

ML 기반의 초기 비디오 압축 파라미터 예측 기술 연구

김재일, 문정미*, 황태승, 유상민*, 박성수
SK 텔레콤, *SK 브로드밴드

jaeilkim@sk.com, moonjmee@sk.com, hwangtaeseung@sk.com, a07274@sk.com,
seongsoo@sk.com

A Study on Machine Learning-Based Techniques for Predicting Initial Video Compression Parameters

Jaeil Kim, * Jeong-mee Moon, Taeseung Hwang, *Sangmin Yoo, Seongsoo Park
SK Telecom, *SK Broadband

요 약

본 논문은 클라우드 기반 컴퓨팅을 활용한 비디오 콘텐츠 인코딩의 최적화를 위해 머신러닝 기반의 초기 압축 파라미터 예측 알고리즘을 제안한다. SKB 상용 시스템의 대규모 비디오 데이터를 활용하여 다양한 특성을 고려한 학습 데이터를 생성하고, Amazon SageMaker Autopilot 를 사용한 ML 알고리즘을 적용한다. 예측 알고리즘 적용을 통해, 인코딩 과정에서 평균 12.85%의 시간 절감과 95.80% 예측 정확도 달성 및 클라우드 환경에서의 비용 절감 및 인코딩 효율성 향상에 기여를 하였다.

I. 서 론

최근 클라우드 기반 컴퓨팅은 인프라 관리의 효율성과 컴퓨팅 자원의 유연한 확장성 등의 장점을 통해 다양한 비즈니스 애플리케이션 개발에 널리 활용되고 있다. 특히, IPTV 나 OTT 같은 VOD 서비스에 사용되는 HTTP Live Streaming (HLS)에서는 적응적 스트리밍(Adaptive Streaming) 서비스를 위한 클라우드 환경 기반 트랜스코딩 방법이 개발 및 운용되고 있다. 이러한 과정에서 클라우드의 유연한 리소스 확장 기능을 활용하여, 각각의 콘텐츠를 짧은 길이의 구간인 청크(Chunk)로 분할하고, 이를 최적화된 인코딩 방법으로 처리하는 접근 방식이 제안되었다[1, 2, 3].

Lin[1]은 CRF(Constant Rate-Factor) 파라미터를 활용하여 인코딩된 비트율이 목표 비트율을 초과하지 않는 단계적 인코딩 방식을 제안하였다. 해당 방법은 CRF 값이 목표 비트율을 넘지 않는 경우, CRF 인코딩 과정에서 저장된 통계 정보를 사용하여 VBR(Variable Bitrate)[3] 인코딩을 수행한다. 반대로 목표 비트레이트를 초과하는 경우에는 CRF 파라미터를 재예측한다. 실험 결과 CRF 인코딩 단계를 최대 4 단계로 제한하였으며, 평균적으로 2.52 단계의 CRF 예측이 필요하다. 제안한 방식은 예측된 CRF(Constant Rate-Factor) 파라미터를 기반으로 한 인코딩 결과가 항상 균일한 품질을 보장하지 못한다는

문제점을 가지고 있다. 더욱이 최종 VBR 인코딩 단계까지 포함하면 평균적으로 3.52 단계의 인코딩 과정이 필요하다.

Cock[2] 논문은 복잡한 입력 콘텐츠에 다양한 CRF 파라미터 범위로 인코딩함으로써 콘텐츠 최적의 비트율 계단(Bitrate Ladder)를 도출하는 방법을 제시한다. 해당 방법은 최적 비트율에 맞는 CRF 파라미터를 사용하여 각 청크를 인코딩하고, 결과 비트율에 따라 단일 단계 또는 이중 단계 VBR 인코딩을 수행한다. 각 타이틀별로 비트율 계단을 생성하는 데 상당한 시간이 소요되며, 최종 VBR 인코딩 단계를 포함할 경우 최소 2 단계의 인코딩 과정이 필요하다.

Btv 에 적용한 이전 논문[3]은 CRF 인코딩 과정을 최대 3 회로 제한하는 방식을 제안하고, VMAF[4]를 사용하여 각 인코딩 단계의 영상 품질이 목표 품질에 달성하는지 비교한다. 목표 품질에 도달하지 못할 경우, interpolation 및 extrapolation 예측을 통해 멀티 단계의 인코딩을 진행한다. 제안하는 방법은 단일 단계 CRF 인코딩 대비 약 1.21 배의 처리 시간을 투입하여 고품질의 균일한 인코딩 결과를 달성하였다. 하지만, IPTV 에 한정된 결과이며, 모바일 서비스를 위한 다양한 환경에서는 여전히 2 단계 이상의 멀티 패스로 인코딩을 진행하였다.

본 논문에서는 이전 연구의 문제점을 제시하고, ML 기반의 초기 압축 파라미터 예측을 통한 단계 최소화 방법을 제안한다.

II. 본론

2.1 이전[3] 연구의 문제점

이전의 연구에서는 문제점들을 파악하기 위하여, 3월부터 5월까지 Btv의 상용 인코딩 시스템에서 멀티패스 인코딩 선택 비율을 조사하였다. 조사 결과는 모바일과 IPTV에서 멀티패스 인코딩이 선택된 비율을 도식화한 그림 1과 같다.

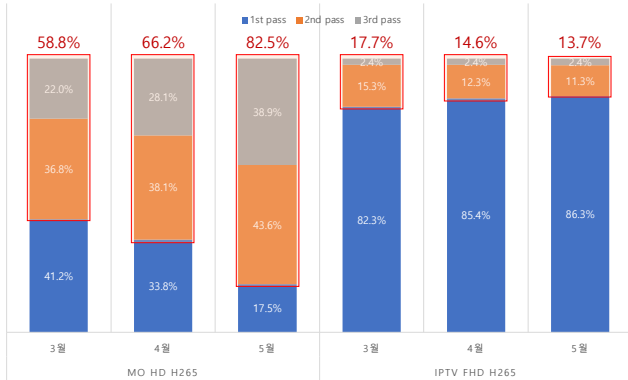


그림 1. IPTV와 모바일의 멀티패스 선택 비율

해당 조사에서 파란색은 첫 번째 패스(1st pass)로 인코딩이 완료된 비율을 나타내며, 주황색은 첫 번째 패스에서 조건을 충족하지 못해 두 번째 패스(2nd pass)까지 실행된 결과를 보여준다. 또한, 세 번째 패스(3rd pass)는 총 세 번의 인코딩 시도 후의 결과이다. 'MO HD H265'는 모바일용 HD 프로파일을, 'IPTV FHD H265'는 IPTV의 FHD 프로파일을 각각 의미한다.

실험 결과를 통해 확인할 수 있듯이, IPTV에서는 주로 화질 달성 여부를 중점으로 1차 패스('1st pass')에서 인코딩이 대부분 완료된다. 반면, 모바일 환경에서는 목표 화질 달성 뿐만 아니라 비트율을 최소화하는 관점에서 인코딩이 진행되므로 멀티패스 인코딩의 비율이 상대적으로 높다. 이 문제를 해결하기 위하여, 본 연구에서는 영상의 특성을 반영한 머신러닝 알고리즘을 통해 초기 인코딩 파라미터를 예측한다. 이를 위해, 학습 데이터를 수집하고 ML 알고리즘을 개발했으며, AWS SageMaker를 사용하여 REST API를 구성하였다. 또한, REST API는 기존 클라우드 시스템에 통합되어 작동하도록 구현되었다. 이렇게 개발된 시스템은 모바일 환경에서의 인코딩 효율을 개선하고, 전반적인 인코딩 프로세스의 최적화에 기여하였다.

2.2 ML 알고리즘 적용을 위한 학습 데이터 생성

머신러닝 알고리즘 개발에 필요한 학습 데이터를 생성하기 위하여, SKB 상용 시스템에서 제공하는

총 5000시간 분량의 영상 데이터를 대상으로 실험을 진행하였다. 이를 위해 5분 간격으로 10초마다 데이터를 추출하였다. 추출된 데이터에서 영상의 특성을 계산하기 위해 사용된 특징 벡터(Feature Vector)는 공간적 인덱스(spatial index), 시간적 인덱스(temporal index), 노이즈 인덱스(noise index), 밝기 변화 인덱스(illumination index)로 구성되었다. 추가적으로, 입력 영상의 프레임 속도(fps), 해상도, 필드 모드(field mode) 등의 정보도 활용하였다.

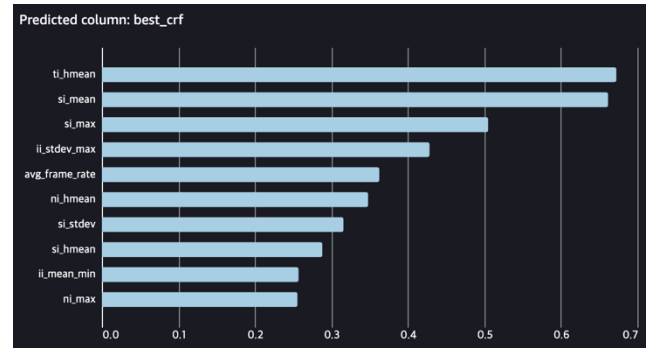


그림 2. CRF와 특징 값의 SHAP 지수

또한, 영상의 특성을 보다 정확하게 반영하기 위해 평균값(mean)뿐만 아니라 최소값(min), 최대값(max), 조화 평균(harmonic mean) 등을 학습 데이터로 활용하였다. 이러한 통계적 값들은 영상의 다양한 특성을 포괄적으로 나타내며, 비디오 압축 파라미터인 CRF(Constant Rate Factor)의 예측에 중요한 역할을 한다. 그림 2에서는 SHAP(Shapley Additive exPlanations) 분석을 통해, 시간 인덱스의 조화 평균(ti_hmean), 공간 인덱스의 평균값(si_mean), 밝기 변화의 최대 표준편차(ii_stdev_max) 등 주요 파라미터들의 영향력을 확인할 수 있다.

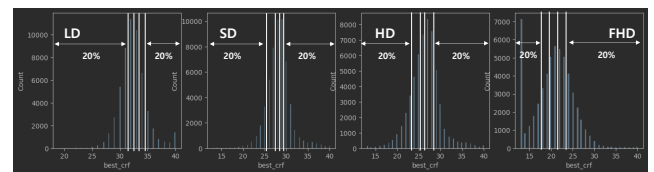


그림 3. 클래스 불균형한 CRF의 확률 분포

본 연구에서 수집된 데이터는 그림 3에 나타나 있는 것처럼, 압축 파라미터인 CRF의 특성상 클래스 불균형(class unbalancing) 문제를 야기할 수 있다. 이러한 문제를 해결하기 위해, 확률 분포를 기반으로 한 퍼센트 비율로 데이터를 샘플링하여, 균일하게 분포된 데이터 세트를 구성하였다. 이러한 접근 방식은 ML 알고리즘의 학습 과정에서 발생할 수 있는 편향을 최소화하고, 보다 정확하고 균형 잡힌 예측 모델을 구축하는 데 도움이 되었다.

2.3 ML 기반 앙상블 학습 적용

본 연구에서는 주어진 학습 데이터를 바탕으로 Amazon SageMaker Autopilot[5] 라이브러리를 사용하여 학습을 진행하였다. 학습 결과는 표 1에 자세하게 나타나 있다. 표 1을 통해 확인할 수 있듯이, 목적함수로는 RMSE(Root Mean Square Error)를 설정하였고, 다양한 알고리즘을 포함하는 앙상블 모델이 가장 뛰어난 성능을 나타냈다. 이 앙상블 모델에는 XGB(XGBoost), GBM(Gradient Boosting Machine), CAT(CatBoost), FASTAI, NN_TORCH(Neural Network with PyTorch), LR(Logistic Regression), RF(Random Forest), XT(Extra Trees) 등이 포함되었다. 추론 시간 또한 1 초 이내로 매우 빠르게 처리되어, 최종적으로 첫 번째 모델을 기반으로 시스템을 구성하였다. 이는 실시간 또는 거의 실시간 처리가 요구되는 응용 프로그램에서도 효과적으로 활용될 수 있음을 보여준다.

표 1. 앙상블 모델 학습 결과

Included model types	Objective: RMSE	추론시간 (Sec)
XGB, GBM, CAT, FASTAI, NN_TORCH, LR, RF, XT	3.635	0.613
XT, RF	3.708	0.16
XT, RF, NN_TORCH, FASTAI, LR	3.666	0.162
XGB, GBM, CAT	3.647	0.093
XGB, GBM, CAT, LR	3.647	0.095
XGB, GBM, CAT, XT, RF, LR	3.647	0.093
XGB, GBM, CAT, XT, RF	3.647	0.093

예측 결과는 그림 4에 자세하게 나타나 있다. 해당 그림에서는 실제 값과 예측된 값 사이의 점들, 그리고 잔차(residual)의 분포를 히스토그램 형태로 보여줍니다. 결과 분석을 통해, 예측에 따른 에러가 0에 가까워지는 것을 확인할 수 있으며, 이는 모델이 데이터의 경향성을 잘 따라가고 있음을 의미한다.

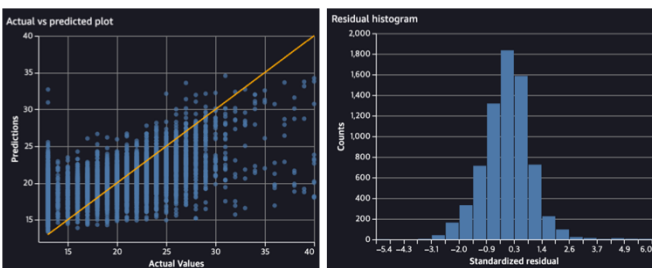


그림 4. 실제값 vs 예측 값과 잔차 히스토그램

2.4 실제 시스템 구현

클라우드 환경에서 실행 시간은 비용과 직접적으로 연관되어 있다. 따라서, 특징 벡터를 최소한의 시간으로 예측하기 위해, 기존에 청크 단위로 나누어 인코딩하는 과정에서 사용되는 Inspector 파라미터의 추출을 병렬 처리 방식으로 진행하였다. 이로 인해 발생하는 시간 증가는 0.33 시간에서 0.37 시간으로, 거의 무시할 수 있는 수준에 불과했다. 이러한 결과는 클라우드 기반의 인코딩 시스템에서 효율성과 비용 효과를 동시에 고려하는 방식이 실용적임을 보여준다.

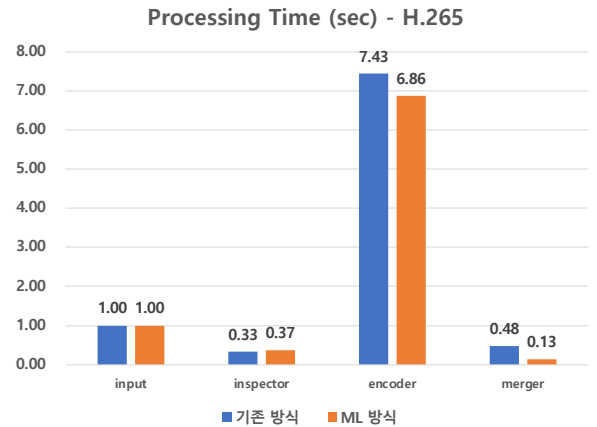


그림 5. 처리 시간 비교 - 기존 방식 vs. ML 방식

2.5 실험 결과

최종 실험은 영화, 드라마, 뮤직비디오, 스포츠, 애니메이션 등 다양한 장르를 포함한 총 100 편의 테스트 콘텐츠를 대상으로 진행되었다. 실험 결과는 표 2에 구체적으로 나타나 있다. 표에서 볼 수 있듯이, 실험 전 평균 1.224의 패스 성공률이 실험 후 1.067로 감소하였다. 이는 인코딩 과정에서 평균적으로 약 12.85%의 시간 절약을 의미한다. 이 결과는 제안된 인코딩 방식이 효율성을 높이는 데 기여함을 보여주며, 다양한 장르의 콘텐츠를 통해 얻은 이러한 결과는 실제 적용 환경에서의 성능 향상을 의미한다.

표 2. ML 알고리즘 예측을 통한 Pass 성공 비율

구분	1pass 통과 비율	2pass 통과 비율	3pass 통과 비율	평균 pass
기존 방법	81.21%	15.20%	3.58%	1.224
제안 방법	95.80%	2.72%	1.48%	1.067

III. 결론

본 논문에서는 기존 연구의 멀티패스 인코딩 방법의 패스를 줄이기 위하여 ML 알고리즘을 개발하였으며,

특징벡트 추출을 위한 계산 시간은 최소화 하면서도 전체 시간 (비용)은 12.85% 절약할 수 있었다.

ACKNOWLEDGMENT

이 과제는 SK Telecom 과 SK Broadband 의 협력 과제로 진행되었으며, 함께해 주신 분들께 감사드립니다.

참 고 문 헌

- [1] Y.-C. Lin, H. Denman, and A. Kokaram, "Multipass encoding for reducing pulsing artifacts in cloud based video transcoding," in IEEE International Conference on Image Processing (ICIP). IEEE, 2015, pp. 907- 911.
- [2] J De Cock, Z Li, M Manohara, A Aaron, "Complexity-based consistent-quality encoding in the cloud," in IEEE International Conference on Image Processing (ICIP). IEEE, 2016, pp. 1484-1488
- [3] J. Moon, J. Kim, T. Hwang, D. Kim, S. Park, "An Uniformalized Quality Encoding in Cloud Transcoding System," in 2023 28th Asia Pacific Conference on Communications. IEEE, 2023, pp. 483 - 488.
- [4] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," The Netflix Tech Blog, 2016.
- [5]<https://docs.aws.amazon.com/sagemaker/latest/dg/autopilot-automate-model-development.html>