아의 별난 아이디어를 시도하는 것이 가치 있을지에 대해서는 매우 회의적이었다.

"신시아는 과학자야. 그러니 신시아는 네가 실험을 수없이 수행해야 한다고 생각할거야. 신시아는 지식을 위한 지식을 얻기를 원하지, 결코 실험 비용에 대해서는 생각하지 않아. 하지만 넌 사업가야, 데비. 먹고 살아야지. 사이트 수익을 극대화하도록 노력해야 해. 재무 상태를 제대로 유지하려면 수익성이 있을 만한 실험들만 해야 해. 지식은 비즈니스에서 이익을 얻을 수 있을 때만 가치가 있어. 네가 진정 변경이 가치가 있을 가능성이 있다고 믿지 않는다면 절대로 수행하지 마. 그리고

새로운 아이디어가 없다면 기존 로고를 그대로 유지하는 게 최고의 전략이야."

큰 규모 실험의 가치에 대한 밥의 회의론은 데비가 초기에 가졌던 우려를 다시 불러일으켰다. 고객을 잃는다는 위협은 신시아의 디자인 실험을 행한 열정에 의해 고양되었을 때 데비가 느꼈던 것보다 훨씬 컸다. 데비는 실험을 하지 않은 채 어떤 변경이 이익이 되는지 어떻게 결정할 수 있는지 명확하지 않았다. 이런 데비의 생각은 기존 로고에 대한 밥의 선호를 무시하고 신시아의 원래 제안을 따라야 하는지 고민하게 만들었다.

신시아와 밥의 의견을 저울질하는 데 여러 시간을 보낸 후에야 데비는 신시아와 밥에게 동기를 부여한 목표 사이에는 근본적인 트레이드오프trade-off이 있을 거라고 생각했다. 소규모 비즈니스에서는 과학자처럼 행동하고 지식을 위한 지식을 모으기 위해 비용을 지출할 여유가 없지만, 그렇다고 현재 이익에 근시안적으로 집중하고 새로운 어떤 아이디어를 아예 실험해보지 않을 수도 없다. 지금까지 본 바와 같이, 데비는 결코 (1) 새로운 것을 배우는 것과 (2) 이미 학습했던 것에서 이익을 얻기 위한 필요 간에 균형을 맞출 수 있는 간단한 길은 없다고 느꼈다.

1.1.3 경영과학 연구원 '오스카'의 조언

운이 좋게도 데비는 조언을 해줄 수 있는 또 한 명의 친구가 있다. 바로 경영과학 연구소의 지역 부서에서 일하고 있는 교수, '오스카'다. 데비는 오스카가 비즈니스 의사 결정에서 저명한 전문가인 것을 알고 있었고, 오스카가 '이익 극대화'와 '실험' 사이에서 균형을 맞추는 것에 대해 답변을 해줄 지식을 가지고 있을 거라 생각했다.

그리고 오스카는 데비의 아이디어에 확실히 흥미가 있었다

"네가 실험에 대한 신시아의 관심과 이익에 대한 밥의 관심 사이에서 균형을 맞출 수 있는 방법을 찾아야만 한다는 점에 대해 동의해. 동료들과 나는 이걸 '탐색-활

용 트레이드오프Explore-Exploit trade-off'이라고 불러."

"그게 뭔데."

"경영과학 연구원들이 이익 극대화와 실험 사이에 균형을 맞춰야 하는 너의 필요를 지칭하는 말이야. 우리는 '실험'을 '탐색exploration'이라 부르고 '이익 극대화'는 '활용exploitation'이라 불러. 사람, 회사 또는 로봇이든 간에 이익 추구 시스템이 균형 맞출 방법을 찾아야만 하는 근본적인 가치들이야. 네가 탐색을 너무 많이 한다면 금전적 손실이 발생해. 반대로 활용만 너무 많이 한다면 너의 사업은 정체되고 새로운 기회를 놓치게 돼."

"그렇담 어떻게 탐색과 활용 간의 균형을 잡을 수 있지?"

"불행히도 간단하게 답변할 수는 없어. 네가 추론하는 것처럼, 두 개의 목표 사이의 균형을 맞출 수 있는 보편적 해법은 없어. 아이디어가 좋은지 나쁜지를 알기 위해선 돈도 잃고 거의 이익을 얻지 못할 수 있다는 위험을 감수하고 탐색을 해야 해. 새로운 아이디어를 탐색하는 것과 기존 아이디어들 중 최선을 활용하는 것 중에 하나를 선택하는 올바른 방법은 네가 처한 상황의 세부 환경에 따라 달라져. 내가 너에게 말할 수 있는 건, 네가 어떤 로고 색상이 최선인지 학습할 수 있는 유일한 방법으로 신시아와 밥 둘 다 당연하게 여기는, A/B 테스팅을 수행하려는 너의 계획이 항상 최선의 선택은 아니라는 거야."

"예를 들어, 크리스마스 시즌과 올해의 남은 기간 동안 계속 운영되는 최고의 디자인이 확실히 있고, 그 디자인을 엄격하게 적용한다면, A/B 테스팅의 시험 기간은 의미가 있어. 그러나 최고의 색상 조합이 핼러윈 시즌에는 흑색/오렌지색이고 크리스마스 시즌에서는 적색/녹색으로 확연히 갈린다고 상상해봐. 만일 이 두 시즌 중 한 시즌에서 이 색상 조합들 가지고 A/B 실험을 한다면 두 대조군 사이에서는 매우 큰 차이가 벌어지게 될 거야. 그리고 색상들은 특정 기간에만 최고의 조합이

기 때문에 테스팅한 시즌을 제외한 다른 기간의 매출은 급격히 떨어질 거야."

"그리고 순수 A/B 실험에는 다른 잠재적인 문제가 있어. 네가 크리스마스 시즌과 핼러윈 시즌 양쪽을 포함하여 실험한다고 가정해 보자. 시즌별로 색상 조합들이 분리된 것을 검증하고 각 시즌에 매우 큰 효과를 내는 색상 조합들이 각각 있는 것을 알 수 있을 거야. 하지만 그럼에도 불구하고 시즌 별로 매출을 증진시키는 색상 조합이 확연히 갈리기 때문에, 이 두 개 색상 조합에 대한 평균 효과를 기대할 수 없을 거야. 너는 의미 있는 실험을 설계할 컨텍스트context가 필요해. 지능적으로 실험할 필요가 있어. 고맙게도 네가 더 나은 실험을 설계하는 것에 도울 수 있는 수많은 알고리즘이 있어."

1.2 탐색-활용 딜레마

이 짧은 이야기가 여러분이 웹 사이트를 최적화하려 할 때 필요한, 완벽하게 다른 두 가지 목표가 있다는 사실을 명확하게 여러분에게 이해시켜줄 것이다. 여러분은 (A) (지금부터 계속 '탐색'이라 부를) 새로운 아이디어를 학습할 필요가 있고, 반면에 (B) 기존 아이디어 중 (지금부터 '활용'이라 부를) 최선의 아이디어의 장점을 취해야만 한다. 과학자 신시아는 탐색을 의미한다. 신시아는 보라색 또는 연두색의 사용과 같이 끔찍한 아이디어를 포함해서 새로운 모든 아이디어에 개방적이었다. 밥은 활용을 의미한다. 밥은 미성숙한 새로운 아이디어에 대해 폐쇄적이었고 기존 것에 과도하게 집착하려 했다.

더 나은 웹 사이트의 구축을 돕기 위해, 필자는 오스카가 데비를 돕기 위해 하려 했던 바로 그것을 할 것이다. '탐색-활용 딜레마Explore-Exploit dilemma'를 풀기 위한 방법에 대해 집중 학습을 제공할 것이다(두 개의 전통적 알고리즘과 하나의 최신 기술 알고리즘에 대해 토론할 것이다). 그리고 '탐색-활용 트레이드오프Exploration-Exploitation trade-off' 주변에서 야기된 거대한 분야에 관한 정보를 더 많이 얻기 위해서는 일반적인 교과서를 참조하면 된다.

그러나 '탐색-활용 트레이드오프'을 풀기 위한 알고리즘 학습을 시작하기 전에, 이 책에서 제시하는 밴디트 알고리즘들과 대다수 웹 개발자들이 새로운 아이디어를 탐색하기 위해 사용하는 전통적인 A/B 실험 방법 간의 차이점에 초점을 맞출 것이다.

2 | 왜 멀티암드 밴디트 알고리즘을 사용하는가

2.1 우리가 시도해볼 것은 무엇인가?

이전 장에서 탐색과 활용이라는 두 가지 핵심 개념을 소개했다. 이 장에서는 웹 사이트 최적화라는 특정 상황을 가지고 탐색과 활용 개념을 보다 확실하게 이해하는데 중점을 두려 한다. 여기서 '웹 사이트 최적화'는 웹 개발자가 웹 사이트에 대한 일련의 변경을 수행하는 단계적 절차를 말한다. 그리고 해당 웹 사이트의 성공success을 증가하도록 각각의 절차가 수립된다. 수많은 웹 개발자들에게 가장 유명한 웹 사이트 최적화 유형은 '검색 엔진 최적화'(간략하게 SEO)다. 검색 엔진 최적화는 검색 엔진 결과에서 사이트의 랭크를 상승시키기 위해 웹 사이트 수정 등을 포함한 절차다. 이 책에서는 SEO를 전혀 다루지 않는다. 그러나 이 책에서 설명하는 알고리즘들은 어느 SEO 기술이 해당 사이트에 가장 적합한지를 결정하기 위한 SEO 개선 방안의 일부로 쉽게 적용이 가능하다.

SEO 또는 웹 사이트의 성공을 높이기 위한 다른 특정 수정 방안에 초점을 맞추는 대신, 웹 사이트(들)에 적용한 수정 사항의 실제 가치를 측정할 수 있는 여러 알고리즘을 설명할 것이다.

그러나 이들 알고리즘을 설명하기 전에 '성공success'의 개념을 명확히 정의해보자. 지금 이 순간부터는 아래와 같은 측정 성과를 설명할 때에만 '성공'이라는 용어를 사용하도록 하자.

트래픽 Traffic

변경 사항은 웹 사이트의 랜딩 페이지[01]에서 트래픽을 증가시켰는가?

01 (역자주) 랜딩 페이지(landing page). 웹 사이트 방문자가 처음 접하는 페이지. 링크, 검색어, 광고 등으로 들어온 사용자(즉, 잠재고객)가 클릭한 후 처음으로 대면하는 페이지를 말한다.

전환 Conversions

변경 사항은 성공적으로 재방문 고객이 된 신규 방문자의 수를 증가시켰는가?

판매 Sales

변경 사항은 신규 고객 또는 기존 고객의 구매 횟수를 증가시켰는가?

CTR Click Through Ratio[02]

변경 사항은 방문자들의 광고 클릭 횟수를 증가시켰는가?

성공에 대한 명확하고 정량적인 측정 외에도 웹 사이트의 성공을 증가시킬 것이라고 예상하는 잠재적인 변경 목록을 가지고 있을 필요가 있다. 지금부터는 성공의 측정을 '보상reward'으로, 잠재적 변경 목록을 '암arm'이라 부르기로 한다. 이들 용어의 역사적 근거에 대해서는 짧게 설명할 것이다. 필자 개인적으로는 알맞은 용어는 아니라고 생각하지만, 관련 주제에 대한 학술 문헌에서 절대적인 표준용어로 사용하고 있고 알고리즘에 대한 논의를 명확하게 할 수 있도록 도와준다.

지금부터는 다른 주제로 관심을 돌려보자. 웹 사이트 최적화를 할 때 새로운 아이디어를 시험하기 위해 왜 귀찮게 밴디트 알고리즘을 사용하는가? A/B 테스팅만으로 충분하지 않은가?

이 질문에 대답하기 위해 일부 세부 사항에 대해 전형적인 A/B 테스팅 설정을 설명하고 왜 이것이 이상적이지 않은지 이유를 말해보겠다.

2.2 비즈니스 과학자: 웹 규모 A/B 테스팅

대부분의 대형 웹 사이트는 이미 새로운 아이디어를 테스트하는 방법을 많이 알고 있다. 데브 널에 관한 짧은 이야기에서 설명한 것처럼, 이들은 새로운 아이디어가

02 (역자주) 인터넷상에서 배너 하나가 노출될 때 클릭되는 횟수를 뜻한다. 보통은 '클릭률'이라고 한다. 예컨대 특정 배너가 1백 번 노출됐을 때 3번 클릭된다면 CTR는 3%가 되는데, 일반적으로 1~1.5%가 광고를 할만한 수치다.

제대로 작동하는지는 통제된 시험 수행을 통해서만 결정할 수 있다는 것을 이해하고 있다.

이런 통제된 시험 유형은 A/B 테스팅이라 부른다. 왜냐하면, 일반적으로 이 테스팅은 접속하는 웹 사용자를 무작위로 그룹 A와 그룹 B의 2개 중 하나에 할당하기 때문이다. 이러한 사용자의 무작위 할당을 웹 개발자가 A 방안이 B 방안보다 더 성공적이라거나 또는 반대의 경우인 것을 확신할 때까지 유지한다. 그 후에 웹 개발자는 모든 미래의 사용자들을 좀 더 성공적인 버전에 할당하고 열등한 버전은 닫아버린다.

신규 아이디어를 시도하기 위한 이런 실험 방식은 과거에는 매우 성공적이었고 앞으로도 수많은 상황에서 계속 성공할 것이다. 그런데 밴디트 알고리즘이 우리에게 무언가를 제공한다는 사실을 왜 믿어야 하는가?

이 질문에 대한 제대로 된 답변을 하기 위해 다시 탐색과 활용의 개념으로 돌아가 보도록 하자. 표준 A/B 테스팅은 다음과 같이 구성된다

- 오직 탐색만을 수행하기 위한 짧은 기간. 이 기간에 동일한 수의 사용자들을 그룹 A와 B에 할당한다.
- 오직 활용만을 수행하기 위한 긴 기간. 이 기간에 모든 사용자들을 좀 더 성공적인 버전으로 보내고, 열등한 것처럼 보이는 방안은 다시는 사용하지 않는다.

왜 이것이 나쁜 전략일까?

- 둘 사이에 부드러운 이전이 가능함에도 불구하고, 탐색에서 활용으로 딱딱 끊어져 점프한다.
- 오직 탐색만 수행하는 단계 동안, 가능한 한 많은 데이터들을 수집하기 위해

열등한 방안을 탐색하는 데 자원을 낭비한다. 그러나 눈에 띄게 열등한 방안에 대한 데이터를 수집하는 건 원하지 않는다.

밴디트 알고리즘은 이들 문제에 대한 해법을 제공한다. (1) 밴디트 알고리즘은 탐색의 양을 갑작스럽게 감소시키는 대신 천천히, 부드럽게 감소시킨다. 그리고 (2) 일반적인 A/B 테스팅에서 열등한 옵션을 과잉 탐색하여 시간을 허비하는 대신 더 나은 옵션을 탐색 과정 동안에는 자원에 초점을 맞추도록 해준다. 사실, 밴디트 알고리즘들은 시간이 지남에 따라 최상의 가용 옵션에 점차 고정되기 때문에 위의 두 문제를 다루는 밴디트 알고리즘은 동일하다. 학술 문헌에서는 최상의 가용 옵션을 설정하는 이러한 절차를 '수렴convergence'이라 한다. 모든 좋은 밴디트 알고리즘들은 결국 수렴한다.

실제로, 이들 두 유형의 개선이 비즈니스에 얼마나 중요할지는 비즈니스 운용 방식의 세세한 사항에 많이 달려있다. 그러나 밴디트 알고리즘이 제공하는 탐색과 활용을 고려한 일반적인 프레임워크는 어떤 일에서든 유용할 것이다. 왜냐하면, 밴디트 알고리즘은 A/B 테스팅도 특별한 경우로 포함하기 때문이다. 표준 A/B 테스팅은 순수 탐색과 순수 활용의 극단적인 두 가지 경우만 사용하는 경우를 말한다. 밴디트 알고리즘은 이들 두 극단적인 상태 사이의 넓고 흥미로운 공간 내에서 동작한다.

밴디트 알고리즘이 어떻게 균형을 이루는지 '엡실론-그리디epsilon-Greedy 알고리즘'을 첫 번째로 살펴보자.

3 | 엡실론-그리디 알고리즘

3.1 엡실론-그리디 알고리즘 소개

탐색-활용 딜레마에 대해 알고리즘적 사고를 시작하기 위해, 탐색과 활용 사이에 트레이드오프가 가능한 가장 간단한 알고리즘을 코딩하는 법을 소개하려 한다. 이 알고리즘은 엡실론-그리디 알고리즘이라 한다. 컴퓨터 과학 분야에서 그리디(탐욕) 알고리즘은, 비록 그 결정이 장기적으로는 나쁜 결과를 초래할지라도, 항상 현재 상태에서 가장 최선의 결과로 보이는 행동을 취하려 한다. 엡실론-그리디 알고리즘은 일반적으로 최선의 가용한 옵션을 활용하기 때문에 그리디 알고리즘과 매우 비슷하다. 그러나 가끔 엡실론-그리디 알고리즘은 다른 가용한 옵션을 탐색하는 점이 다르다. 보이는 것처럼, 알고리즘의 이름에 있는 엡실론이란 용어는 활용 대신 탐색하는 확률을 의미한다.

좀 더 자세히 살펴보면, 엡실론-그리디 알고리즘은 순전히 무작위적인 실험에 대한 신시아의 이상과 이익을 극대화하려는 밥의 본능 사이에서 무작위적으로 왔다 갔다하며 수행된다.

이익을 높이기 위해 웹 사이트 로고 색상을 변경하는 앞장의 예제를 가지고 아이디어를 더 구체화해보자. 데브는 녹색과 적색의 두 색상 사이에서 숙고 중이고, 그녀는 신규 방문자가 회원 가입할 확률을 최대화하는 색상을 찾기를 원한다고 가정한다. 엡실론-그리디 알고리즘은 각 신규 잠재 고객들에게 순차적으로 적용되는 다음의 절차(그림 3-1 참조)를 사용하여 최적의 색상 로고를 찾으려 시도한다.

- 신규 방문자가 사이트에 오면, 알고리즘은 엡실론의 확률로 뒷면이 나오는 동전을 던진다. (확률 X의 의미는 어떤 일이 100 * X 퍼센트만큼 발생하는 것이

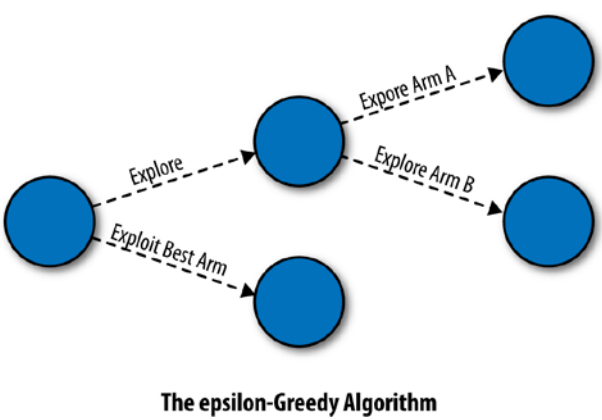
다. 0.01의 확률로 동전이 뒷면이 나온다는 말은 뒷면이 나올 확률이 1%라는 것이다.)

- 동전이 앞면이 나온다면, 알고리즘은 활용을 수행한다. 활용을 위해, 알고리즘은 이력을 기록하는데 사용하는 데이터 소스에서 녹색과 적색 로고 양쪽에 대한 변환 비율 이력을 찾아본다. 과거에 어떤 색상이 더 높은 성공 비율을 가졌는지 결정한 후에, 알고리즘은 신규 방문자에게 가장 성공적인 이력을 가진 색상을 보여주기로 정한다.

- 동전이 앞면 대신 뒷면이 나왔다면, 알고리즘은 탐색을 수행한다. 탐색은 고려되는 두 색상에 대한 무작위적인 실험을 포함하기 때문에, 알고리즘은 그들 중 선택하기 위한 두 번째 동전을 던질 필요가 있다. 첫 번째 동전과 달리, 두 번째 동전에서 앞면이 나올 확률은 50%라고 가정한다. 일단 두 번째 동전을 던지고 나면, 알고리즘은 이 절차의 마지막 단계로 이동할 수 있다.

  - 두 번째 동전이 앞면이 나온다면, 신규 방문자에게 녹색 로고를 보여준다.
  - 두 번째 동전이 뒷면이 나온다면, 신규 방문자에게 적색 로고를 보여준다.

그림 3-1. 엡실론-그리디 알고리즘 선택 절차



이 알고리즘은 한동안 신규 사용자들을 사이트에 늘어놓은 후에, (A) 현재 알려진 최선의 옵션을 수행하는 것과 (B) 가용한 모든 옵션 사이에서 임의로 탐색하는 것 사이에서 왔다 갔다하며 수행한다.

- (1 - 엡실론)의 확률로, 엡실론-그리디 알고리즘은 알려진 최선의 옵션을 활용한다.
- (엡실론 / 2)의 확률로, 엡실론-그리디 알고리즘은 알려진 최선의 옵션을 탐색한다.
- (엡실론 / 20의 확률로, 엡실론-그리디 알고리즘은 알려진 최악의 옵션을 탐색한다.

이것이 엡실론-그리디 알고리즘의 전부이다. 지금까지 설명된 내용에서 빼먹은 중요한 개념은 없다. 이제부터는 실제 사이트에서 이 알고리즘을 어떻게 적용하는지 명확하게 하기 위해 파이썬으로 알고리즘을 구현할 것이다. 다음 장에서는 다른 시나리오들에서 이 알고리즘이 어떻게 동작하는지에 대한 직관력 개발에 도움이 될 엡실론-그리디 알고리즘용 단위 테스트 프레임워크를 구축할 것이다.

3.2 로고 선택 문제의 추상화

3.2.1 암은 무엇인가?

엡실론-그리디 알고리즘에 대한 코드를 작성하기 전에, 녹색 로고와 적색 로고를 비교하기 원하는 예제를 추상화하기 위해, 이 책의 잔여 부분에서 사용할 전문 용어 2개를 소개한다.

첫 번째로, 골라야 하는 색상이 단지 두 개가 아닌 수백 개 또는 수천 개일 때의 확률을 고려한다. 일반적으로, N개의 고정된 다른 옵션 세트를 가지고 이들을 열거할 수 있어서 녹색 로고는 옵션1, 적색 로고는 옵션2, 그리고 다른 색상 로고는 옵

션 N이라 칭할 수 있다고 가정한다. 역사적인 이유로 인해, 이들 옵션들은 전통적으로 암을 의미한다. 그래서 옵션 1, 옵션 2, 옵션 N 대신 암 1, 암 2, 암 N으로 표현한다. 그러나 핵심 개념은 선택한 단어에 상관없이 동일하다.

왜 옵션이 전통적으로 암이라 불리는지 설명하는 것이 이 책에 나오는 전문 용어들을 이해하는 데 도움이 될 것이다. 암이라는 이 용어는 이 책에서 설명하는 알고리즘 디자인의 이면에 있는 근본 동기를 나타낸다. 이들 알고리즘들은 원래 이상적인 도박사가 가상의 카지노에서 가능한 많은 돈을 따는 법을 설명하기 위해 발명되었다. 이 가상적인 카지노에는, 오직 슬롯머신이라는 한 가지 종류의 게임만 있고, 이것은 때로는 고객의 돈을 잃게 하는 머신의 성향 때문에 원-암드 밴디트(단일 팔을 가진 도적)라 불렸다. 이 카지노는 슬롯머신만 있었지만, 각각 다른 배당 스케줄을 가진 많은 종류의 슬롯머신이 있었기 때문에 계속 방문하기에 흥미로운 곳이었다.

예를 들어, 이 가상 카지노의 슬롯머신 일부는 100번 레버를 당기면 그 중 한 차례 5불을 지불한다. 반면 다른 머신은 1000번 레버를 당기면 한 차례 25불을 지불한다. 어떤 이유에서든지, 원래의 수학자들은 그들의 사고 실험에서 서로 다른 슬롯머신들을 마치 수많은 암을 가진 한대의 거대한 슬롯머신으로 취급하기로 결정했다. 이럼으로써 그들의 문제에서 옵션들을 암으로서 참조할 수 있게 되었다. 또한 이런 사고 실험을 멀티암드 밴디트 문제라 부르게 되었다. 오늘날, 우리는 여전히 이런 알고리즘들을 밴디트 알고리즘이라 부르고 있고, 그래서 역사적인 명명 방식을 아는 것이 왜 옵션을 암으로 나타내는지 아는 데 도움을 준다.

3.2.2 보상은 무엇인가?

지금까지 암이 무엇인지 설명하여 엡실론-알고리즘의 추상화 준비의 절반을 마쳤다. 다음으로 보상을 정의해보자. 보상은 단순히 성공의 측정이다. 고객이 광고를 클릭하거나 사용자로 등록하는 것을 말한다. 중요한 것은 단순하게 (A) 보상은 수학적으로 기록할 수 있는 정량적이라는 것과 (B) 보상은 적은 것보다 많은 것이 좋

다는 것이다.

3.2.3 밴디트 문제는 무엇인가?

지금까지 암과 보상에 대해 정의했고, 이제 이 책에서 구현할 모든 알고리즘들의 동기가 되는 밴디트 문제의 추상적인 개념을 설명할 수 있다.

- 당길 수 있는 N개의 암 세트를 가진 밴디트라 불리는 복잡한 슬롯머신을 마주하고 있다.
- 레버를 당기면 어떤 주어진 암은 보상을 줄 것이다. 그러나 이들 보상들은 신뢰할 수 없기 때문에, 이것이 도박이라 불리는 이유다. 암 1은 단지 1%의 보상 1단위를 줄 것이고, 반면 암 2는 단지 3%의 보상 1단위를 줄 것이다. 어떤 특정 암을 어떤 특정한 때에 당기는 것은 위험하다.
- 언제 암을 당기는 것이 위험하다는 것뿐만 아니라, 암들의 보상 비율도 모르는 채로 시작할 수밖에 없다. 실험적으로 모르는 암을 실제로 당겨서 이 비율을 가늠해야만 한다.

지금까지 기술한 문제는 단지 통계학에서 문제일 뿐이다. 어떤 암이 가장 높은 평균 보상을 가지는지 계산하여 위험에 대처해야 할 필요가 있다. 각 암을 매우 많이 당기고 되돌려 받은 보상의 평균값을 연산해서 평균 보상을 계산할 수 있다. 그러나 실제 밴디트 문제는 더욱 복잡하고 또한 더욱 현실적이다.

밴디트 문제를 특별하게 만드는 것은 각 암으로부터 보상에 대한 단지 적은 양의 정보만을 받는 것이다.

- 단지 실제로 당기는 암에 의해 주어지는 보상에 관해서만 찾을 수 있다. 어떤 암을 당기든 간에, 당기지 않는 다른 암에 대한 정보는 놓칠 수밖에 없다. 현실 세계와 비슷하게, 취했던 경로에 대해서만 학습할 수 있고 취하지 않았던 경