

Mathematical Data Science HW1

20180617 You SeungWoo

September 9, 2023

Problem 1

”너도밤나무집의 비밀”에서,
홈즈가 얻은 데이터:

- 루캐슬은 헌터에게 평균 2배 이상의 연봉을 제안했다. 단, 몇 가지 괴상한 조건도 같이 제시했다.
 - 밝은 파란색 옷을 입어야 한다.
 - 응접실에서 바깥을 등진 채로 창가의 의자에 한 시간 정도 앉아있다.
 - 머리를 짧게 잘라야 한다.
- 루캐슬의 전 부인은 사망했고, 루캐슬은 재혼했다. 부인은 많아야 서른, 루캐슬은 40대 중반이다.
- 앨리스는 루캐슬과 전 부인의 딸이며, 필라델피아에 있다고 한다.
- 루캐슬은 헌터가 창가에 앉아있는 동안 재미있는 얘기를 해서 그녀를 웃긴다.
- 헌터는 창문 쪽으로 얼굴이 향하지 않도록 해야 한다.
 - 헌터는 거울로 뒤에 한 남자가 있는 것을 확인한다. 이를 눈치챈 루캐슬은 서둘러 그를 쫓아낸다.
 - 이후 그녀는 창가에 앉지 않고, 드레스도 입지 않았다.
- 루캐슬에게는 카를로라는 맹견이 있다. 매일 밤, 톨러가 카를로를 풀어 밤에 사람이 침입하지 못하도록 한다.
- 방 안의 서랍에 헌터의 자른 머리와 완전히 같은 머리카락 다발이 들어있었다.
- 집 안에는 사용하지 않는 건물이 있으며, 이 문은 항상 잠겨있다.
- 루캐슬은 그 건물에 출입한다. 건물 안에서는 사람의 발소리가 들린다.
- 루캐슬은 방의 존재를 숨기려는 것처럼 말하고 행동한다.

문제 해결의 알고리즘화:

1. 데이터를 종합해 각 인물별로 정리한다.
2. 각 인물의 연결 관계를 정리한다.
3. 여러 사물과 인물의 연결 관계를 정리한다.
4. 특별한 사건으로부터, 일반적인 가능성을 일차적으로 추정한다.
5. 추정한 가설이 앞서 정리한 관계에 부합하는지 확인한다. 그렇지 않으면 해당 가설을 기각한다.
6. 가설이 너무 많거나 적은 경우, 다시 새로운 데이터를 수집해 앞의 단계를 반복한다.

홈즈의 데이터 활용:

- 홈즈는 감정이 아닌 데이터에 기반한 추리를 진행한다. 데이터가 불완전하거나 불충분한 경우, 해당 연결 고리를 찾고자 더 많은 정보를 모은다. 또한 그들의 심리와 행동, 활동 시간까지 파악해 단서를 찾고 추리를 이어간다.
- 루캐슬은 많은 급료를 지급하면서까지 헌터를 고집했고, 헌터에게 마치 변장하는 것과 같이 과도한 외형적 변화를 요구했다. 이는 헌터와 비슷한 누군가의 대역으로 헌터를 이용하려는 목적이 강하다고 추정할 수 있다.
- 루캐슬의 말에 의하면 앨리스는 집에 없다. 하지만 빈 건물에 누군가 있고, 그 인물은 루캐슬과 어느정도 연결점이 있으며, 헌터와 동일한 머리카락도 발견되었다. 이를 통해 건물의 인물이 앨리스일 것이라 추론할 수 있다.
- 루캐슬은 헌터에게 창 밖을 보지 못하게 했다. 창 밖에는 어떤 남자가 그녀를 주시하고 있었다. 또한 루캐슬은 창가에서만 그녀를 웃기려고 시도한다. 이는 남자에게 그녀가 다른 사람이라는 점을 알아차리지 못하게 하면서, 그녀의 행복한 모습을 공개하려는 의도로 분석할 수 있다.
- 밤에 사나운 개를 풀어 누군가의 침입을 막는다. 보안이 중요하다는 뜻으로, 루캐슬에게 지켜야만 하는 것이 있음을 의미한다. 위의 내용을 종합했을 때, 그 남자로부터 앨리스를 지키는 행위인 것으로 연결해볼 수 있다.

홈즈가 왓슨에게 한 말의 의미:

- 진흙 없이 벽돌을 만들 수 없는 것처럼, 데이터 없이 판단을 내릴 수 없음을 의미한다. 홈즈는 헌터의 말을 듣고 정상적이지 않다고 생각했지만, 이를 판단할 수 있는 근거가 없기에 불안감 속에서 해당 발언을 했다.

Problem 2

The SIR model is the following system of quadratic ODEs:

$$\frac{dS}{dt} = -\beta SI, \tag{1}$$

$$\frac{dI}{dt} = \beta SI - \nu I, \tag{2}$$

$$\frac{dR}{dt} = -\nu I, \tag{3}$$

where the disease transmission rate $\beta > 0$ and the recovery rate $\nu > 0$ (or in other words, the duration of infection $D = 1/\nu$).

The bi-linear incidence term $\beta S I$ for the number of new infected individuals per unit time corresponds to homogeneous mixing of the infected and susceptible classes. The total population size should remain constant, and this easily follows from the SIR system: that the sum of the left hand sides of the three equations is the derivative of the total population size by sum of the right hand sides is zero. We denote the total population size by N . Since $R(t) = N - S(t) - I(t)$, the system can be reduced to a system of two ODEs: (1) and (2).

Suppose that each infected individual has κ contacts (each sufficient for transmission) per unit time and κ is independent of the population size. Then $\kappa S/N$ of these contacts are with susceptible individuals. If the fraction τ of adequate contacts result in transmission, then each infected individual infects $\kappa\tau S/N$ susceptible individuals per unit time. Thus $\beta = b/N$ where $b = \kappa\tau$. The parameter τ is called the *transmissibility* of the infectious disease.