

## Q-learning 강화학습을 이용한 가격 예측에 관한 연구

A Study on the price prediction using Q-learning

---

저자 (Authors)	임희주, 이주현 Heeju Lim, Joohyun Lee
출처 (Source)	<a href="#">한국통신학회 학술대회논문집</a> , 2019.6, 1019-1020(2 pages) <a href="#">Proceedings of Symposium of the Korean Institute of communications and Information Sciences</a> , 2019.6, 1019-1020(2 pages)
발행처 (Publisher)	<a href="#">한국통신학회</a> Korea Institute Of Communication Sciences
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09234682">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09234682</a>
APA Style	임희주, 이주현 (2019). Q-learning 강화학습을 이용한 가격 예측에 관한 연구. 한국통신학회 학술대회논문집, 1019-1020
이용정보 (Accessed)	한국외국어대학교 203.253.93.*** 2021/04/02 13:32 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# Q-learning 강화학습을 이용한 가격 예측에 관한 연구

임희주, 이주현\*  
한양대학교 \*한양대학교

heju0509@hanyang.ac.kr, \*joohyunlee@hanyang.ac.kr

## A Study on the price prediction using Q-learning

Heeju Lim\*, Joohyun Lee\*

\*Division of Electrical Engineering, Hanyang Univ.

### 요 약

본 논문은 강화학습을 통해 2013 년도부터 2017 년도 까지의 식재료 가격을 이용하여 올해의 가격을 예측하도록 agent 에 학습시켰고, 2018 년도 식재료 자료를 이용하여 학습한 agent 가 얼마나 정확하게 학습했는지 확인하였으며 그 유용성을 발전시키는 방향에 대하여 논의하였다.

### I. 서 론

학생이나 직장인이 원하는 음식을 음식점에서 식사를 사 먹을 때와는 달리 가정에서의 식사는 조리하는 이의 선택에 전적으로 달려있다. 보통의 주부들이 음식을 하는데 있어서 가족 구성원들의 선호도 또한 중요한 고려대상이지만 식비라는 금전적인 부분도 무시할 수 없다. 실제 주부들이 식재료를 구매하는데 있어서 다른 재료보다 할인되는 재료를 구매하려고 하며 이를 이용한 음식을 준비하는 경우가 많다. 즉 메뉴를 정하는데 있어서 메뉴의 식재료 가격이 영향을 미치는 것인데 본논문에서는 그러한 식비에 따른 메뉴 선택에 도움을 주고자 재료에 따른 음식 가격을 예측하고자 하였다.

### II. 본론

본논문에서는 이를 위해 서울시 홈페이지에 올라와 있는 롯데마트 서울역 지점의 식료품 가격 동향을 데이터 셋으로 사용하였다

다음은 고등어 가격으로 확인한 2017 년과 2018 년 가격 변동 그래프이다.

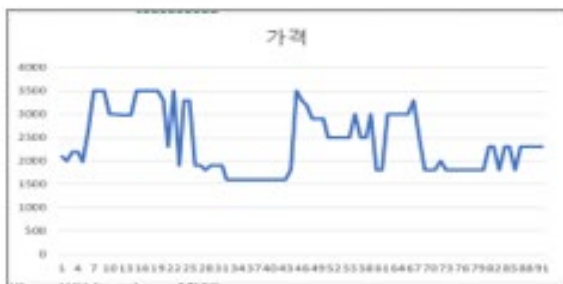


그림 2-1 2017 고등어 가격 변동 그래프

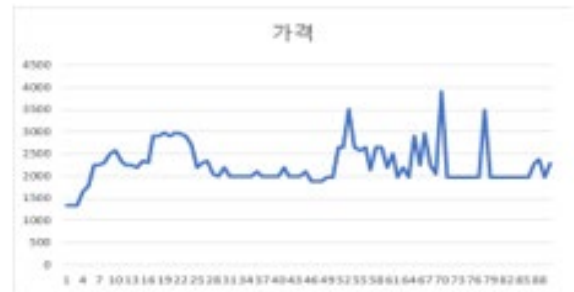


그림 2-2 2018 고등어 가격 변동 그래프

두 가격 변동 그래프를 확인하면 대략적으로 비슷한 시기에 가격의 상승과 하락이 있는 것을 볼 수 있다. 데이터셋으로 참고한 2013 년부터 2017 년까지의 자료를 토대로 이러한 경향성을 학습시켜서 연 평균가격과 비교하여 선택한 날짜의 가격이 평균가격보다 높을 것인지 낮을 것인지 예측할 수 있을 것으로 판단하였다.

학습은 Q-Learning 알고리즘 [2,3]을 2013 ~ 2017 년도 식재료 가격을 트레이닝 셋으로 하여 활용하였다.

```
Initialize  $Q(s, a), \forall s \in S, a \in A(s)$ , arbitrarily, and  $Q(\text{terminal-state}, \cdot) = 0$ 
Repeat (for each episode):
  Initialize  $S$ 
  Repeat (for each step of episode):
    Choose  $A$  from  $S$  using policy derived from  $Q$  (e.g.,  $\epsilon$ -greedy)
    Take action  $A$ , observe  $R, S'$ 
     $Q(S, A) \leftarrow Q(S, A) + \alpha [R + \gamma \max_a Q(S', a) - Q(S, A)]$ 
     $S \leftarrow S'$ 
  until  $S$  is terminal
```

다음과 같은 본연의 Q-Learning 의 알고리즘에서 State 는 시간  $t$  를 날짜로 두어 모든 날짜를 1 일부터 365 일까지로 두었고 state 는 for 문으로 카운트하여 시간이 지나면 next state 로 넘어가도록 하였다. 각 state 에서는 Low, Middle, High 의 3 가지 action 이 가능하며 이것은 연 평균가격과 비교하여 10%이내의 값일 경우에는 Middle, 그보다 클 경우에는 High, 작은 값이면 Low 를 의미한다. 13 년부터 17 년까지의 데이터셋으로 평균 가격과 비교하여 High, Middle, Low 3 가지로 분류한 보상테이블을 만든 후 action 이

보상테이블의 값과 같을 때는 1 값을, 그렇지 않을 때는 0 값을 반환하도록 하였다. 그리하여 Q-learning 알고리즘에서 reward 를 1 또는 0 을 받도록 하여 Q 테이블의 값을 업데이트 하도록 하였다. 이때 learning rate 를 현재에 가까워질 때마다 높게 설정하여 먼 과거의 가격변동 추이보다는 가까운 과거에서의 가격 변화에 영향을 더 많이 받도록 하였다. 탐험과 탐사를 적절히 사용하기 위해 E-greedy 알고리즘을 사용하여 첫번째 action 은 랜덤하게 선택하고 그 이후의 action 에 대해서는 q 테이블의 값이 큰 값을 선택하도록 했다.

각 년도에 따른 episode 의 수는 1000 으로 하여 학습시켰고 트레이닝셋으로 학습이 끝난 후에는 최종적으로 선택되는 action 들이 알고리즘이 선택한 최적의 action 이라 가정하고 이를 2018 년 자료를 테스트 셋으로 이용하여 학습이 잘 되었는지 판별했다.

### III. 결론

최적으로 학습된 action 들과 테스트 셋인 2018 년 롯데마트 서울역점 물가와 비교한 결과 45%에서 55%의 정확도를 얻었다. 이는 물가의 예측이 대략 절반 정도의 확률로 가능하다는 것을 의미한다.

```
고등어조림 : 0.5506849315068493
명태찌개 : 0.4794520547945205
건절이 : 0.44931506849315067
후박죽 : 0.4794520547945205
궁금한 날짜는 몇월입니까?
10
며칠인지도 말해주세요
27
고등어조림 : 저렴한편입니다
명태찌개 : 저렴한편입니다
건절이 : 비싼편입니다
후박죽 : 저렴한편입니다
```

. 이는 물가의 예측이 대략 절반 정도의 확률로 가능하다는 것을 이렇게 성공률이 절반 정도인 이유로 물가가 변하는 환경이 non-deterministic world 이며 트레이닝 셋이 부족하기 때문이라는 결론을 얻었다. 트레이닝셋으로 사용한 2013 년도 자료부터 2017 년도 자료를 이용하였는데 각 연도에서 재료의 가격 변화 추이를 학습하는 것이다 보니 5 개년도의 자료로는 그 이상 학습시키는 것에 어려움이 있었다. 뿐만 아니라 예측한대로 결과가 나오는 deterministic world 의 경우와는 달리 물가가 오를지 내릴지 혹은 유지할지 모르기 때문에 이전 년도의 데이터로 학습하더라도 그 예측이 우연한 경험일 수 있기에 확신할 수 없다. 이러한 non-deterministic 의 환경에서 사람의 경우 최종 목표, 학습에 다다른 것이 매우 어렵지만 강화 학습 알고리즘을 통해 사람이 찾기 힘든 패턴을 학습시키는 것이 가능하다. 이러한 이유에서 약 50%의 예측 성공률이 낮지 않다고 볼 수 있다.

향후엔 알고리즘의 예측 방향을 좀더 개선하여 각 년도에서 월별 물가를 학습시키는 방향으로 나아가서 효율적인 예측이 가능하도록 할 수 있다. 연 평균 가격과 비교하여 원하는 날짜의 가격을 파악하는 것은 주부들이 비교하고자 하는 날짜 간의 가격비교에 어려움이 있을 수 있기에 월 평균 가격과 비교하여 입력된 날짜의 가격이 비싸질 것인지 저렴해질 것인지 알려준다면 보다 효과적으로, 소비자가 원하는 가격 추이를 알려줄 수 있을 것으로 보인다.

### 참 고 문 헌

- [1] 서울특별시 물가정보  
<http://mulga.seoul.go.kr/info/price02.jsp#>
- [2] 오일석 "Machine Learning", pp. 509-510.
- [3] Richard S. Sutton and Andrew G. Barto "Reinforcement Learning", pp 131-132.