TPC AI 벤치마크 개발 동향



김효실 한국정보통신기술협회 소프트웨어시험인증연구소 책임연구원

1. 머리말

TPC(Transaction Processing Performance Council, www.tpc.org)는 트랜잭션(transaction) 처리와 데이터베이스 산업 표준을 제정하고 객관적이고 검증 가능한 성능 데이터를 업계에 알리고자 1988년에 설립된 비영리 단체다. 마이크로소프트, HPE, IBM, 오라클, 시스코, 인텔, 레노버 등이 정회원사로 있으며, TTA는 2017년부터 정회원사로 활동하고 있다.

최근 TPC는 AI 관련 벤치마크(TPCx-AI) 제정을 준비 중에 있으며, 본고에서는 TPCx-AI 소개, 개발 진행현황 및 TTA의 기여 활동에 대해 설명하고자 한다.

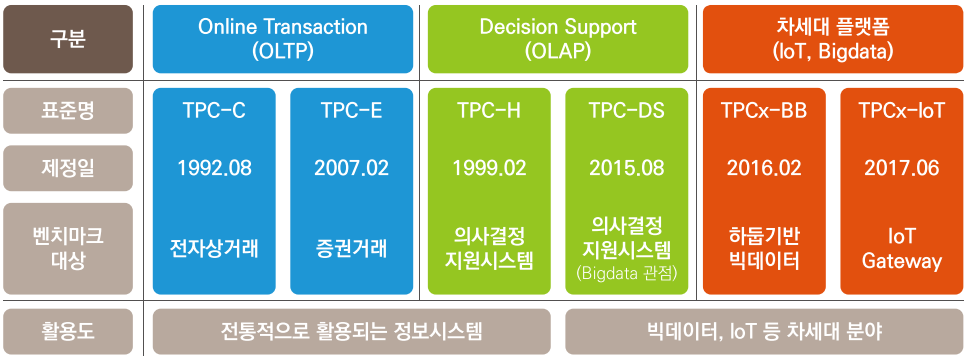
2. 개요

TPC는 TPC-C, TPC-H 등 전통적인 OLTP(Online Transaction Processing), OLAP(Online Analytical Processing) 시스템의 데이터 처리 성능을 측정하는 벤치마크를 제공한다. 그뿐만 아니라, TPCx-BB, TPCx-IoT와 같은 빅데이터, IoT 데이터 처리 성능을 측정하는 차세대 분야의 벤치마크도 제공하고 있으며, 현재 AI 분야 벤치마크 또한 개발 중이다.

AI 성능 관련 벤치마크 개발을 목적으로 2017년에 발족된 TPCx-AI 분과회의는 현재 벤치마크 툴킷 개발을 거의 완료한 상태이며 2020년 하반기에 벤치마크 출시를 목표로 활발히 활동 중이다. TTA는 올해부터 TPCx-AI 분과회의 위원으로 활발히 참여 중이며, 개발 단계 툴킷의 출시 전 테스트, 국내 AI 관련 기업의 개발 동태 관련 설문조사 같은 활동으로 해당 분과에 기여하고 있다.

3. TPCx-AI 벤치마크 소개 및 개발 진행 현황

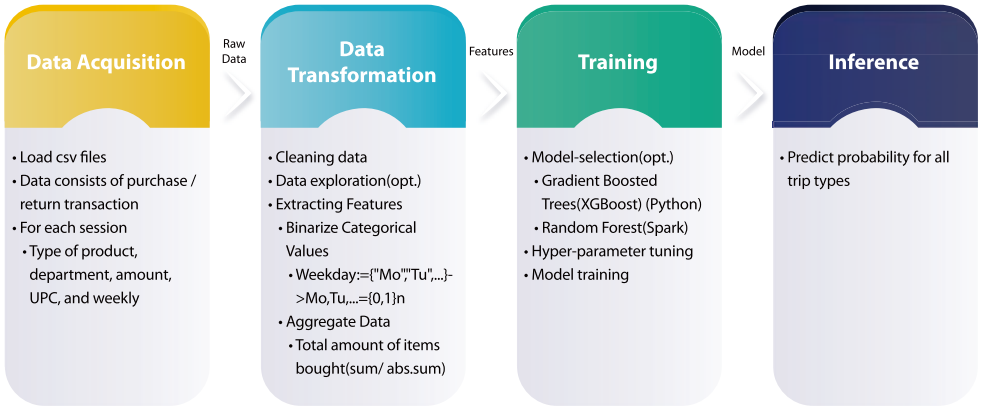
TPCx-AI 툴킷은 머신러닝 파이프라인을 데이터 생성, 데이터 로딩, 훈련(Training), 추론(Inference), 및 채점(Scoring)의 5단계로 정의



[그림 1] TPC 주요 벤치마크



[그림 2] TPCx-AI 머신러닝 파이프라인



[그림 3] 여행 종류 분류에 대한 머신러닝 파이프라인 예시

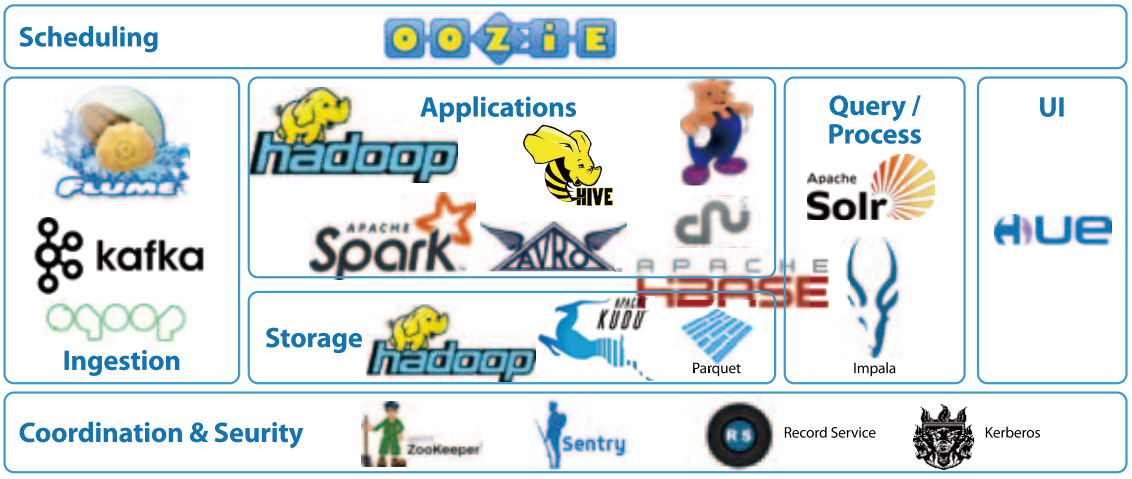
하고, 훈련 및 훈련 및 추론에서 소요되는 시간을 측정해서 성능을 비교한다. AI 분야에서 가장 많이 활용되는 8개의 유스케이스를 정의하고, 각 유스케이스에 대해 머신러닝 파이프라인의 모든 단계를 거친 후 성능이 결정되는 구조이다.

TPCx-AI 툴킷에서 현재 사용되고 있는 8개의 유스케이스는 다음과 같다.

TPCx-AI 유스케이스

1. 여행 종류 분류(Trip-Type Classification)
2. 데이터 센터에서 하드웨어 고장 탐지(Detect hardware failure in data center)
3. 고객 분류(Customer Segmentation)
4. 판매 예측(Sales Prediction)
5. 제품 추천(Product Recommendation)
6. 스팸 탐지(SPAM Detection)
7. 가격 제안(Price Suggestion)
8. 자금 세탁 탐지(Detect Money Laundering)

8개의 유스케이스에서 사용되는 알고리즘은 K-평균(K-means), 아리마(ARIMA, Auto-regressive Integrated Moving Average), 서포트 벡터 머신(Support vector machines), 그라디언트 부스트 트리(Gradient boosted trees),



※출처: http://www.comworld.co.kr/news/articleView.html?idxno=49159

[그림 4] 클라우데라 제공 패키지

나이브 베이즈(Naive bayes), 교대 최소 제곱법(Alternating least squares), 로지스틱 회귀(Logistic regression) 등이다.

추후, 언어 처리, 얼굴 인식을 포함한 2~3개 정도의 유스케이스를 추가하여 최종 벤치마크가 출시될 예정이다.

4. TPCx-AI 벤치마크 실행 및 TTA의 기여 활동

TPCx-AI 벤치마크 툴킷은 한 대의 독립적인 서버나 여러 대의 서버를 묶어서 구성하는 클러스터 단위에서 각각 실행 가능하다. 한 대의 서버에서 실행하기 위해서는 파이썬과 해당하는 라이브러리(Tensorflow, Sklearn 등)를 설치하면 된다. 클러스터 환경에서 실행하려면 클라우데라를 설치하여 하둡(Hadoop), 스파크(Spark), 주키퍼(Zookeeper) 등 서버 간 자원 할당/관리가 가능하도록 테스트 환경을 구축해야 한다.

최근 클라우데라가 CDH(Cloudera's Distribution for Hadoop) 서비스를 중지하고 CDP(Cloudera Data Platform)를 출시하면서, TTA는 새로운 플랫폼인 클라우데라 CDP에서 벤치마크 툴킷을 테스트하고 있다.

동일한 클라우드 환경에서 클라우데라 CDH와 CDP를 설치하고 벤치마크 툴킷 테스트 결과 성능이 유사하게 나왔다. 따라서 CDH에서 CDP로 옮겨가는 것에 무리가 없어 보인다.

5. TPCx-AI vs. MLPerf

TPCx-AI 벤치마크는 또 다른 머신러닝 관련 벤치마크인 MLPerf(mlperf.org)와 종종 비교된다. MLPerf는 바이두, 구글, 하버드, 스탠포드 등 유수의 기업 및 대학이 2018년에 설립한 단체다. 기계학습 관련 하드웨어, 소프트웨어, 서비스에서 훈련과 추론 과정의 성능을 측정한다.

두 벤치마크의 가장 큰 차이점은 TPCx-AI 벤치마크는 머신러닝 파이프라인 전체가 성능의 비교대상이 되는 반면 MLPerf는 훈련과 추론 과정에서의 성능이 각각 따로 측정된다는 점이다. 즉, TPCx-AI 벤치마크는 머신러닝 워크

로드를 실행하는 전체 시스템의 성능을 비교하는 데 초점을 두는 반면, MLPerf는 훈련 또는 추론 과정에 특화된 머신러닝 소프트웨어, 하드웨어 가속기, 클라우드 플랫폼의 성능을 비교하는 데 초점을 둔다. TPCx-AI는 데이터 크기(스케일 팩터, Scale factor)별로 성능 비교를 할 수 있지만, MLPerf는 데이터 크기가 고정되어 있다. MLPerf는 딥러닝 관련 워크로드를 다루지만 TPCx-AI는 딥러닝 및 전통적인 머신러닝 유스케이스를 모두 다룬다는 점도 다르다.

6. 맺음말

최근 4차 산업혁명과 더불어, 국내에 수많은 AI 관련 사업 및 과제가 등장했다. 이에 AI 성능을 측정하기 위한 벤치마크의 제정과 도입은 필수다. 현재 TPCx-AI 툴킷은 개발이 거의 완료된 상태이며, 스펙(Specification) 작성, 가격 정책 결정, 세부 메트릭 결정 등이 아직 회의 논제로 남아있다. 신규 벤치마크를 빠르게 도입하기 위해 TTA는 TPCx-AI 분과위원회와 긴밀하게 업무를 추진하고 있으며, 예정대로 2020년 하반기에 TPCx-AI가 제정이 완료되어 2021년에는 TTA에서 TPCx-AI 인증이 나오기를 기대한다.