

AI 기반 브랜치 예측 기법 연구 동향

이주원¹, 김유신¹, 오현영^{2*}
¹가천대학교 AI 소프트웨어학부 학부생
²가천대학교 AI 소프트웨어학부 교수

{ljw0509, kysle32, hyoh}@gachon.ac.kr

A Survey on AI-Based Branch Prediction Techniques

Ju-Won Lee, Yoo-Shin Kim, Hyunyoung Oh
Dept. of AI · Software, Gachon University

요 약

브랜치 예측은 CPU 파이프라인 성능 최적화의 핵심이다. 본 논문은 퍼셉트론, CNN, SRNN, 그리고 강화학습을 이용한 AI 기반 브랜치 예측 기법의 최신 동향을 조사한다. 이러한 기법들은 제어 헤저드에 대한 예측 정확도를 향상시켜 파이프라인 성능을 개선한다. 특히 SRNN 기반 예측기는 초기 학습 단계에서 높은 정확도와 낮은 지연 시간을 보이며, 강화학습 접근법은 복잡한 브랜치 패턴에서 효과적인 성능을 보여준다. 본 연구는 이러한 AI 기반 방법들이 파이프라인 최적화에 미치는 영향을 분석한다.

1. 서론

컴퓨터 구조의 주요 목표 중 하나는 CPU 성능을 극대화하는 것이다. 명령어를 병렬로 처리하기 위해 파이프라이닝 기법이 널리 사용되지만, 이는 CPU 성능을 저하시키는 헤저드로 인해 어려움을 겪을 수 있다. 특히 브랜치 예측 실패는 파이프라인 지연과 플러시를 유발하여 특히 문제가 된다. 전통적인 결정론적 방식과 규칙 기반 예측 방법이 이 문제를 해결하기 위해 사용되어 왔지만, 복잡한 브랜치 패턴을 처리하는 데 한계를 보였다. 최근 연구에서는 브랜치 예측을 위한 더 효율적인 해결책으로 AI 기반 예측 방법을 도입했다. 이 논문은 이러한 AI 기반 브랜치 예측 기법의 최신 동향과 파이프라인 최적화에 대한 효과를 조사하는 것을 목표로 한다.

2. 최근 연구 및 대표 논문

2.1 퍼셉트론 기반 브랜치 예측

퍼셉트론 기반 브랜치 예측 기법은 단일 층 신경망 모델을 사용하여 브랜치 history와 결과 간의 상관관계를 학습한다. Jimenez et al.[1]이 제안한 퍼셉트론 예측기는 각 브랜치 주소에 대해 가중치 벡터를 유지하고, 입력과 가중치의 내적을 계산하여 예측을 수행한다. 이 방법은 선형 분리 가능한 패턴에 대해 높은 정확도를 보이지만, 복잡한 비선형 패턴에 대해서는 한계가 있다. Nain S et al.[2]의 연구에서는

이 퍼셉트론 모델을 활용해 명령어 수준 병렬성(ILP) 프로세서에서 제어 헤저드를 해결하는 방법을 제시했다. 이 접근 방식은 복잡한 패턴을 학습하여 브랜치 예측 실패를 줄이고, 파이프라인 성능을 10-15% 향상시키는 결과를 보였다.

2.2 컨볼루션 신경망(CNN) 기반 브랜치 예측

Mao Y et al.[3]은 LeNet부터 ResNet-50까지 다양한 CNN 구조를 브랜치 예측에 적용하여 실험했다. CNN은 컨볼루션 레이어와 풀링 레이어를 통해 로컬 특징을 효과적으로 추출하고, 이를 통해 브랜치 패턴의 계층적 표현을 학습할 수 있다. 이 방법은 제어 헤저드와 데이터 헤저드 모두에 대응할 수 있으며, 약 12%의 성능 향상을 달성했다. 실험 결과, CNN의 깊이가 증가할수록 예측 정확도가 향상되는 것을 확인했으며, ResNet-50 구조가 가장 높은 성능을 보였다.

2.3 순환 신경망(SRNN) 기반 동적 브랜치 예측

Zhang L[4]이 제안한 SRNN 기반 예측기는 RNN의 장점을 살리면서도 병렬 처리를 통해 지연 시간을 줄일 수 있는 구조를 제시했다. SRNN은 입력 시퀀스를 여러 개의 부분 시퀀스로 나누고, 각 부분 시퀀스에 대해 병렬로 RNN을 실행한 후 결과를 병합하는 방식으로 동작한다. 이 방법은 파이프라인 성능을 15% 이상 향상시키며, 복잡한 명령어 흐름에서도 높은 예측

* 교신저자

정확도를 유지하는 것으로 보고되었다. 특히 학습 초기 단계에서 기존의 퍼셉트론 예측기보다 2.34% 높은 예측 정확도를 보였다.

2.4 강화학습을 활용한 브랜치 최적화

최근에는 강화학습을 활용한 브랜치 관련 연구가 진행되고 있다. Parsonson et al.[5]은 분기 한정법 (Branch-and-Bound) 알고리즘에서의 분기 변수 선택에 강화학습을 적용하는 "retro branching" 방법을 제안했다. 이 접근 방식은 브랜치 예측 문제에도 유용한 통찰을 제공할 수 있다. 이 연구는 긴 에피소드, 큰 상태-행동 공간, 부분 관찰 가능성 등 강화학습을 적용하기 어려운 문제들을 해결하기 위한 새로운 방법을 제시했다. 검색 트리를 여러 개의 경로로 분해하여 에이전트가 더 짧고 예측 가능한 궤적에서 학습할 수 있게 했으며, 전문가 지침이나 사전 훈련 없이도 분기 변수 선택을 학습할 수 있음을 보였다. 이 방법은 기존의 최신 강화학습 알고리즘보다 3-5 배 우수한 성능을 달성했으며, 모방 학습 방법의 성능에 20% 이내로 근접했다. 이러한 접근 방식은 브랜치 예측에서도 긴 시퀀스의 명령어 실행 과정을 더 짧고 관리 가능한 세그먼트로 나누어 학습하는 방식으로 적용될 수 있는 가능성을 보여준다.

3. 각 방식 간의 비교

각 논문에 소개된 내용을 바탕으로 (표 1)의 비교표를 작성하였다. 다만, 강화학습 기반 방식[5]이 제시하는 성능 지표는 분기 변수 선택에 대한 것이므로 다른 연구들과 공정한 비교가 어려워 제외하였다.

(표 1) 각 브랜치 예측 방식 비교

	퍼셉트론 기반 [2]	CNN 기반 [3]	SRNN 기반 [4]	강화학습 기반[5]
성능향상	10-15%	~12%	15% 이상	-
학습시간	빠름 (일반적으로 몇 초에서 수 분 내외)	중간 (수십 분에서 몇 시간)	빠름 (수십 초에서 수 분)	느림 (탐색 공간과 에피소드의 길이에 따라 수 시간에서 수일)
장점	선형 분기 가능한 패턴에 효과적, 하드웨어 구현이 용이	공간적 상관관계를 효과적으로 학습, 깊은 네트워크로 복잡한 패턴 포착 가능, 데이터 해저드	시간적 의존성을 효과적으로 모델링, 병렬 처리로 지연 시간 감소, 동적 환경에서 높은 적응	사전 지식 없이 새로운 패턴 발견 가능, 동적 환경에 지속적 적응, 복잡한 의사

		와 제어 해저드 모두 처리	력	결정 트리 최적화 가능
단점	복잡한 비선형 패턴에 대해 성능 제한적, 긴 브랜치 히스토리 처리에 한계	학습에 큰 데이터셋 필요, 계산 복잡도가 높음, 하드웨어 구현 시 자원 소모가 큼	구현 복잡도가 높음, 긴 시퀀스에서 학습 불안정성 가능	학습 시간이 길고 불안정할 수 있음, 대규모 상태-행동 공간에서 탐색 어려움
구현난이도	낮음 (구현이 간단하고 계산 효율적)	중간 (높은 메모리와 계산 자원 요구)	높음 (병렬화된 순환 구조 필요)	높음 (하이퍼파라미터 튜닝에 민감)

4. 결론

AI 기반 브랜치 예측 기법은 파이프라인 해저드 관리에 중요한 전환점을 가져왔다. 각 방법은 고유한 장단점을 가지고 있어, 특정 응용 분야나 하드웨어 제약 조건에 따라 적합한 방법이 달라질 수 있다. 향후 연구에서는 이러한 기술들이 더욱 통합되고 발전함에 따라, 파이프라인 해저드 관리에서 AI의 역할이 더욱 커질 것으로 예상된다. 특히 각 방법의 장점을 결합한 하이브리드 모델, 실시간 온칩 학습 알고리즘, 효율적인 하드웨어 구현 방법 등이 유망한 연구 방향이 될 것이다.

사사문구

이 논문은 2024년도 정부(산업통상자원부)의 재원으로 한국산업기술기획평가원의 지원(No. RS-2024-00406121, 자동차보안취약점기반위험분석시스템개발(R&D))과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. RS-2022-00166529)을 받고 과기정통부 정보통신기획평가원의 정보보호핵심원천기술개발사업(No. RS-2024-00337414)으로 수행한 결과임.

참고문헌

[1] D. A. Jimenez and C. Lin, "Dynamic branch prediction with perceptrons," HPCA, 2001

[2] Nain S, Chaudhary P. "A Neural Network-Based Approach for the Performance Evaluation of Branch Prediction in Instruction-Level Parallelism Processors," The Journal of Supercomputing, 2021.

[3] Mao Y et al., "Exploring Convolution Neural Network for Branch Prediction," IEEE Access, 2020.

[4] Zhang L et al., "A Dynamic Branch Predictor Based on Parallel Structure of SRNN," IEEE Access, 2020.

[5] Christopher W. F. Parsonson et al., "Reinforcement Learning for Branch-and-Bound Optimisation Using Retrospective Trajectories," AAAI, 2023