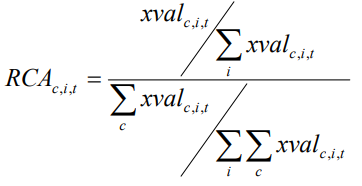
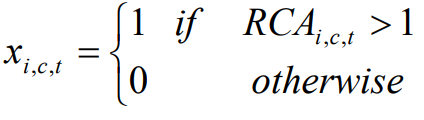
행렬을 활용한 계산비용 단축

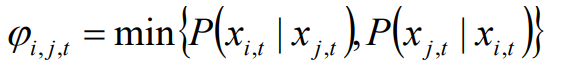
인하대학교 경제학과 김성운

4학년 학부연구생 활동을 하면서 무역상품 간 근접도를 계산하여, 한 국가의 산업지식 인프라를 측정하는 연구활동을 진행했습니다. 무역상품 간 근접도를 계산하는 과정에서 행렬의 구조적 특성과 논문의 확률개념을 결합하여 새로운 계산방법을 고안하였고 실제 계산비용을 대폭 절감하는 성과를 얻었습니다.

**1. 논문 리뷰**

상품간 근접도란 상품간 존재하는 이질성을 거리개념(Hausmann and Klinger, 2006)을 수치화한 것입니다. A라는 상품과 B라는 상품끼리 동질성이 높으면 1에 수렴(즉 A와 A의 거리는 1), 반대로 서로 이질적이면 0에 수렴합니다. 상품간 근접도는 어떤 나라가 A상품을 잘 만들고 있다면, A상품과 근접도가 가까운 상품 또한 잘 만들 것이라는 가정하에 출발합니다. 한 국가가 특정 상품에 특화되어 있는지 판단하기 위하여 경제학 개념인 비교우위 개념을 차용합니다. 비교우위를 판단하기 위해 RCA(Revealed Comparative Advantage, Balassa, 1986)를 사용합니다.

해당 식[[1]](#footnote-1)을 직관적으로 이해하자면 세계 전체 시장에서 해당 상품 무역액이 차지하는 비중보다 한 국가안에서 수출하는 해당상품의 무역 비중이 클 경우, 그 국가는 해당상품에 RCA가 있다라고 말할 수 있습니다.

논의를 간단히 하기 위해서 1을 기준으로 1보다 크면 1(비교우위를 가지고 있음.)을, 그렇지 않은 경우 0(비교우위를 가지고 있지않음.)으로 변환합니다

i 상품과 j 상품의 근접도는 조건부 확률을 통해 구합니다. 예를 들어 자동차와 트랙터가 있다고 생각해봅시다. 자동차와 트랙터의 경우 엔진, 미션, 타이어, 조향과 같이 유사한 생산투입요소를 공유하고 있습니다. 상품의 근접도란 상품 자체의 근접도이지 생산의 근접도가 아니기 때문에 단순히 교집합으로써 확률을 구할 수 없습니다. 자동차와 트랙터의 예시에서 자동차와 트랙터의 생산투입은 유사하지만 전혀 별개의 시장을 가지고 있습니다. 즉 생산에서 발생하는 유사성을 제거하기 위해서 조건부확률을 구하고 그 낮은 값을 차용하여 실제 상품간 근접도를 측정합니다. 예를 들어 자동차를 잘 만들었을 때 트랙터를 잘 만들 확률과 트랙터를 잘 만들었을 때 자동차를 잘 만들 확률을 비교하였을 때 낮은 값이 자동차와 트랙터사이의 근접도를 더 정확하게 나타낼 수 있습니다.[[2]](#footnote-2) 이를 확률공간으로 표현하면 다음과 같습니다.

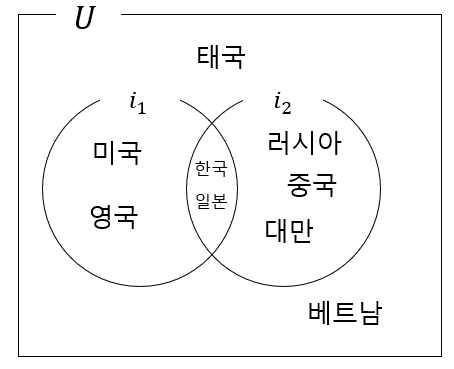
**<그림 1. 자동차 및 트랙터의 확률공간>**

그림 1 : = 전체 국가 집합, = 자동차를 비교우위(RCA >1)가지는 국가 집합, = 트랙터를 비교우위(RCA >1)가지는 국가 집합 이라고 정의하면 자동차에 비교우위가 있을 때 트랙터에도 비교우위가 있을 조건부 확률은 1/2이고, 트랙터에 비교우위가 있을 때 자동차에도 비교우위가 있을 조건부확률은 2/5이므로 자동차와 트랙터의 상품간 근접도는 2/5입니다.

**2. 계산과정 효율화**

상품간 근접도를 계산하는 과정에서 행렬을 이용하여 계산비용을 획기적으로 감소시켰습니다. 먼저 행은 국가, 열은 상품명으로 하는 행렬을 생성하고 각 국가(c)와 상품(i)에 대한 비교우위여부를 원소로 취하는 RCA행렬을 선언합니다.

이때 를 실시하게 되면 i\*i 정방 행렬이 도출됩니다. 이때 의 원소는 상품와 상품 모두 RCA를 가지는 국가 수로 정의됩니다.

은 대칭행렬이며 상품간 교집합 행렬로 표현됩니다. 이때 상품간 근접도를 구하기 위해서는 행렬의 대각성분으로 대각성분을 포함하는 열과 행의 원소를 나누어 주면 각 원소는 조건부 확률을 나타냅니다. 이때 조건부확률의 비대칭성을 확보하기 위해서는 각 대각성분을 중심으로 나누기를 실시할 때 이전 대각성분으로 나눈 값보다 작으면 이전 값을 삭제하고 새로운 값을 할당합니다. 대각성분을 포함하는 행, 열에 대해 대각성분으로 나누어 주고, 그 값이 이전 값보다 낮을 경우 교체하는 방법으로 대각벡터의 원소개수만큼 반복하게 되면 그 행렬이 바로 상품 간 근접도 행렬로 표현됩니다. 이를 의사코드로 작성하면 다음과 같습니다.

①. 대각성분 하나를 뽑는다.

②. 대각성분을 포함하는 행, 열을 대각성분으로 나누어준다.

③. 나눈 값을 새로운 행렬에 할당한다.

만약 이전값보다 높으면 할당하지 않는다.

④ 이를 대각성분만큼 반복한다.

⑤. 행렬을 출력한다.

이를 행렬로 표현하면 다음과 같습니다.

**3. 코드 결과**

R에서 제공하는 system.time()함수를 통해 계산시간을 측정하였습니다. 다음표는 각 계산법을 100번 반복수행하여 얻은 결과값을 정리하였습니다.

**<표 1. 각 계산법 계산시간 측정(단위 : 초)>**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Min** | **Mean** | **Max** |
| 기존 계산법 | 76.25 | 79.11 | 111.69 |
| 행렬 계산법 | 0.17 | 0.23 | 0.41 |

**4. 정리**

논문에서 사용한 확률개념과 이를 표현한 행렬사이의 연결고리를 파악하여 논문개념을 효율적으로 이해할 수 있었고 더불어 계산시간을 획기적으로 단축하여 원활한 코드수행에 도움이 많이 되었습니다.

Reference

1, B. Balassa, The Review of Economics and Statistics, 68, 315 (1986)

2. Ricardo Hausmann and Bailey Klinger, Structural Transformation and Patterns of Comparative Advantage in the Product Space, CID Working Paper No. 128, August 2006

1. c = 국가코드, i = 상품코드, t = 시간 [↑](#footnote-ref-1)
2. 특정상품을 생산하는 기술이 다른 상품생산에 영향을 미치는 것을 생산편의라고 정의하면, 생산편의는 항상 양의 값을 가지기 때문에(특정상품을 잘 만든다고 해서, 다른 상품을 못 만들게하는 경우는 매우 희박하므로) 작은 값을 상품간 근접도로 보는 것이 타당하다. [↑](#footnote-ref-2)