RNN – Automata Paper Review

Index

1. Intro

Automata(automaton) are a relatively self-operating machine, or a machine or control mechanism designed to automatically follow a predetermined sequence of operations or respond to predetermined instructions. Nowadays, Artificial intelligence and Machine learning are the most popular Research subject in 21C. In these perspective, Recurrent Neural Networks(RNNs) are central to deep learning, and natural language processing in particular. RNNs are a class of neural networks used to process sequences of arbitrary lengths. However, while they have been shown to reasonably approximate a variety of languages, what they eventually learn is unclear. In these papers, the authors analyzed deterministic finite automaton (DFA) in RNNs and how RNNs trained to recognize regular formal languages represent knowledge in their hidden state by using automata theory. I will review these two papers by using automata theory that learned in class

2. Contented

1. Paper 1 Extracting Automata from Recurrent Neural Networks Using Queries and Counterexamples.

: RNNs 는 언어의 변화에 대해 합리적이게 대략적으로 보여준다. 즉 그들이 학습하는 것은 명확하지 않습니다. 이러한 명확하지 않은 것에 대해 확실한 규칙을 뽑아내기 위한 시도가 있었습니다. 이러한 시도에서 유한한 알파벳 Σ으로 훈련된 주어진 RNN-acceptor R 에서 저자들은 R 과 굉장히 비슷한 연속적인 방법을 특정하는 DFA A 를 추출하려고 합니다. 저자는 특징을 나타내는 DFA 를 추출하기 위해 훈련된 RNN 을 L* algorithm 을 위한 teacher 로 임명합니다. 여기서 L* Algorithm 은 어떠한 정규 언어 L 을 위해 최소한으로 적당한 teacher 로부터 온 DFA 를 학습하기 위한 exact learning algorithm 입니다. 이를 통해 저자들은 RNN 에서 exact learning 을 통한 DFA 를 추출하는 새로운 기법을 제시합니다. 이 방법은 네트워크의 내부 구성에 대한 가정을 하지 않기에 RNN architecture 어디에든 쉽게 적용할 수 가 있습니다. 여기서 저자들이 지금 산업의 근간이 될 수도 있는 Deep learning 에서 중요한 RNN 에 대한 새로운 시각을 오토마타를 통해 전달 하였습니다. 기존 RNN 은 대략적으로 보여주었던 것을 이제 DFA 를 추출해 이 DFA 에서 주어진 alphabet 과 transition 그리고 state 들을 통해 regular language 를 확실히 알 수 있다는 뜻입니다. 여기서 DFA 는 비록 추상적인 수학적

개념이지만 이는 자주 H/W 와 S/W 에서 다양한 특정 문제를 풀기 위해 구현이 됩니다. 이를 봤을 때 앞으로 Deep Learning 에서 모호성을 줄이는 것에 도움이 되지 않을까 생각합니다.

2. Representing Formal Languages : A Comparison Between Finite Automata and Recurrent Neural Networks.

: 저자는 이 논문에서 본 논문에서는 훈련 된 RNN 이 문법 구조를 어떻게 표현하는지 이해하는 새로운 방법을 제안합니다. 이를 동일한 언어 인식 과제를 해결하는 유한 오토마타와 비교합니다. 공식 언어를 인식하도록 훈련 된 RNN 의 내부 지식 표현을 전통적으로 동일한 공식 언어를 정의하는 데 사용되는 오토마타 이론 모델의 상태에 쉽게 매핑 할 수 있을까에 대해 실험을 함으로써 보이려고 합니다. 저자는 paper 에서 RNN 은 주어진 공식 언어에서 무작위로 생성된 문자열의 긍정적 및 부정적 예의 데이터 세트에 대해 훈련합니다. 훈련 된 RNN 의 숨겨진 상태를 표준 FA 의 상태로 매핑하는 동 형사상. 동일한 언어를 받아들이는 FA 가 무한히 많기 때문에, 저자는 여기서 동일한 언어를 인식하는 가장 최소한의 state 인 MDFA 에 초점을 맞췄습니다. 저자는 본문을 통해 RNN 이 정규 형식 언어가 숨겨진 상태에서 지식을 나타내는 것을 인식하도록 훈련 된 방법을 연구했습니다. 특히, 내부 표현을 언어를 정확하게 인식하는 표준 최소 DFA 로 해독 할 수 있는지 여부를 물었고 ground truth 라고 보일 수 있습니다. 저자는 선형 함수가 그러한 디코딩을 수행하는 데 현저히 좋은 일을 한다는 것을 보였습니다. 이것을 통해 저자는 이 논문이 Neural Network 가 정규 논리 개념을 배울 때의 근간의 단계라고 생각합니다. 저자는 이 논문을 위해 RNN 이 정규 언어를 학습할 때의 내부 방법을 오토마타의 DFA 와 비교하는 것을 볼 수 있습니다. 이것은 눈에 잘 보이지 않는 RNN 의 인식을 우리가 수학적으로 정의하고 또 알 수 있는 DFA 를 이용해 RNN 을 분석하려 한 것을 알 수 있습니다. 이에 대해 오토마타는 주어진 alphabet 과 그에 대한 transition 에 대해 결과를 우리가 사고 할 수 있는 방식으로 나타내는 학문을 알았고 컴퓨터 과학에서 애매하고 가시적이지 않은 부분에서 오토마타를 이용한다면 효과를 얻을 수 있다는 것을 배울 수 있었습니다.

4. Reference

[1] Gail Weiss, Yoav Goldberg, Eran Yahav. Extracting Automata from Recurrent Neural Networks Using Queries and Counterexamples. In Proceedings of the 35th International Conference on Machine Learning, PMLR 80:5247-5256, 2018.

[2] Joshua J, Michalenko, Ameesh Shah, Abhinav Verma, Richard G. Baraniuk, Swarat Chaudhuri, Ankit B. Patel. Representing Formal Languages: A Comparison Between Finite Automata and Recurrent Neural Networks. In ICLR 2019 Conference Blind Submission.

[3] Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan "Hona" Cernocky, Sanjeev Khudanpur. Recurrent nueral network based language model. In INTERSPEECH 2010 11th Annual Conference of the International Speech Communication Association.

[4] J. E. Hopcroft and J. D. Ullman. Introduction to Automata Theory, Languages and Computation. Addison-Wesley, 1979.

[5] L. Miclet and C. de la Higuera. Grammatical Inference: Learning Syntax from Sentences. Springer, 1996.
[6] C. W. Omlin and C. L. Giles. Constructing deterministic finite-state automata in recurrent neural networks. Journal of the Association of Computing Machinery, JACM, 43(6):pages 937–972, 1996a.

[7] C. W. Omlin and C. L. Giles. Extraction of rules from discrete-time recurrent neural networks. Neural Networks, 9(1):41–52, 1996b.

[8] Zeng, Z., Goodman, R. M., and Smyth, P. Learning fi- nite state machines with self-clustering recurrent net- works. Neural Computation, 5(6):976–990, 1993. doi: 10. 1162/neco.1993.5.6.976. URL https://doi.org/ 10.1162/neco.1993.5.6.976.

[9] Tomita, M. Dynamic construction of finite automata from examples using hill-climbing. In Proceedings of the Fourth Annual Conference of the Cognitive Science Soci- ety, pp. 105–108, Ann Arbor, Michigan, 1982.

[10] Cechin, A. L., Simon, D. R. P., and Stertz, K. State au- tomata extraction from recurrent neural nets using k- means and fuzzy clustering. In Proceedings of the XXIII International Conference of the Chilean Computer Sci- ence Society, SCCC '03,

pp. 73–78, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-2008-1. URL http://dl.acm.org/citation.cfm? id=950790.951318.

[11] Adam Niewola, Leszek Podsedkowski. L* Algorithm – A Linear Computational Complexity Graph Searching Algorithm for Path Planning. Journal of Intelligent & Robotic Systems volume 91, pages425–444(2018)