ISSN 2383-630X(Print) / ISSN 2383-6296(Online)
Journal of KIISE, Vol. 45, No. 12, pp. 1287-1291, 2018. 12
https://doi.org/10.5626/JOK.2018.45.12.1287

시간차 학습을 이용한 단어 감정 값 측정법 연구

(Measuring Semantic Orientation of Words using Temporal Difference Learning)

김 영 삼 †　　신 효 필 ††
(Youngsam Kim)　　(Hyopil Shin)

요 약 시간차(temporal-difference) 학습은 강화학습의 핵심적인 알고리즘으로 마르코프 체인 모형에서 상태의 가치를 실시간으로 측정하는데 유용한 방법론을 제공한다. 이 방법론에서 활용되는 마르코프 모형은 감쇄 비(discount factor)를 사용하여 보상이 주어지는 시점과 가까운 상태일수록 보상 값에 대해 더 많은 가중치를 주게 된다. 본 논문에서는 텍스트의 어떤 어휘가 갖는 감정 값을 측정하는데 있어 시간차 학습이 기존의 베이즈 확률을 이용하는 방법보다 상대적으로 유용함을 보이고자 한다. 이는 시간차 학습이 본질적으로 점증적(incremental) 처리이며 감쇄 비를 통해 부여할 감정 값의 가중치를 조절할 수 있기 때문이다. 본 논문은 영화평 자료를 이용하여 이 방법의 효과를 간접적인 방법과 직접적인 방법 모두에서 검증하였으며, 이 방법이 대용량의 자료에 적용 가능함(scalable)을 보이기 위해 비동기 병렬처리 방식으로도 이 방법의 효과가 유지됨을 검증하였다.

키워드: 시간차 학습, 강화학습, 단어 감정 값, 점증적 처리, 비동기 병렬처리

Abstract Temporal-difference(TD) learning is a core algorithm of reinforcement learning, which employs models of Markov process. In the TD methods, rewards are always discounted by a discount factor and states receive these discounted values as their rewards. In this paper, we attempted to estimate a semantic orientation of words in texts using the TD-based methods and examined the effectiveness of the proposed methods by comparing them to existing feature selection methods (indirect approach) and Bayes probabilities (direct approach). The TD-based estimation would be useful for tasks of social opinion mining, since TD learning is inherently an on-line method. In order to show our approach is scalable to huge data, the estimation method is also evaluated using asynchronous parallel processing.

Keywords: temporal-difference learning, reinforcement learning, semantic orientation of words, incremental processing, asynchronous parallel processing

† 학생회원 : 서울대학교 협동과정 인지과학
youngsamy@gmail.com
†† 정 회 원 : 서울대학교 언어학과 교수(Seoul Nat'l Univ.)
hpshin@snu.ac.kr
(Corresponding author임)

논문접수 : 2018년 8월 31일
(Received 31 August 2018)
논문수정 : 2018년 9월 21일
(Revised 21 September 2018)
심사완료 : 2018년 10월 4일
(Accepted 4 October 2018)

정보과학회논문지 제45권 제12호(2018. 12)

1. 서 론

측정된 단어나 절의 감정 값(semantic orientation /sentiment polarity)은 감정 분석 과정에서 유용하게 이용될 수 있는데, 초창기 감정 분석연구에서 이 값들은 사람이 직접 주석한 값들을 이용하였다[1-3]. 이후 Turney[4]는 비지도(unsupervised) 방법을 이용해서 감정 값을 측정하는 연구를 하였으나, 개인이 얻기는 불가능한 대규모 말뭉치 자료가 필요하다는 한계점이 있었다. 그리고 Potts[5]는 영화평 자료에서 주석된 감정 레이블에 따른 개별 단어의 베이즈 확률을 계산하여 사용할 수 있음을 보였다.

본 논문에서는 특정 어휘의 감정 값은 그 대상에 대한 대중의 호/불호를 반영할 수 있다는 점과 트위터와 같은 소셜 블로그의 텍스트에서 감정가를 나타내는 메타태그를 많이 쓴다는 점에 착안하여, 시간차 학습(temporal-difference learning) 알고리즘인 TD 람다 모형을 이용한 단어 감정가 측정법을 제안하고자 한다.

TD 람다 모형은 강화학습의 핵심적인 알고리즘으로 기존의 여러 예측 과제 실험에서 그 효율성과 정확성이 입증되었다[6-10]. TD 기반 방법은 경험을 바탕으로 하여 어떤 상태에서 기대할 수 있는 보상에 대한 예측치를 제공하는데, 기계학습의 다른 방법들과 구별되는 특징은 이 방법이 본질적으로 점증적(incremental)이라는 데 있다. 즉, 이 방법은 텍스트의 어떤 지점에서도 예측값을 갱신(update)할 수 있다는 장점을 갖는데, 이는 기존의 방법들이 단어의 값을 갱신하기 위해서는 적어도 텍스트나 문장을 하나의 단위로 정해야 한다는 점에서 매우 대조적이다[11,12].

근래에는 스마트폰과 같은 모바일 기기의 발달과 범람으로 인해 트위터나 페이스북과 같은 소셜 블로깅 서비스 사용이 보편화 되었고, 이를 이용한 소셜 의견분석 연구 또한 활발하다. 하지만 여기에는 여러 복잡한 기술적, 이론적 문제가 존재하는데, 우선 이런 문제에서 목표 데이터 셋이 빅데이터인 경우가 많다는 점과 실시간 분석이 용이하지 않다는 점이 있다.

본 논문에서 제안하는 TD 방법은 주어진 텍스트에 대한 완전한 실시간 처리가 가능하므로 실시간 분석이 가능할 뿐 아니라, 텍스트 중간에 주어지는 감정 관련 이모티콘과 같은 태그를 예측 값 측정에 유용하게 활용할 수 있다는 장점이 있다. 또한 본 연구에서는 비동기 병렬처리에 기반한 Asynchronous TD 방법을 이용한 실험을 통해 이 방법이 빅데이터와 같은 문제에 대해 scalable하다는 점을 보일 것이다. 강화학습에 비동기 병렬처리를 이용하는 기본적 전략은 다수의 학습 에이전트가 주어진 데이터를 나누어 학습한 결과를 비동기 방식으로 공유하는 것이다[6-8].

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 TD 람다 알고리즘과 그 부수적 조건들을 설명하고, 3장에서 연구의 전반적인 방법을 설명한다. 4장에서는 TD 모형에 따라 계산된 단어 감정가를 이용한 어휘 자질선택의 효과를 기술하며, Stanford Sentiment Treebank에서 주석된 단어 감정가를 TD 모형에 의해 측정된 감정가와 직접적으로 그 상관성을 살펴본다. 그리고 비동기 병렬처리 방식으로 계산한 TD 기반 단어 감정가 실험결과를 기술한다. 마지막으로 5장에서는 결론을 요약한다.

2. 연구 배경

Sutton[9-11]은 TD람다 모형을 통해 가장 단순한 형태의 시간차 학습과 몬테 카를로 학습을 람다라는 파라미터를 도입하여 통합된 모형으로 만들었다. 그의 모형에서 특정 상태에 대한 가치 함수 V는 마르코프 보상과정을 통해 근사 되는데, 마르코프 보상과정(Markov Reward Process)은 네 개의 구성요소로 이루어진 튜플 (S, P, R, γ)로 S는 상태공간, P는 상태 이행확률 함수, R은 실수의 보상 값을, $\gamma \in [0,1]$는 보상 값에 대한 감쇄비를 나타낸다.

본 연구의 TD 학습은 이 MRP(Markov Reward Process)를 사용하며 가장 단순한 TD 학습인 TD(0)는 식 (1)을 따라 상태 값을 측정한다.

$$V(S_t) = V(S_t) + \alpha(R_{t+1} + \gamma V(S_{t+1}) - V(S_t)) \quad (1)$$

여기서 그리스 알파벳 α는 학습률을 나타낸다.

몬테 카를로 학습에 해당하는 TD(1)은 의미적으로는 식 (2)와 같으나, TD 람다 모형에서는 이 값이 기존의 몬테 카를로 모형에서와는 달리 점증적으로 계산될 수 있다는 점이 다르다.

$$R_t = \sum_{s=t}^{T-1} \gamma^{s-t} R_{t+1}$$

$$V(S_t) = V(S_t) + \alpha(R_t - V(S_t)) \quad (2)$$

본 연구에서 사용된 TD 람다 알고리즘은 아래와 같은데 여기 제시된 알고리즘은 Sutton과 Barto[11]에서 제시된 것과는 다르게 보다 빠른 계산을 위해 수정된 것이다.

적격 흔적도(eligibility traces) 개념은 TD 람다 모형에서 상태 값의 N-step 백업을 backward view 방식으로 구현하기 위해 도입된 것으로, 파라미터 람다를 통해 조절된다. 람다가 0이면 이는 가장 단순한 TD 모형인 식 (1)을 따르며, 람다가 1이고 보상 감쇄비도 1일 때 식 (2)와 같은 몬테 카를로 방법이 된다.

적격 흔적도에는 몇 가지 유형들이 있는데, Sutton과 Barto[11]에서는 replacing traces (알고리즘 1의 7번째 줄)과 accumulating traces가 식 (3)과 같이 소개되어 있다.

알고리즘 1. 빠른 TD 람다(replacing traces)
Algorithm 1. Fast TD(λ) with replacing traces

```
Initialize V(s) arbitrarily and let e(s) = 0 for all s ∈ S;
H ← new hash table;
repeat
    while s_t not at end of the episode do
        observe reward, r, and s_{t+1};
        δ ← r + γV(s_{t+1}) − V(s_t);
        e(s_t) ← 1;
        if H not contains s_t then
            insert s_t into H;
        for all h ∈ H do
            if e(h) ≤ 0.001 then
                e(h) ← 0;
                remove h from H;
                continue;
            V(h) ← V(h) + αδe(h);
            e(h) ← γλe(h);
until the episode is terminal;
```

$$e(S_t) = e(S_t) + 1 \quad (3)$$

보다 근래의 연구에서 True Online TD 모형과 함께 Dutch traces [13,14]가 발표되었는데 이는 식 (4)와 같다.

$$e(S_t) = (1-\alpha)e(S_t) + 1 \quad (4)$$

본 연구에서는 위 세 가지 적격 흔적도 방법들을 이용한 TD 람다 모형과 True Online TD 모형을 TD 기반 측정 방법들로 활용하였다.

3. 연구 방법

본 연구의 강화학습 환경은 그림 1과 같다. 그림 중간의 이모티콘처럼 단어가 아니라 긍정 감정을 표현하는 이모티콘이 등장하면 +1의 보상 값을 제공하는 양상을 나타낸다. 이 학습 환경에서 상태공간은 어휘 집합으로 정의되며, 한 에피소드는 하나의 감정 레이블이 주어진 한 텍스트로 정의되었다.

본 연구의 실험은 세 종류의 실험들로 이루어진다. 첫 번째 실험에서는 Naive Bayes(NB) 분류기를 이용한 영화평 자료(polarity dataset)[15]의 감정분석 과제에서 단어 자질 셋을 선택하는데 TD 방법을 이용하는 조건

그림 1 본 연구에서 가정된 마르코프 결정 과정 도식. 마지막 검은 원은 텍스트의 감정분류 레이블을 가리키며 화살표 위/아래의 숫자는 보상 값을 가리킨다.

Fig. 1 Example of MRP used in this research. The last black state indicates the label of the text and numbers are rewards given for the transitions

들과 기존의 자질선택 방법들[16,17]을 비교하였다. 단어의 감정가는 알고리즘 1을 따라 매 단계마다 갱신되는 상태 값의 점증평균(incremental mean)을 사용하였으며, 감정가 측정이 끝나면 최대/최소 긍정 값을 가진 단어를 각각 5,000개씩 선택해 총 1만개 어휘를 자질 집합으로 이용하여 감정분류 과제에 이용하였다.

두 번째 실험에서는 Stanford Sentiment Treebank [18]를 이용하여, 각 단어에 주석된 감정가(0~1)와 TD 방법에 의해 측정된 감정가와의 상관관계를 살펴보았다. 이 실험의 방법과 절차는 실험 1과 동일하고 비교조건으로는 Potts[5]의 베이즈 확률공식을 통한 방법을 이용하였는데, 그 이유는 실험 1의 비교방법들은 자질선택 과제에만 적합하고 감정가 측정에는 활용할 수 없기 때문이다.

마지막 세 번째 실험에서는 비동기 병렬처리 방식으로 TD 방법을 이용하는 것이 가능한지를 살펴보았다. 이를 위해 Amazon Web Service의 C5 인스턴스를 이용하여 다코어 머신에서 각 프로세서가 동일한 상태 값 벡터를 공유하면서 영화평 자료(sentence polarity dataset) [19]의 단어의 감정가 갱신을 수행하도록 하였다.

4. 실험 결과

표 1은 첫 번째 실험 결과를 보여준다. 각 방법마다 10-fold cross 검증집합을 만들고 정확도의 평균을 최종 정확도로 기록하였다. TD 방법들의 모수는 사전 파일럿 실험을 통해 최적 모수 값을 사용하였는데, 최적 흔적 감쇄비인 람다 값은 모두 1이었다. 사용된 보상 감쇄비는 모두 0.99였다. 학습률은 accumulating traces 조건과 True Online TD 조건에는 0.1을 사용했으며, replacing 조건에는 0.3을, Dutch 흔적도 조건에는 0.2를 이용하였다.

표 1의 결과에서 가장 높은 정확도를 기록한 방법은

표 1 어휘 자질 선택과제의 최종 정확도 결과(10-fold cross validation)

Table 1 Averages of accuracies of 10-fold cross validation sets

Method	NB Accuracy
TD(1) with accumulation	**0.84**
TD(1) with replacing	0.83
TD(1) with Dutch	0.83
True Online TD(1)	0.78
Simple Averages	0.64
Document Frequency	0.67
Averaged TF-IDF	0.69
χ^2 statistic	0.66
Information Gain	0.83

accumulating traces를 사용한 TD(1) 방법이었으며, 가장 낮은 성능을 기록한 것은 Simple Averages 조건으로 이 조건은 각 단어가 출현한 텍스트의 감정 레이블 값을 모두 합한 후에 평균을 구하는 방법으로 가장 단순한 단어 감정가 계산방식이다.

두 번째 실험결과인 표 2는 Stanford Sentiment Treebank의 모든 단어(21,684개)의 주석된 감정가와 TD 방법 및 비교방법에 의한 측정가와의 상관관계를 계산한 결과를 보여준다(True Online TD 조건은 실험에서 부족한 성능으로 인해 제외되었다). 그리고 표 3은 주석된 감정가에서 가장 높은 분산을 갖는 세 개의 형태소 태그(JJS, RBR, JJ)에 속하는 4,532 단어셋에 대한 결과를 표시한다. 표 3과 4에서 볼 수 있듯이, TD 방법은 베이즈 확률을 사용한 비교조건보다 더 높은 상관관계를 기록하였다.

마지막 실험인 비동기 병렬처리를 통한 실험결과는 표 4에 제시되었다. Replacing traces를 사용한 TD 방법을 대표방법으로 이용하였으며, 구체적인 실험방법은 실험 1과 같다. 또한 본 결과가 바닥효과에 의한 것이 아니라는 것을 보이기 위한 비교 방법으로 LSTM 기반 분류조건이 이용되었는데, 그 정확도는 0.75였다(2 hidden layers, embedding dimensions: 100, hidden dimensions: 256, learning rate: 0.01, epochs: 7).

표 2 모든 단어(24,684개)에 대한 상관관계 결과표
Table 2 Correlations of the total word set

Method	Pearson	Spearman
Bayes Prob.	0.21	0.2
TD(1) with replacing	0.24	0.21
TD(1) with Dutch	0.24	0.21
TD(1) with accumulation	0.24	0.21

표 3 형용사 및 부사 태그를 가진 단어셋에 대한 상관관계 결과표(4,532개)
Table 3 Correlations of 4,532 words (JJS, RBR, JJ tagged)

Method	Pearson	Spearman
Bayes Prob.	0.32	0.3
TD(1) with replacing	0.38	0.35
TD(1) with Dutch	0.38	0.35
TD(1) with accumulation	0.38	0.34

표 4 비동기 병렬처리를 이용한 감정분류 정확도 결과
Table 4 Accuracy results of sentiment analysis using asynchronous parallel processing

Processors	1	2	4	8	16
Time (sec)	53.6	28.6	14.6	7.3	3.9
Accuracy	0.76	0.76	0.76	0.75	0.75

표 4의 결과에 따르면, 프로세서 수가 증가함에 따라 처리시간은 선형적으로 감소하였으며, 정확도는 거의 변하지 않았다.

5. 결 론

우리는 temporal-difference 람다 방법을 이용해 단어의 감정가 측정을 실험하였으며, 기계학습에서 기존에 사용되던 방법들에 비하여 더 나은 수행을 보였음을 간접적으로, 그리고 직접적으로 관찰하였다. 또한 비동기 병렬처리를 통한 temporal-difference 기반 감정가 측정 실험을 통해 이 방법이 매우 큰 데이터에 대해서도 scalable 하다는 점을 확인하였다. 본 연구의 방법론은 방대한 실시간 SNS 데이터에 대한 소셜 의견 분석과 같은 과제에 적합하게 이용될 수 있을 것으로 보인다.

References

[1] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," *Proc. of the eighth conference on European chapter of the Association for Computational Linguistics*, pp. 174-181, 1997.

[2] J. Wiebe, "Learning subjective adjectives from corpora," *AAAI/IAAI* 20, pp. 735-740, 2000.

[3] M. Taboada, C. Anthony, and K. Voll, "Methods for Creating Semantic Orientation Dictionaries," *LREC,* pp. 427-432, 2006.

[4] P.D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," *Proc. of the 40th annual meeting on association for computational linguistics*, pp. 417-424, 2002.

[5] C. Potts, "On the negativity of negation," *Semantics and Linguistic Theory*, Vol. 20, pp. 636-659, 2010.

[6] V. Mnih, A. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," *International Conference on Machine Learning*, pp. 1928-1937, 2016.

[7] M. Babaeizadeh, I. Frosio, S. Tyree, J. Clemons, and J. Kautz, "Reinforcement learning through asynchronous advantage actor-critic on a gpu," *International Conference on Learning Representations*, 2017.

[8] D. Horgan, J. Quan, D. Budden, G. Barth-Maron, M. Hessel, H. Van Hasselt, D. Silver, "Distributed Prioritized Experience Replay," *International Conference on Learning Representations*, 2018.

[9] R. S. Sutton, *Temporal credit assignment in reinforcement learning*, Ph.D. Dissertation, University of Massachusetts, Amherst, MA. 1984.

[10] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine learning*, Vol. 3,

No. 1, pp. 9-44, 1988.
[11] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 1998.
[12] C. Watkins and P. Dayan, "Q-learning," *Machine learning*, Vol. 8, No. 3,4, pp. 279-292, 1992.
[13] H. Seijen and R. S. Sutton, "True Online TD (lambda)," *PMLR*, pp. 692-700, 2014.
[14] H. Seijen, A R. Mahmood, P. Pilarski, M. Machado, R. Sutton, "True Online Temporal-Difference Learning," *Journal of Machine Learning Research*, Vol. 17, No. 145, pp. 1-40, 2016.
[15] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," *ACL*, pp. 271-278, 2004.
[16] Y. Yang and J. O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," *Proc. of the Fourteenth International Conference on Machine Learning*, pp. 412-420, San Francisco, CA, USA, 1997.
[17] C. Lee and G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization," *Information processing & management*, Vol. 42, No. 1, pp. 155-165, 2006.
[18] R. Socher, A. Perelygin., J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, & C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," *Proc. of the conference on empirical methods in natural language processing*, pp. 1631-1642, 2013.
[19] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," *Proc. of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics*, pp. 115-124, 2005.

김 영 삼

2005년 대전대학교 철학과 졸업(학사) 2007년 서울대학교 협동과정 인지과학 졸업(석사). 2018년 서울대학교 협동과정 인지과학 졸업(박사). 관심분야는 강화학습, 자연어처리

신 효 필

1988년 서울대학교 언어학과 졸업(학사) 1990년 서울대학교 언어학과 졸업(석사) 1994년 서울대학교 언어학과 졸업(박사) 1997년 12월 University of Missouri, Computer Science 졸업(석사). 1998년 1월~2001년 1월 Computing Research Lab, New Mexico State University. 2001년 1월~2001년 12월 YY Technology in Silicon Valley. 2001년 9월~2003년 2월 서울대학교 전자공학부 BK교수. 2003년~현재 서울대학교 인문대학 언어학과 교수. 관심분야는 딥러닝/강화학습을 이용한 자연언어처리