



공학석사 학위논문

CNN 기반 유정배치 최적화 성능 향 상을 위한 초기 샘플링 기법

Initial sampling methods to improve CNNbased well placement optimization

2022 년 8 월

서울대학교 대학원

에너지시스템공학부

손 창 균

CNN 기반 유정배치 최적화 성능 향상을 위한 초기 샘플링 기법

Initial sampling methods to improve CNN-

based well placement optimization

지도 교수 최종근

이 논문을 공학석사 학위논문으로 제출함 2022 년 8월

> 서울대학교 대학원 에너지시스템공학부 손 창 균

손창균의 공학석사 학위논문을 인준함 2022 년 8 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

유전개발계획의 주요 설계변수에는 유정배치와 운영조건이 있고 그 중에서 유정배치는 유가스의 회수율 및 사업의 경제성에 매우 중요하다. 이와 관련된 저류층 인자들의 경우 불균질적으로 분포하고 비선형적인 상관관계를 가지기 때문에 직관만으로는 복잡한 유체 거동을 파악하기 어렵고 시뮬레이션과 수학적 알고리즘을 통해 최적화를 수행한다. 그러나 저류층 시뮬레이션과 추계학적 광역탐색 알고리즘은 많은 계산량이 요구되기 때문에 이를 대체하기 위한 프록시 모델의 개발이 필요하다.

답러닝 기반 프록시 모델의 경우 학습 자료의 한계로 인해 유정배치 최적화가 진행되면서 예측성능이 저하되는 문제가 있다. 이를 해결하기 위해 최적화 과정의 중간에 프록시 모델을 재학습시키는 방법이 사용되었지만 이는 그 시점과 기준이 불명확하다는 한계점이 있다.

본 연구에서는 프록시 모델이 최적화 전과정에서 안정적인 예측 성능을 확보하기 위한 초기 학습 자료 샘플링 방법에 대해 제시하고자 한다. 해당 방법은 두 단계에 걸쳐 샘플링을 수행한다. 첫 번째 단계에서 얻은 샘플들로부터 유정의 배치(개수, 종류, 위치)에 대한 잠재적인 생산성(품질) 정보를 취득한다. 그리고 이 품질 정보를 확률로서 반영하여 두 번째 샘플들을 생성한다. 이 때 품질 정보로는 NPV(Net Present Value)값과 유정자료인 WOPT(Well Oil Production Total), WWIT(Well Water Injection Total)을 사용하였다.

제안한 방법의 효과성을 검증하기 위해 여러 샘플링 방법에 대해 CNN(Convolutional Neural Network) 프록시 모델을 구축하고 3D 저류층 모델인 PUNQ-S3와 Egg model에서 PSO(Particle Swarm Optimization) 알고리즘으로 3회씩의 최적화를 수행하여 그 결과를 비교하였다. 유정배치를 임의로 결정하는 random sampling 방법은 최적해의 예측 오차 평균이 PUNQ-S3 model에서 11.17%, Egg model에서 6.16%인 반면 제안한 방법은 각각 2.14%, 1.11%로 향상된 최적화 성능을 보여주었다.

주요어 : 유정배치 최적화, 프록시 모델, CNN, 2-stage sampling, 품질지도

학 번:2020-22103

제	1 장	서 론	1
제	2 장 제 1 제 2 제 3	이론적 배경 절 Convolutional Neural Network 절 Streamline Simulation 절 Particle Swarm Optimization	8
제	3 장 제 1 제 2 제 3 제 4	연구 방법 절 저류층 모델 개요 절 유정배치 샘플링 절 CNN 프록시 모델 구축 절 프록시 모델을 이용한 유정배치 최적화	20 20 28 36 41
제	4 장 제 1 제 2	연구 결과 절 PUNQ-S3 모델 절 Egg 모델	 43 43 50
제	5 장	결 론	58
참.	고문헌	<u>1</u>	60
Ab	strac	xt	65

표 목차

[Table 3.1]	Petrophysical parameters of PUNQ-S3 model	.22
[Table 3.2]	Petrophysical parameters of Egg model	.23
[Table 3.3]	Simulation parameters	.36
[Table 3.4]	Structure of CNN proxy model	.39
[Table 4.1]	The NPVs of final solutions in PUNQ-S3 model	.50
[Table 4.2]	The NPVs of final solutions in Egg model	.57

그림 목차

[Figure 1.1] Integrated reservoir modeling1
[Figure 1.2] 3D representation of the 2D quality map6
[Figure 2.1] The typical structure of DNN
[Figure 2.2] The example of convolution10
[Figure 2.3] The example of pooling11
[Figure 2.4] Types of activation function12
[Figure 2.5] The typical structure of CNN13
[Figure 2.6] The process of streamline simulation14
[Figure 2.7] Schematic of streamline tracing in a 2D grid15
[Figure 2.8] A well placement and corresponding TOF maps17
[Figure 2.9] Schematic of updating a particle in PSO algorithm19
[Figure 3.1] The top structure of PUNQ-S3 model21
[Figure 3.2] The log permeability of PUNQ-S3 model21
[Figure 3.3] The log permeability of Egg model23
[Figure 3.4] Numerical simulation of a 1D waterflooding24
[Figure 3.5] Welge technique at breakthrough when M >125
[Figure 3.6] Modification of relative permeability curves27
[Figure 3.7] The example of random sampling29
[Figure 3.8] The analysis on well numbers and types of primary samples
[Figure 3.9] Quality map obtained from the expected NPV32
[Figure 3.10] CDFs of well numbers and positions

[Figure 3.11] Schematics of proposed sampling methods35
[Figure 3.12] Schematic of CNN proxy model
[Figure 3.13] Flow chart of PSO algorithm using proxy model42
[Figure 4.1] NPV histogram of samples (PUNQ-S3)44
[Figure 4.2] The NPV boxplot of each sampling method (PUNQ-S3) 45
[Figure 4.3] The well location distribution in PUNQ-S3 model46
[Figure 4.4] The performance of CNN proxy models in PUNQ-S3
[Figure 4.5] Optimal well placements in PUNQ-S3 model48
[Figure 4.6] The NPVs of optimal solutions in each generation (PUNQ-S3)
[Figure 4.7] NPV histogram of samples (Egg)51
[Figure 4.8] The NPV boxplot of each sampling method (Egg)52
[Figure 4.9] The well location distribution in Egg model53
[Figure 4.10] The performance of CNN proxy models in Egg54
[Figure 4.11] Optimal well placements in Egg model55
[Figure 4.12] The NPVs of optimal solutions in each generation (Egg)

제1장 서론

석유가 매장되어 있는 유전을 어떻게 개발할 것인지에 대한 계획은 석유의 회수율 및 사업의 경제성에 매우 중요하다. 유전개발사업은 그 규모와 사업의 특성상 초기 투자비용이 크고 추후 수정이 어렵기 때문에 초기에 최적의 계획을 수립하는 것이 중요하다. 유전개발계획의 주요 설계변수들로는 유정의 개수, 위치, 종류, 운영조건 등이 있고 개발의 대상이 되는 저류층의 인자들을 고려하여 순현재가치(NPV, net present value) 또는 누적오일생산량이 최대가 되도록 이러한 설계변수들을 결정한다.

이 때, 저류층 인자들은 그 수가 많을 뿐만 아니라 불균질적으로 분 포하고 인자 간에 비선형적인 상관관계를 가지고 있기 때문에 직관만으 로는 저류층 내부에서의 복잡한 유동을 파악하기 어렵다. 컴퓨터 성능의 향상과 빅데이터 처리 기술의 등장은 Figure 1.1처럼 복합적인 저류층 모델을 구축하고 저류층 시뮬레이션으로 저류층의 반응을 현실적으로 모사할 수 있게 만들었다. 이에 저류층 시뮬레이션과 수학적인 최적화 알고리즘을 통해 최적화를 수행하는 방법이 가능해졌고 현재까지 많은 연구가 진행되어 왔다(Bittencourt와 Roland, 1997; Litvak 등, 2007; Isebor 등, 2014a).



Figure 1.1 Integrated reservoir modeling (Jaber et al., 2019)

유전개발계획 최적화 문제는 유정배치 최적화와 유정운영 최적화 그리고 둘을 결합한 동시최적화 문제로 나뉜다(Udy 등, 2017). 유정배치 최적화를 통해 유정의 위치, 종류, 개수를 결정하고 유정운영 최적화를 통해 각 유정의 시간별 공저압력 혹은 유량을 결정한다. 최적화 변수의 특성에 따라 최적화 알고리즘의 성능이 달라지기 때문에 적절한 최적화 알고리즘을 선정하는 것이 중요하다.

유정배치 최적화의 경우에는 최적화 변수의 개수가 많고 격자별로 할당되기 때문에 이산적이라는 특징을 가진다(Onwunalu와 Durlofsky, 2010). 그리고 유체투과율과 같은 저류층 인자들은 불균질하게 분포하기 때문에 목적함수가 부드럽지 않고 무수히 많은 지역해를 가지게 된다. 유정배치 최적화 문제에 적합한 알고리즘으로는 기존의 기울기 기반 알고리즘을 보완한 adjoint gradient-based method (Sarma와 Wen, 2008; Forouzanfar 등, 2010) 혹은 추계학적 광역탐색기법인 PSO(Particle Swarm Optimization) 알고리즘 (Onwunalu와 Durlofsky, 2010; Cheng 등, 2012; Harb 등, 2020), genetic 알고리즘(Yeten 등, 2003; Tukur 등, 2019), CMA-ES (Covariance Matrix Adaptation Evolution Strategy)알고리즘 (Bouzarkouna 등, 2012; Awotunde와 Carlos, 2014), harmony search 알고리즘(Afshari 등, 2011) 등이 사용된다.

그리고 유정운영 최적화에서는 최적화 변수가 시계열 형태의 연속 적인 실수값을 가지고 저류층, 파이프라인, 생산설비에서의 제약조건이 복잡하게 얽혀 있다는 특징을 가진다(Kosmidis 등, 2004). 이에 유정운영 최적화 문제에 사용되는 알고리즘으로는 local search 알고리즘(Ciaurri 등, 2012; Jansen과 Durlofsky, 2017)과 gradientbased 알고리즘 (Carroll과 Horne, 1992; Dutta-Roy와 Kattapuram, 1997; Chen과 Reynolds, 2018) 등이 있다.

동시최적화 문제에서 주 관심사는 서로 다른 특성을 가지는 유정의 배치와 운영조건을 최적화 변수로서 함께 고려하는 것이다. 이를 해결하 기 위해 다양한 알고리즘들을 결합한 하이브리드 알고리즘 연구가 진행되었다. Bellout 등(2012)은 기울기 기반 알고리즘인 sequential quadratic programming과 지역탐색 알고리즘인 pattern-search 알고리즘을 결합하는 방법을 제안하였다. 그리고 Yang 등(2018)은 추계학적 광역탐색 알고리즘인 DE(Differential Evolution) 알고리즘과 지역탐색 알고리즘인 MADS(Mesh Adaptive Direct Search)를 결합하여 기존의 생산정을 주입정으로 전환하는 시나리오를 추가적으로 고려한 동시최적화 연구를 수행하였다.

이번 연구에서는 유정배치 최적화에 초점을 맞추었으며 최적화 결과가 지역해에 수렴하는 문제를 방지하기 위해 대표적인 추계학적 광역탐색 알고리즘인 PSO 알고리즘을 사용하였다. PSO 알고리즘은 간단하여 구현하기가 쉽다는 장점이 있지만 유정배치 최적화와 같은 복잡한 문제에 적용하기 위해서는 수정이 필요하다. 입자해가 일반적인 유정배치 규칙이나 제약조건에 부합하도록 PSO를 수정하는 시도가 많이 있었으며 Harb 등(2020)은 유정 인근에 다른 유정이 배치되지 않도록 탐색 공간을 제한하는 방법을 사용하였고 Isebor 등(2014b)과 Kim 등(2018)은 필터기법을 사용하여 제약조건을 만족시키는 해집단의 영역으로 해들을 이동시키는 방법을 사용하였다. 이 외에도 많은 수의 변수들을 효과적으로 최적화하기 위해 Kim 등(2019)은 최적화 변수들을 연결시키는 매개변수의 도입과 함께 군집을 분할하여 개별적으로 최적화를 수행하는 cooperative micro-PSO 알고리즘을 제안하였고 Isebor 등(2014a, 2014b)은 PSO-MADS 하이브리드 알고리즘을 최적화에 사용하였다.

하지만 PSO를 비롯한 추계학적 광역탐색 알고리즘은 광역해에 수렴하기까지 많은 횟수의 시행이 요구된다. 특히 저류층 시뮬레이션은 각 격자사이에서의 유동을 수치해석적으로 계산하므로 저류층 모델의 격자 수가 많을 경우에 상당한 시간이 소요된다. 이러한 상황을 극복하기 위해 parallel processing, local grid refinements 등 저류층 시뮬레이션의 계산속도를 증진시키려는 여러 시도가 있었으나 큰 효과를 보여주지 못하였고 최근에는 저류층 시뮬레이션을 프록시 모델로 대체하는 연구가 큰 이목을 받고 있다(Jaber 등, 2019).

프록시 모델은 특정 인자들만을 입력받아 저류층 반응을 모사하는 수학적, 통계적 모델로 정의되고(Denney, 2010) 빠른 계산이 가능 하다는 것과 주요 인자들간의 상호작용과 저류층의 거동에 미치는 영향을 심층적으로 파악할 수 있다는 장점을 지닌다. 하지만 프록시 모델 구축에 사용된 저류층 모델과 상이한 모델을 대상으로는 예측 성능이 떨어질 수 있어 사전 검증과정이 필수적이다.

프록시 모델은 다양한 학문, 산업 분야에서 적용되고 있으며 저류층 공학 분야에서는 주로 확률 예측, 위험도 분석, 민감도 분석, 히스토리 매칭, 생산 최적화 등에 활용된다. 프록시 모델의 종류에는 물리 기반 프록시 모델과 데이터 기반 프록시 모델이 있다. 물리 기반 프록시 모델은 저류층 내의 유체 거동을 간단한 형태로 모사한 후 계산을 수행하는 방식으로 차수감소모델(Jansen과 Durlofsky, 2017), fast marching method(Zhang 등, 2016), 유선 시뮬레이션(Afshari 등, 2011) 등이 있다. 이러한 모델들은 간략화된 수식을 사용하기 때문에 기존의 저류층 시뮬레이션과 비교하여 계산량 측면에서 이점이 크면서도 물리 이론에 기반하여 모델을 구축하므로 물리적으로 타당한 계산 결과를 얻을 수 있다. 그러나 모델을 너무 간소화하게 되면 계산 결과가 부정확해지기 때문에 정확도와 계산량 간에 trade-off 관계가 존재한다.

데이터 기반 프록시 모델은 자료들 간에 내재된 관계를 직접 이용하여 저류층 반응을 모사하는 모델이며 모델을 구축하는 방법으로는 통계적 기법 또는 인공지능 기법이 사용된다. 그 예로 Goodwin(2015) 은 통계적 기법인 DOE(Design Of Experiments)와 RSM(Response Surface Methodologies)를 이용하여 저류층 반응에 영향을 주는 주요 인자들을 판명하고 그 영향과 인자들 간의 상호작용으로부터 프록시 모델인 반응표면모델을 구축하였다. 그리고 Alenezi와 Mohaghegh (2016)은 ANN(Artificial Neural Network) 기법을 사용하여 수공법 (waterflooding)하에서의 저류층 시뮬레이션의 반응을 학습하였고 각 격자의 압력과 물포화도를 예측하는 연구를 수행하였다.

특히 ANN과 같은 머신러닝 기법으로 구축한 프록시 모델의 경우 저류층 인자들간의 복잡한 상호작용에 대한 구체적인 해석 없이도 모델 구축이 가능하고 응용이 간단하다는 장점이 있다. 하지만 머신러닝 모델은 내부 인자들의 수학적 관계를 이해하기 어려운 블랙박스 모델이기 때문에 모델 자체를 개선하기 어려워 학습 데이터의 질이 모델의 예측성능에 매우 큰 영향을 미친다. 따라서 효과적인 학습을 위해 저류층 시뮬레이션에서의 정적인자(유체투과율, 유정위치 등)와 함께 저류층 대류의 정량적 정보를 제공하는 동적 자료들을 학습 자료에 추가하여 예측성능을 증진시키려는 연구들이 진행되었다. Nwachukwu 등(2018)은 fast marching method로 유정으로부터 각 격자까지 압력이 전파되는 시간인 DTOF(Diffusive Time Of Flight)를 계산하였고 이를 학습 자료로 사용하여 프록시 모델이 저류층의 공간적 연결성과 불확실성을 반영할 수 있도록 하였다. 그리고 Kim 등(2020)은 유선 시뮬레이션으로 계산한 TOF(Time Of Flight) 정보를 프록시 모델 학습에 사용하였고 프록시 모델이 예측한 NPV값을 목적함수로 하여 유정배치를 최적화하는 연구를 수행하였다. 이들은 최적화 과정 중간에 프록시 모델을 재학습시켜 프록시 모델의 예측 성능이 최적해 부근에서 저하되는 문제를 해결하였지만 재학습을 하는 시점과 기준이 불명확하고 재학습에 소요되는 시간으로 인해 프록시 모델의 이점이 경감된다는 한계가 존재하였다.

Lee(2021)는 초기 학습 자료의 샘플링 방법 자체를 개선함으로써 위의 문제를 보완하고자 하였고 여러 방법으로 유정의 배치를 샘플링 하여 최적화 성능을 비교하였다. 그리고 그 결과로부터 높은 NPV를 나타내는 유정배치를 샘플링하는 것이 프록시 모델의 유정배치 최적화 성능에 있어 중요하다는 것을 밝혀내었다. 본 연구에서는 이에 착안하여 품질지도(quality map)의 개념을 응용한 유정배치 샘플링 방법을 제안하고자 한다. 품질지도는 Figure 1.2와 같이 저류층 모델에서 각 격자의 잠재적인 생산성을 나타내는 2D 자료로 적은 수의 시뮬레이션 결과만을 가지고 효과적으로 유정의 위치를 선택하기 위해 사용된다 (Badru, 2003; da Cruz 등, 2004; Taware 등, 2012).



Figure 1.2 3D representation of the 2D quality map (Badru, 2003)

본 연구의 목표를 요약하자면 다음과 같다.

- 지류층 각 격자의 잠재적인 생산성을 나타내는 품질지도를 바탕으로 2-stage sampling을 수행하여 생산성이 높은 유정배치 샘플들을 생성
- 2. 주어진 유정배치에 대해 NPV를 예측하는 프록시 모델을 구축
- 개발한 프록시 모델이 최적화 전과정에서 합리적인 예측성능을 유지하는지 검증

본 연구에서 사용한 샘플링 방법은 두 단계에 걸쳐 샘플링이 이루어지며 첫 번째 단계에서는 난수를 생성하여 유정의 위치, 종류, 개수를 임의로 선정한다. 첫 번째 유정배치 샘플들로부터 유정의 개수와 종류 그리고 각 생산정 및 주입정의 위치에 대한 품질지도를 통계적으로 계산하여 만들고 이 품질지도를 확률로서 반영하여 두 번째 유정배치 샘플들을 생성한다.

그 후 유정배치와 TOF map을 입력받아 NPV를 예측하는 CNN 프록시 모델을 구축하고 CNN 프록시 모델이 예측한 NPV값을 PSO 알고리즘의 목적함수로 하여 최적의 유정배치를 찾는다. 이 때 각 위치의 품질을 평가하는 기준으로써 NPV값 또는 유정자료를 사용할 수 있으며 본 연구에서는 각각의 품질지도를 사용한 경우에 대해 최적화를 수행하고 그 결과를 비교할 것이다.

1장에서는 유정배치 최적화 문제의 중요성과 특징, 현재까지 진행된 관련 연구들, 저류층 시뮬레이션을 대체할 수 있는 프록시 모델의 개념, 그리고 초기 학습 데이터 샘플링의 중요성을 소개한다. 2장에서는 본 연구에서 사용된 기법들의 이론적 배경들을 정리할 것이다. 3장에서는 본 연구에서 제안한 샘플링 방법과 전체적인 절차에 대해 서술하고 4장에서 두 가지의 3D 저류층 모델에 적용해본 결과를 정리하고자 한다. 5장에서는 본 연구의 결론에 대해 논의할 것이다.

제 2 장 이론적 배경

제 1 절 Convolutional Neural Network

CNN(Convolutional Neural Network)는 이미지 자료 학습에 특화된 딥러닝 알고리즘의 일종으로 기존의 DNN(Deep Neural Network)의 한계를 극복하고자 고안되었다(Lecun 등, 1998). DNN은 입력층과 출력층 사이에 다수의 은닉층으로 구성된 구조를 가지고 있으며 각 층의 뉴런들은 인접한 층의 뉴런들과 모두 연결되어 있다(Figure 2.1). 이러한 특징으로 인해 DNN의 은닉층은 전연결층 (fully connected layer)이라고도 불린다. DNN에서 뉴런 간의 연결은 가중치, 편향, 활성화 함수로 표현되며 각 층에서는 식 2.1과 같은 수학적인 연산이 수행된다.

$$h(I) = f(w \times I + b) \tag{2.1}$$

여기서 I는 이전 층에서의 계산 결과이며 각 층은 이 I를 입력 받아 가중치(w)와 편향(b) 값들로 행렬계산을 수행하고 그 결과를 활성화 함수에 대입하여 나온 결과 h(I)를 다음 층으로 입력시킨다.



Figure 2.1 The typical structure of DNN

DNN은 각 층의 뉴런들이 1차원 배열로 구성된 구조를 가지고 있기 때문에 이미지 자료를 DNN에 학습시키기 위해서는 2차원 배열을 1차원 배열로 변형시켜야 하고 이 과정에서 본래 자료의 공간적, 지역적 정보가 손실된다. 따라서 손실되는 정보들로 인해 신경망이 자료의 특징을 제대로 추출할 수 없게 되고 그 결과 학습 성능이 낮아지는 문제점이 발생한다. CNN은 이러한 점에 착안하여 이미지의 형태를 그대로 보존한 상태로 연산을 수행하며 이미지의 공간 정보를 효과적으로 학습하는 것이 가능하다. 비유적으로 표현하자면 DNN이 인간의 뇌신경을 모방한 딥러닝 기법이었다면 CNN은 인간의 시신경을 모방한 기법이라고 할 수 있다.

입력 자료가 2차원의 이미지 자료이므로 CNN에는 이미지 자료에 대한 연산을 수행하는 컨볼루션층과 풀링층이 존재한다. 컨볼루션층에는 3x3, 5x5 등의 작은 크기를 가지는 행렬인 필터(커널)들이 존재한다. 이 필터가 입력된 2차원 배열 위에서 움직이면서 내적계산을 반복하여 수행함으로써 새로운 2차원 배열을 만들어낸다. 곧 컨볼루션은 입력 이미지의 모든 영역에 같은 필터를 반복 적용해 지역적인 특징들을 추출하는 것이고 그 결과로 얻어진 이미지를 특징지도(feature map)라 한다. 입력 이미지 I에 대해 pxq 크기의 r번째 필터 k'로 구해진 특징지도를 h'이라 할 때, h'의 특정 지점 (x, y)에서의 값은 식 2.2로 계산된다.

$$h_{x,y}^{r}(I) = (k^{r} * I)_{x,y} = f\left(\sum_{i=1}^{p} \sum_{j=1}^{q} k_{i,j}^{r} I_{x+i,y+j} + b\right)$$
(2.2)

Figure 2.2는 5x5의 입력 이미지에 대해 3x3의 필터를 가로 또는 세로로 한 칸씩 이동시켜가며 컨볼루션 연산을 수행한 예시를 보여준다.



Figure 2.2 The example of convolution

컨볼루션 연산 후에는 이미지의 크기가 줄어들고 그 정도는 필터의 크기와 필터의 이동간격(stride)에 의해 결정된다. 컨볼루션은 입력 이미지가 여러 개일 때도 가능하며 입력 이미지들은 높이 방향으로 배열되어 3차원 텐서 형태가 된다. 이 때 필터도 동일한 높이를 가지는 3차원 텐서를 사용하여 특징지도는 마찬가지로 2차원 배열이 된다. 컨볼루션층이 여러 개의 필터를 가질 경우 각 필터로 얻은 특징지도가 다시 높이 방향으로 배열되어 새로운 3차원 텐서가 만들어진다. 정리하자면 mxn의 입력 이미지 u개에 대해 pxq의 필터 v개를 각각 s만큼 이동시키면서 컨볼루션 연산을 수행하면 $\left(\frac{m-p}{s}+1\right) \times \left(\frac{n-q}{s}+1\right)$ 크기의 특징지도가 v개가 생성된다.

풀링층은 이미지의 크기를 줄임으로써 계산량을 줄이고 특정 feature를 강조하는 역할을 한다. 풀링의 방법에는 선택 영역에서의 최댓값을 해당 영역의 대푯값으로 설정하는 max pooling과 평균값을 가져오는 average pooling 등이 있다(Figure 2.3). CNN에서는 주로 max pooling이 사용되며 이는 뉴런이 가장 큰 신호에 반응하는 것과 유사하다고 할 수 있다. 그 외에도 풀링 과정을 통해 노이즈 감소, 이미지 요소들의 이동 및 회전에 의한 영향 완화, 과적합 방지 등의 이점을 얻을 수 있다.

10



Figure 2.3 The example of pooling

입력 이미지 자료가 CNN의 각 층에서 연산을 거쳐 출력층에 도달하는 과정을 피드포워드(feedforward)라 한다. 그리고 CNN이 예측한 결과와 정답 자료의 차이를 나타내는 손실함수(loss function)의 값이 최소가 되도록 CNN의 파라미터들을 학습시키며 이를 역전파 (backpropagation)라 한다. 이 때 학습되는 파라미터들에는 식 2.2에서 필터를 구성하는 값들인 $k_{i,j}$ 과 편향값 b가 있다. 역전파 과정은 손실함수의 각 파라미터들에 대한 편미분값을 계산하고 기울기 기반 기법으로 파라미터들을 갱신한다. 대표적으로 사용되는 기울기 기반 기법에는 현재 위치와 이전 진행 방향을 고려하여 다음 진행 방향과 보폭을 유연하게 조절하는 Adam(Adaptive moment estimation)기법이 있다(Kingma와 Ba, 2015).

CNN 파라미터들을 효과적으로 학습시키기 위해 컨볼루션과 풀링 외에도 여러 보조적인 연산이 수행된다. Figure 2.2를 보면 알 수 있듯 컨볼루션을 수행하면 본래 이미지보다 사이즈가 작아진다. 그렇기 때문에 컨볼루션층들을 지날수록 가장자리의 정보들이 사라지고 가능한 컨볼루션 연산의 수가 제한된다. 이러한 문제점을 해결하기 위해 패딩이 수행된다. 패딩은 입력 이미지의 가장자리에 특정 값으로 설정된 픽셀들을 추가함으로써 이미지의 크기를 유지시킨다. CNN에서는 주로 이의 값을 추가하는 zero-padding을 사용한다.

11

활성화 함수는 선형 함수인 컨볼루션에 비선형성을 추가하기 위해 사용한다. 활성화 함수로는 다양한 종류의 함수가 사용되며(Figure 2.4) 초기에는 sigmoid나 tanh가 사용되었으나 두 함수들은 끝부분에서 미분값이 0에 수렴하기 때문에 기울기 소실이 발생하고 파라미터의 학습이 잘 되지 않는다는 문제가 생긴다. 따라서 CNN의 활성화 함수로는 음수값이 입력되었을 때 0을 출력하고 0과 양수값은 그대로 출력하는 ReLU(Rectified Linear Unit) 함수가 주로 사용되며 최근에는 Leaky ReLU나 ELU와 같이 ReLU를 발전시킨 함수들도 시도되고 있다.



Figure 2.4 Types of activation function

이 외에도 기울기 소실 문제와 과적합 문제를 방지하기 위해 각 층의 입력값이 평균 0과 분산 1의 분포를 따르도록 변환하는 배치 정규화와 역전파 과정에서 일부 노드를 의도적으로 배제하는 드롭아웃 등이 사용된다.

위에서 소개한 각종 연산들이 수행된 이미지 자료는 1차원 배열로 변환되어 전연결층에 입력된다. 그리고 전연결층에서는 입력된 이미지의 label을 분류하는 classification 혹은 특정 값을 예측하는 regression이 수행된다. 결론적으로 CNN의 전체적인 구조는 Figure 2.5와 같고 CNN은 컨볼루션과 풀링을 반복하여 이미지 자료에서 특징을 찾아내고 그 특징을 전연결층에 보내 출력값을 예측하는 프록시 모델이다.



Figure 2.5 The typical structure of CNN

제 2 절 Streamline Simulation

유선 시뮬레이션(Streamline simulation)은 저류층에서의 3차원 유체 유동을 각 유선들을 따르는 여러 개의 1차원 유체 유동으로 모사하는 시뮬레이션 기법이다. 유선 시뮬레이션은 유선이라는 유체역학 적 개념을 사용하여 저류층의 유동을 간단한 형태로 근사하여 계산하는 것이기 때문에 넓은 의미에서 물리정보 기반 프록시 모델이라고 할 수 있다.

이러한 유선 시뮬레이션은 기존의 저류층 시뮬레이션과 비교하였을 때 여러 장점을 가진다. 그 중 하나는 계산속도가 빠르고 계산과정에서 수치분산오차가 억제된다는 것이다. 이는 1차원 유동이 3차원 유동보다 계산이 간단하고 수치해석적으로 해를 구하는 것이 아니라 이론해를 직접 구하기 때문이다. 또한 1차원 좌표계로 변환하는 과정에서 저류층의 불균질성이 고려되기 때문에 격자의 크기나 방향에 의한 영향을 받지 않는다(Datta-Gupta, 2000). 이로 인해 유선 시뮬레이션 은 대상 저류층이 크고 불균질성이 심할수록 효과적이다.



Figure 2.6 The process of streamline simulation

유선 시뮬레이션이 수행되는 과정은 위의 Figure 2.6과 같다. 유선을 추적하기 위해서 먼저 유동방정식을 풀어 각 지점에서의 압력을 계산한다. 그리고 얻어진 압력장을 Darcy 방정식에 대입하여 속도장을 구한다. 유선은 유체가 흐르는 각 지점에서 속도벡터의 접선방향을 이은 선으로 정의되므로 속도장으로부터 각 격자를 지나는 유선을 추적할 수 있다. 세부적으로는 격자 내부에서 속도가 x방향과 y방향으로 독립적, 선형적으로 변한다고 가정하여 격자 내에서의 속도값을 선형보간법으로 계산한다. 따라서 격자로 들어오는 입구좌표를 알게 되면 출구좌표를 계산할 수 있게 되고 이 출구좌표가 다음 격자에서의 입구좌표가 된다. 이 과정을 반복함으로써 유선을 그리고 Figure 2.7은 한 격자 내에서의 유선을 추적하는 모식도이다.



Figure 2.7 Schematic of streamline tracing in a 2D grid (Batycky 등, 1997)

만약 비압축성의 단상 유체가 고정된 운영조건에서 유동하면 유선은 변하지 않는다. 따라서 유선 시뮬레이션을 짧은 기간 동안만 수행해도 되기 때문에 계산효율이 증가한다. 반면 압축성이 큰 유체의 경우 시간에 따라 유동성이 변하기 때문에 저류층의 반응을 정확하게 모사하기 위해서는 유선의 잦은 갱신이 필요하다.

유선을 구한 후에는 좌표변환에 사용되는 물리량인 TOF(Time Of Flight)를 계산한다. TOF는 유선을 따라 흐르는 유체가 각 격자까지 도달하는데 걸린 시간을 의미하며 식 2.3으로 계산된다.

$$\tau = \int_0^s \frac{\phi d\zeta}{|\vec{u}|} \tag{2.3}$$

여기서 s는 유선(ζ)내의 한 지점이고 ϕ 는 공극률, \vec{u} 는 각 지점에서의 속도장을 의미한다.

Figure 2.7을 보면 격자 내에서의 유선의 경로와 각 지점에서의 속도값을 알 수 있으므로 유선이 해당 격자로 들어가서 빠져나오기까지 소요된 거주시간을 계산할 수 있다. 유선 시뮬레이션은 유선이 통과한 격자들에 대해 격자마다 머물렀던 거주시간의 누적값으로 TOF값을 계산한다(Pollock, 1988). TOF는 유정으로부터 거리가 멀거나 해당 위치까지의 유동이 어려울수록 증가한다. 이러한 성질로 인해 TOF는 단순한 물리적 거리뿐만 아니라 유체투과율, 공극률과 같은 유동용이도 및 유정과의 수리학적 연결성을 대표하는 거리인자로 활용할 수 있다(Batycky 등, 1997).

유선 시뮬레이션은 TOF를 계산한 후 3차원 좌표계를 1차원 TOF 좌표계로 변환하고 1차원 advection-diffusion equation의 이론해 로부터 물포화도와 같은 물리량을 각 격자별로 계산한다. 그러나 본 연구에서는 유선 시뮬레이션으로 저류층의 반응을 계산하는 것보다는 TOF map 자체를 얻는 용도로 사용하였다.

TOF map은 유선의 시작점이 생산정(P)인지 주입정(I)인지에 따라 두 개의 TOF map(TOFP, TOFI)을 얻을 수 있으며 이들은 각각 배수구역과 배유구역을 나타낸다. Figure 2.8은 임의의 유정배치에 대해 TOFP와 TOFI 지도가 어떻게 나타나는지를 보여준다.



(a) 유정배치 (빨: 생산정, 검: 주입정)



(b) TOFP 지도



(c) TOFI 지도 Figure 2.8 A well placement and corresponding TOF maps

유선은 저류층 모델의 구조와 불균질성 그리고 유정들의 위치를 모두 고려하여 추적되므로 Figure 2.8에서와 같이 TOF 지도는 저류층 모델 정보와 유체 유동에 관한 동적 정보를 반영한다. 본 연구에서는 이러한 TOF 지도를 CNN 프록시 모델에 학습시킴으로써 프록시 모델의 예측 성능과 저류층의 불확실성에 대한 강건성을 향상시키고자 하였다 (Kim 등, 2020).

제 3 절 Particle Swarm Optimization

PSO는 해집단을 이용하는 추계학적인 광역탐색 알고리즘이다. 이 알고리즘은 Kennedy와 Eberhart(1995)에 의해 제안되었고 집단 생활을 하는 생물의 행동방식을 모방하였다. PSO는 무작위로 생성된 Np개 후보해들로 고차원 탐색영역을 탐색해 나간다. 이 때 개별해를 입자(particle)라고 하며 전체 집단을 군집(swarm)이라 한다. 입자들은 각자가 기억하고 있는 탐색과정과 군집이 모은 정보를 조합하여 반복적으로 탐색을 수행한다. 이러한 방법은 계산량이 많지만 각 입자들이 정보를 상호 교환하면서 최적화를 수행하기 때문에 하나의 입자가 지역해에 빠지더라도 군집 전체적으로는 광역해에 수렴할 수 있다는 장점을 가진다. 현재 많은 최적화 문제에 이 알고리즘이 적용되고 있으며 Onwunalu와 Durlofsky(2010), Kim 등(2020)이 유정의 위치와 종류를 최적화하기 위해 사용하였다.

PSO에서 각 입자해들은 속도의 개념에 기반하여 자신의 위치를 업데이트한다. k번째 세대에서 i번째 입자의 위치벡터(x_i^k)는 식 2.4에 의해 다음 세대로 갱신되고 이 때의 속도벡터(v_i^{k+1})는 식 2.5에 의해 구해진다.

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{2.4}$$

$$v_i^{k+1} = w \cdot v_i^k + c_1 \cdot r_1 \cdot \left(x_{i,pbest}^k - x_i^k \right) + c_2 \cdot r_2 \cdot \left(x_{i,gbest}^k - x_i^k \right)$$
(2.5)

식 2.5를 보면 속도벡터는 세 요소를 통해 결정되고 이들을 각각 관성(inertia), 인지(cognitive), 사회(social) 요소라 한다. 관성항은 기존 속도벡터를 반영하고 인지항에서는 개별 입자해인 x^k_i가 지금까지 탐색한 해들 중 가장 좋은 해인 x^k_{i,pbest} (personal best)를 반영한다. 그리고 사회항은 군집 전체가 탐색한 해들 중 가장 좋은 해인 x^k_{i,gbest} (global best)를 반영한다.

그리고 w, c₁, c₂는 각 항에 적용되는 가중치로 PSO 알고리즘의 수렴 속도와 안정성에 영향을 준다. r₁과 r₂는 균일분포 [0,1]에서 추출된 난수로서 탐색에 임의성을 부여하고 지역해로 수렴할 가능성을 더욱 줄인다(Bellout 등, 2012). Figure 2.9는 PSO 알고리즘에서 개별 입자해가 업데이트되는 과정을 나타내는 모식도이다.



Figure 2.9 Schematic of updating a particle in PSO algorithm

제 3 장 연구 방법

제 1 절 저류층 모델 개요

본 연구에서는 프록시 모델을 이용한 유정배치 최적화의 성능을 검증하기 위해 3D 저류층 벤치마크 모델인 PUNQ-S3 모델과 Egg 모델을 사용하였다. 본 절에서는 본격적인 연구방법의 설명에 들어가기에 앞서 사용한 모델들의 개요에 대해 설명하고자 한다.

(1) PUNQ-S3 model

PUNQ-S3 모델은 Elf Exploration Production 사에서 수행한 실제 운영데이터에 기반한 저류층 모델로 다양한 기법의 성능을 검증하는 용도로 사용된다(Floris 등, 2001). 해당 모델은 19x28x5의 그리드로 구성되어 있으며 각 격자의 크기는 x, y, z 방향으로 180m, 180m, 15m이고 그 중에서 1761개의 그리드만 활성화시켜 저류층의 불규칙 적인 형태를 구현한다. 저류층의 상부에는 gas cap이 존재하고 저류층의 동쪽과 남쪽에는 단층이 존재한다. 그리고 저류층의 서쪽과 북쪽에는 대수층이 존재하는데 본 연구의 경우 주입정의 효과를 확인하기 위해 대수층을 제거한 상태의 모델을 사용하였다.

Figure 3.1은 PUNQ-S3 모델의 상부면 구조를 보여주고 Figure 3.2는 각 층의 유체투과율 지도를 보여준다. PUNQ-S3 모델에서 1, 3, 5층은 공극률이 큰 sand층이 셰일층에 둘러싸여 있고 2, 4층은 공극률 이 낮은 셰일층으로 구성되어 있다. 그 외의 저류층 및 운영 조건은 Table 3.1에 정리하였다.

20



Figure 3.1 The top structure of PUNQ-S3 model



Figure 3.2 The log permeability of PUNQ-S3 model

Parameters	Values	
Reference depth, m	2355	
Initial reservoir pressure,	234.46	
Oil-Water Contact, m	2395	
Gas-Oil Contact, m	2355	
Rock compressibility, 1/bar		0.00045
Formation volume factor, rm ³ /sm ³	Oil	1.2
	Water	1.0042
	Gas	0.0044
Fluid compressibility, 1/bar	Oil	1.0725E-4
	Water	5.43E-5
	Gas	4.586E-3
Fluid viscosity, cp	Oil	1.46
	Water	0.5
	Gas	0.0133
Fluid density, kg/sm ³	Oil	912
	Water	1000
	Gas	0.8266

Table 3.1 Petrophysical parameters of PUNQ-S3 model

(2) Egg model

Egg 모델은 유체투과율의 불균질성이 매우 큰 3차원 채널 저류층 모델로 채널형태의 퇴적환경이나 불확실성이 큰 저류층을 대상으로 하는 기법의 검증에 사용된다(van Essen 등, 2009; Jansen 등, 2014). 해당 모델은 60x60x7의 그리드로 구성되어 있으며 본 연구에서는 계산의 편 의성을 위해 상부의 2개 층만 사용하였다. 각 격자의 크기는 x, y, z 방 향으로 118ft, 118ft, 20ft이고 그 중에서 5092개의 그리드만 활성화되 어 있다.

Egg 모델은 불확실성을 고려하기 위해 101개의 등가모델을 가지고 있고 등가모델들의 로그유체투과율 분포의 평균은 5이고 분산은 1.5이 다. 본 연구에서는 등가모델들 중 임의로 한 개 모델을 선정하여 사용하 였다. Egg 모델의 경우 PUNQ-S3 모델과 달리 저류층에 가스가 존재 하지 않으며 압축성이 거의 없는 오일과 물이 공극을 차지하고 있다. Figure 3.3은 사용한 Egg 모델의 유체투과율 지도를 나타내고 Table 3.2는 시뮬레이션에 사용된 기타 저류층 및 운영 조건들을 나타낸다.



Parameters	Values	
Reference depth, m		4000
Initial reservoir pressure,	400	
Initial water saturation, fr	0.1	
Initial porosity, fraction	0.2	
Oil-Water Contact, m		5000
Rock compressibility, 1/bar		3E-5
Formation volume factor, rm ³ /sm ³	Oil	1
	Water	1
Fluid compressibility, 1/bar	Oil	1E-5
	Water	1E-5
Fluid viscosity, cp	Oil	3
	Water	1
Fluid density, kg/sm ³	Oil	900
	Water	1000

Table 3.2 Petrophysical parameters of Egg model

(3) Mobility ratio를 고려한 상대유체투과율 곡선 수정

상대유체투과율은 하나의 유체가 존재하는 공극매질에 다른 종류의 유체가 유입되었을 때의 유효유체투과율과 절대유체투과율의 비로 다상 유동 시스템에서 매우 중요한 인자 중 하나이다. 저류층 반응을 정확하게 모사하기 위해서는 실제 저류층 조건에서의 상대유체투과율 곡선을 사용하여야 하나 대다수의 시뮬레이션에서는 실험실 조건에서 구한 곡선을 사용한다. 실험실에서 상대유체투과율을 측정할 때는 연속적인 곡선을 얻기 위해 17cp 정도의 매우 높은 점성도를 가지는 인공 오일이 사용된다. 따라서 실제 저류층 유체를 사용한 것이 아니기 때문에 상대유체투과율이 왜곡되어 저류층 시뮬레이션이 유체 유동을 현실적으로 모사하지 못하게 된다.

Figure 3.4는 1차원 waterflooding을 저류층 시뮬레이션으로 구현한 결과이다. Buckley-Leverett equation의 이론적 해에서는 불연속적인 shock front가 하나 존재하지만 시뮬레이션 결과를 보면 불연속면이 순차적으로 다수 나타났다. 이는 일종의 수치분산 오차가 발생한 결과이며 높은 점성의 오일로 측정한 상대유체투과율 곡선을 사용하면서 물의 mobility가 과대평가되었기 때문이다.



Figure 3.4 Numerical simulation of a 1D waterflooding (Dake, 2001)

Dake(2001)에 따르면 이러한 문제는 fractional flow 개념을 도입함으로써 해결할 수 있다. Fractional flow는 전체 유량 중 물의 유량의 비율을 나타내며 식 3.1으로 계산한다. Fractional flow는 상대유체투과율의 함수이고 상대유체투과율은 물포화도에 따라 변하므로 따라서 물포화도의 함수로 나타낼 수 있다. 곡선의 개형은 물과 오일의 mobility ratio에 따라 결정되며 mobility ratio는 식 3.2로 계산된다.

$$f_{w} = \frac{q_{w}}{q_{w} + q_{o}} = \frac{1}{1 + \frac{\mu_{w}}{\mu_{o}} \frac{k_{ro}}{k_{rw}}}$$
(3.1)

$$M = \frac{k'_{rw}}{\mu_w} / \frac{k'_{ro}}{\mu_o}$$
(3.2)

여기서 μ는 유체의 점성도이고 kr은 상대유체투과율이다. 그리고 kr'은 곡선의 끝점에서의 상대유체투과율값이다.

Waterflooding 상황에서 물은 포화도가 water breakthrough 값에 도달한 후부터 오일을 밀어내기 시작하며 이 지점에서 shock front가 형성된다. M>1일 때, fractional flow 곡선은 S자 형태를 띠며 Welge(1952)는 Figure 3.5처럼 초기 물포화도(Swc)에서 곡선에 접선을 그으면 그 접점이 water breakthrough saturation과 같다는 것을 밝혀냈다. 접점의 좌표는 fractional flow 곡선 위의 한 점에서의 순간기울기가 그 점과 초기 물포화도를 연결한 직선의 기울기와 같아지는 지점으로부터 찾아낼 수 있다.



Figure 3.5 Welge technique at breakthrough when M > 1 (Dake, 2001)

Water breakthrough saturation값을 구한 후에는 기존의 상대유체투과율 곡선에서 breakthrough saturation보다 작은 영역의 값은 0의 값을 가지도록 수정하면 물의 비현실적인 이동을 제한할 수 있고 앞선 수치분산 오차를 제거할 수 있게 된다. 본 연구에서는 PUNQ-S3 모델과 Egg 모델 각각에 대해 위 과정을 수행하였다. Mobility ratio 값은 PUNQ-S3 모델이 3.24, Egg 모델이 2.81이고 water breakthrough saturation은 각각 0.722, 0.542로 계산되었다.

물과 오일만 존재하는 Egg 모델과 달리 PUNQ-S3 모델은 gas cap이 존재하고 3상 유동에서의 상대유체투과율을 수정하는 것은 매우 복잡한 과정이 수반된다. 또한 해당 모델에서 유정의 완결과 생산은 하부층에서만 이루어지기 때문에 가스는 유동에 직접적으로 참여하는 것보다 drive로서의 역할이 주도적이라고 가정할 수 있다. 따라서 PUNQ-S3 모델의 경우에는 계산의 효율성을 고려하여 물과 오일에 대한 상대유체투과율 곡선만 수정하였다. 상대유체투과율 수정 시 사용된 그래프들과 결과는 Figure 3.6에 제시하였다.



Figure 3.6 Modification of relative permeability curves

제 2 절 유정배치 샘플링

CNN 프록시 모델은 정답 데이터로부터 입력 자료와 출력 자료 사이의 관계를 학습하는 지도학습이 사용되기 때문에 학습 데이터를 샘플링하는 과정이 선행되어야 한다. 머신러닝 기법은 학습 데이터의 질이 모델의 예측 성능에 매우 중요하고 Lee(2021)에 따르면 초기에 생산성이 높은 유정배치를 샘플링하여 프록시 모델에 학습시키는 것이 추후 프록시 모델을 이용한 유정배치 최적화의 성능 향상에 영향을 미친다.

본 연구에서는 유정배치 샘플링 방법 3가지를 시도하였고 각각 2000개의 샘플을 학습 데이터로 사용하였을 때 프록시 모델의 최적화 성능을 비교하였다. 첫 번째 방법은 Kim 등(2020)이 시도한 방법으로 난수를 생성하여 유정의 위치, 종류, 개수를 임의로 선택하는 방법이다. 두 번째와 세 번째 방법은 본 연구에서 제안하는 샘플링 방법들로 이 방법들은 두 단계에 걸쳐 유정배치를 샘플링한다. 첫 번째 단계에서는 총 샘플 수의 절반만큼 random sampling 방법으로 생성하고 첫 단계의 샘플들로부터 유정배치에 대한 품질 정보를 획득한다. 그리고 이 품질 정보들을 다음 단계 샘플링 시의 확률로서 반영하여 보다 개선된 배치가 추가적으로 생성되도록 하였다.

이 때 두 번째 샘플링 방법은 NPV의 기댓값을 품질값으로 사용하였고 세 번째 샘플링 방법은 NPV 기댓값과 더불어 각 유정의 오일생산량과 물주입량을 품질값으로 사용하였다. 또한 세 번째 방법은 다른 유정의 존재로 인해 각 위치의 품질값이 변동된다는 것을 고려하여 이전에 배치된 유정의 위치에 따라 다른 품질지도를 사용하였다. 각각의 세 방법을 순서대로 random sampling, 2-stage sampling, modified 2stage sampling으로 명명하였고 세부적인 방법은 다음과 같다.

(1) Random sampling

하나의 유정배치 샘플을 생성하기 위해서는 유정들의 개수, 종류, 위치를 결정하여야 한다. 그리고 random sampling 방법은 이들을 난수를 생성하여 임의로 설정한다. 각 유정은 x좌표, y좌표, 유정타입의
세 값을 가진다. 저류층의 평면 격자 수가 가로로 nx개, 세로로 ny개 일 때, x좌표와 y좌표는 각각 1에서 nx 중 하나의 정수값, 1에서 ny 중 하나의 정수값을 가진다. 그리고 유정타입은 -1에서 1사이의 실수값을 갖는다. 유정타입의 경우 생성된 난수가 1/3보다 크면 생산정, -1/3보다 작으면 주입정으로 설정하며 그 사이의 값을 가지면 해당 위치에는 유정을 설치하지 않고 일종의 더미 데이터로 취급한다. 이를 모든 유정에 대해 실시함으로써 유정들의 배치가 임의로 결정된다. 본 연구에서는 설치 가능한 유정의 최대 수를 12개로 설정하였고 따라서 하나의 유정배치 샘플은 36개의 데이터를 가지는 배열로 표현된다.

Figure 3.7은 간단한 예를 들기 위해 설치 가능한 유정의 수를 3개로 축소시켰을 때의 데이터 배열과 그 때의 유정배치 지도이다. 그림을 보면 3개의 유정 중 하나는 더미 데이터이기 때문에 실제로 시뮬레이션 상에 설치되는 유정은 두개임을 알 수 있다.



Figure 3.7 The example of random sampling (red: producer, black: injector)

유정배치 샘플을 생성한 후에는 비현실적인 배치를 제외시키기 위해 샘플이 제약조건을 만족하는지를 검사한다. 이 때 사용되는 제약조건은 총 3가지로 다음과 같다.

- ① 유정 주변으로 40acre의 넓이에 해당하는 원 안에는 다른 유정이 존재하지 않도록 유정간 최소 거리를 유지
- ② 유정은 활성화된 그리드에만 위치
- ③ 생산정은 저류층 가장자리 그리드에는 위치시킬 수 없음

만약 위의 제약조건을 위배하는 샘플이 존재한다면 해당 샘플을 제거하고 새로운 샘플을 추가로 생성하였다. 이 제약조건은 2-stage sampling과 modified 2-stage sampling에도 똑같이 적용되고 PSO 알고리즘에서 입자해들을 업데이트하는 과정에도 적용된다.

(2) 2-stage sampling

2-stage sampling은 품질지도를 만들기 위해 총 샘플 수의 절반인 1000개는 random sampling 방법으로 생성한다. 이 때 유정의 개수에 따른 영향을 분석하기 위해 유정 개수가 3개인 케이스부터 12개인 케이스까지 10가지의 케이스가 동일한 샘플 수를 가지도록 하였다. 따라서 샘플들은 유정이 설치되는 개수별로 100개씩 생성된다.

유정의 개수를 임의로 선정하였던 random sampling과 달리 이번엔 샘플별로 유정의 개수가 고정되어 있기 때문에 기존의 코드를 조금 수정하여 적용하였다. 샘플은 여전히 36개의 데이터 배열로 나타나고 유정의 종류는 -1에서 1 사이의 난수값에 의해 결정된다. 그러나 원래 -1/3과 1/3을 기준으로 주입정, 유정 없음, 생산정으로 구분하였다면 이번에는 0을 기준으로 양수면 생산정으로 결정하고 음수면 주입정으로 결정하였다. 그리고 고정된 개수 외의 유정들에 대해서는 0값을 부여하여 결과적으로 샘플별로 정해진 개수의 유정들만 설치되도록 하였다. 그 후 random sampling한 샘플 집합을 통계적으로 분석하여 유정의 개수, 종류, 위치 각각에 대한 NPV 기댓값을 계산하였다. 먼저 유정의 개수와 종류의 경우 random sample 집합을 개수별로 분류하고 각 부분집합의 샘플들이 가지는 NPV의 평균값을 해당 부분집합의 NPV 기댓값으로 설정하였다. 추가적으로 같은 방법을 사용하여 각 부분집합별로 생산정 수와 주입정 수의 비율에 따른 NPV 기댓값을 계산하였다. 그러나 유정 종류 간의 개수비의 경우 모든 케이스를 고려하면 다음 단계의 샘플링 과정이 너무 번거롭게 된다. 따라서 생산정과 주입정의 비율은 NPV 기댓값이 가장 높을 때의 정보를 사용하였다.

Figure 3.8은 PUNQ-S3 모델과 Egg 모델에 대해 유정 개수에 따른 NPV 기댓값과 개수별로 가장 NPV 기댓값이 높았던 유정 종류간 비율을 나타낸 그래프들이다. 그림을 보면 PUNQ-S3 모델은 유정이 12개일 때 NPV 기댓값이 가장 높고 Egg 모델은 11개일 때 가장 높았다. 그리고 PUNQ-S3 모델에서 생산정의 비율이 유독 높았던 이유는 gas drive로 인해 물 주입정의 효과가 크지 않았기 때문으로 보인다.



Figure 3.8 The analysis on well numbers and types of primary samples

그 다음으로 유정의 위치의 경우 각 위치에 대한 NPV 기댓값으로 이루어진 품질지도를 얻는다. Random sampling한 집합에서 특정 위치에 생산정 혹은 주입정이 있는 샘플들만 탐색하고 그 샘플들의 NPV의 평균값을 해당 위치의 NPV 기댓값으로 정하여 품질지도를 만든다. 따라서 품질지도는 생산정과 주입정에 대해 총 2개가 만들어지고 Figure 3.9는 그 결과를 나타낸다. da Cruz(2004)에서 제안한 품질지도는 생산정 하나를 격자마다 위치시켜보며 누적오일 생산량을 계산하여 구하지만 이 방법은 품질지도를 만들기만 하는데도 상당한 시뮬레이션 수행이 요구된다. 따라서 본 연구에서 사용한 방법은 프록시 모델의 학습데이터를 얻는 것과 동시에 품질지도를 만들 수 있다는 장점이 있다.



Figure 3.9 Quality map obtained from the expected NPV (left: producer, right: injector)

위와 같이 유정의 개수, 종류, 위치에 대한 품질 정보를 획득하였고 이를 확률로서 반영하여 다음 샘플링을 수행하였다. 각 유정의 개수와 위치가 NPV 기댓값에 비례하는 확률을 가지도록 Figure 3.10과 같은 CDF (Cumulative Distribution Function)을 만들고 0에서 1까지의 난수를 생성하여 그 값에 해당하는 CDF값을 갖는 것을 선택하였다. 그리고 유정의 종류에 대해서는 앞 단계 샘플링에서 0이 유정의 종류를 나누는 기준값이었다면 이 기준값을 비율에 맞도록 조정하여 그 확률을 반영해주었다. 그 결과 두 번째 샘플링 단계에서는 NPV 기댓값이 높은 유정 개수, 종류, 위치가 중점적으로 선택된다.



Figure 3.10 CDFs of well numbers and positions

(3) Modified 2-stage sampling

Modified 2-stage sampling 방법은 앞선 2-stage sampling을 보다 개선한 방법이라고 할 수 있다. 유정의 개수와 종류에 관해서는 2-stage sampling과 똑같은 방법을 사용하였고 올바른 비교를 위해 동일한 random sample 집합으로부터 품질 정보를 획득하였기 때문에 두 방법 사이의 차이는 없다. 두 방법의 차이는 유정 위치에 대한 품질지도의 획득 방법에 있으며 2-stage sampling 방법이 생산정과 주입정에 대해 한 개씩의 품질지도를 사용하여 모든 유정의 위치를 선택하였다면 modified 2-stage sampling 방법은 유정별로 각기 다른 품질지도를 사용한다. 이는 각 위치의 품질이 절대적이지 않으며 품질값이 주변 유정들의 배치에 따라 변동된다는 사실을 고려하기 위해서이다(Badru, 2003).

Modified 2-stage sampling의 경우 각 격자별로 품질지도를 생성한다. Random sample 집합 중에서 특정 격자에 유정이 위치해 있는 샘플들만 추려낸 후 그 샘플들로부터 자신의 품질이 아닌 다른 유정들의 품질을 결정한다. 그리고 개별적인 품질을 얻기 위해 NPV를 사용하지 않고 생산정은 누적오일생산량(WOPT, Well Oil Production Total), 주입정은 누적물주입량(WWIT, Well Water Injection Total)의 기댓값을 품질값으로 사용하였다. 따라서 각 품질지도는 특정 격자에 유정이 위치하였다고 가정했을 때의 조건부 품질 정보를 담고 있다.

위의 방법으로 모든 격자에 대해 품질지도를 만들고 그 외에 추가로 2-stage sampling과 유사한 방법으로 각 격자의 WOPT 기댓값과 WWIT 기댓값으로 구성된 품질지도를 만든다. 따라서 PUNQ-S3 모델은 533*2개의 품질지도가 만들어지고 Egg 모델은 3601*2개의 품질지도가 만들어진다. 여기서 용이한 설명을 위해 앞으로 각 격자에 대해 만든 품질지도는 조건부 품질지도, 전체 필드에 대해 만든 품질지도는 전체 품질지도로 명명하였다.

Modified 2-stage sampling에서 유정들은 순차적으로 선정된다. 먼저 유정의 개수와 종류가 결정되면 첫 번째 유정의 위치를 결정한다. 첫 번째 유정은 사전에 설치된 유정이 없기 때문에 전체 품질지도에 상응하는 CDF를 사용하여 위치를 결정한다. 두 번째 유정은 첫 번째 유정의 위치에서 얻은 조건부 품질지도를 사용하고 세 번째 유정은 두 번째 유정의 위치에서 얻은 조건부 품질지도를 사용한다. 이 과정을 정해진 개수의 유정을 모두 선정할 때까지 반복한다. Modified 2-stage sampling의 대략적인 과정을 2-stage sampling의 과정과 비교하여 Figure 3.11에 나타내었다.

34



Figure 3.11 Schematics of proposed sampling methods

제 3 절 CNN 프록시 모델 구축

CNN 프록시 모델 학습에 사용되는 데이터는 입력 자료인 유정배치 지도, TOF 지도와 출력자료인 NPV값이다. 세 가지 방법으로 샘플링한 유정배치를 가지고 유선 시뮬레이션을 통해 TOF 지도를 얻고 저류층 시뮬레이션으로 NPV를 계산한다. 저류층 시뮬레이션은 Schlumberger 사의 Eclipse 100을 사용하였고 유선 시뮬레이션은 Texas A&M에서 개발한 TRACE3D를 사용하였다.

저류층 시뮬레이션은 3600일 동안 수행하였으며 유선은 시간에 따라 변화가 크지 않으므로 유선 시뮬레이션은 초기 30일 동안만 수행한 결과를 사용하였다. 본 연구에서 사용한 Egg 모델은 비압축성의 물과 오일로만 구성된 저류층이고 PUNQ-S3 모델도 상부에 가스층이 존재하나 gas drive로서의 역할이 주도적이며 생산 초기에는 오일의 단상 흐름이 지배적이다. 따라서 유선을 갱신하지 않고 초기의 유선을 따라 TOF를 계산하였다. 그리고 시뮬레이션에 사용된 기타 인자들은 Table 3.3에 제시하였다.

Table 3.3 Simulation parameters

parameters	Values			
Production BHP, barsa	100			
Injection BHP, barsa	400			
Oil price, \$/STB	60			
Gas price, \$/Mcf	4			
Water disposal cost, \$/STB	3			
Water injection cost, \$/STB	5			
Drill cost, \$ MM/well	2			
Discount rate, %	10			

(a) PUNQ-S3

(b) Egg

parameters	Values
Production BHP, barsa	300
Injection BHP, barsa	500
Oil price, \$/STB	60
Water disposal cost, \$/STB	3
Water injection cost, \$/STB	5
Drill cost, \$ MM/well	2
Discount rate, %	10

CNN 프록시 모델의 학습 데이터 셋은 다음의 과정을 따라 구축하였다. 유정배치 지도의 경우 0이 배경값인 2차원 배열에서 생산정의 위치에는 1을 대입하고 주입정의 위치에는 -1을 대입하여 이산적인 2차원 배열 데이터로 가공하였다. 그리고 TOF 지도의 경우 PUNQ-S3 모델과 Egg 모델 모두에서 두 개의 층에서 유정의 완결이 이루어졌기 때문에 해당 두 층의 TOF 지도를 사용하였다.

본 연구에서 사용한 유선 시뮬레이션 프로그램인 TRACE3D의 경우 TOF값이 약 10만 부근까지 계산되지만 이러한 TOF 지도를 CNN의 입력자료로 사용할 시 컨볼루션 연산과 풀링 연산에서 과도하게 큰 값으로 인해 결과를 왜곡시킬 수 있기 때문에 저류층 시뮬레이션을 수행한 기간인 3600일을 TOF의 상한값으로 설정하였다.

그리고 TOF 지도는 생산정을 시작점으로 잡은 TOFP 지도와 주입정을 시작점으로 잡은 TOFI 지도의 두 종류가 있다. 결과적으로 CNN 프록시 모델에는 유정배치 지도 한 개와 TOF 지도 4개가 입력된다.

출력자료인 NPV값의 경우 Table 3.3에 정리한 인자들과 식 3.3으로 계산하였다.

$$NPV = \sum_{t=1}^{T} \frac{\left(C_o Q_o^t + C_g Q_g^t - C_{wp} Q_{wp}^t - C_{wi} Q_{wi}^t\right)}{(1+d)^t} - N \times C_d$$
(3.3)

여기서 C_o, C_g, C_{wp}, C_{wi}, C_d는 각각 oil price, gas price, water disposal cost, water injection cost, drill cost이다. Q_t는 [t-1, t] 기간 동안 생산 혹은 주입된 유량을 의미한다. T는 총 운영 기간이고 d는 연간 할인율, N은 설치된 유정 수를 의미한다.

이렇게 만들어진 학습자료 2000개는 훈련 데이터 70%, 검증 데이터 15%, 시험 데이터 15%로 나누어진다. 프록시 모델은 훈련 데이터를 가지고 모델이 예측한 값들과 실제 값들 사이의 평균 제곱근 편차(RMSE, Root Mean Squared Error)를 식 3.4로 계산하고 이 값이 최소가 되도록 모델의 파라미터들을 업데이트한다. 이 때 파라미터 최적화 알고리즘으로는 Adam을 사용하였다. 검증 데이터는 모델이 훈련 데이터에 과적합되는 것을 방지하기 위해 학습종료시점을 결정하기 위해 사용하였으며 검증 데이터에 대한 RMSE가 최소였을 때를 최종 모델로 결정하였다. 그리고 시험 데이터를 사용하여 프록시 모델의 예측 성능을 평가하였으며 평가기준은 예측값과 실제값의 일치 정도를 나타내는 결정계수(R²)이다(식 3.5).

$$RMSE = \sqrt{E\left[\left(y_i - F(x_i)\right)^2\right]}$$
(3.4)

$$R^{2} = 1 - \frac{\sum_{i} (y_{i} - F(x_{i}))^{2}}{\sum_{i} (y_{i} - \bar{y})^{2}}$$
(3.5)

여기서 x는 프록시 모델(F)에 입력되는 자료이고 y는 정답 자료, \bar{y} 는 정답 자료들의 평균이다. 구축한 CNN 프록시 모델의 대략적인 모식도는 Figure 3.12와 같고 구조를 Table 3.4에 정리하였다.



Figure 3.12 Schematic of CNN proxy model

(a) PUNQ-S3				
Layer (type)	Output shape			
Input	(28, 19, 5)			
Convolution	(28, 19, 128)			
BN	(28, 19, 128)			
ReLU	(28, 19, 128)			
Convolution	(28, 19, 128)			
BN	(28, 19, 128)			
ReLU	(28, 19, 128)			
Average pooling	(14, 9, 128)			
Convolution	(14, 9, 256)			
BN	(14, 9, 256)			
ReLU	(14, 9, 256)			
Average pooling	(7, 4, 256)			
Convolution	(7, 4, 256)			
BN	(7, 4, 256)			
ReLU	(7, 4, 256)			
Average pooling	(3, 2, 256)			
Convolution	(3, 2, 128)			
BN	(3, 2, 128)			
ReLU	(3, 2, 128)			
Average pooling	(1, 1, 128)			
Flatten	(1, 128)			
FC	(1, 128)			
BN	(1, 128)			
ReLU	(1, 128)			
Dropout (0.2)	(1, 128)			
FC	(1, 1)			
Linear	(1, 1)			

Table 3.4 Structure of CNN proxy model

Layer (type)	Output shape		
Input	(60, 60, 5)		
Convolution	(60, 60, 128)		
BN	(60, 60, 128)		
ReLU	(60, 60, 128)		
Average pooling	(30, 30, 128)		
Convolution	(30, 30, 128)		
BN	(30, 30, 128)		
ReLU	(30, 30, 128)		
Average pooling	(15, 15, 128)		
Convolution	(15, 15, 256)		
BN	(15, 15, 256)		
ReLU	(15, 15, 256)		
Average pooling	(7, 7, 256)		
Convolution	(7, 7, 256)		
BN	(7, 7, 256)		
ReLU	(7, 7, 256)		
Average pooling	(3, 3, 256)		
Convolution	(3, 3, 128)		
BN	(3, 3, 128)		
ReLU	(3, 3, 128)		
Average pooling	(1, 1, 128)		
Flatten	(1, 128)		
FC	(1, 128)		
BN	(1, 128)		
ReLU	(1, 128)		
Dropout (0.2)	(1, 128)		
FC	(1, 1)		
Linear	(1, 1)		

(b) Egg

제 4 절 프록시 모델을 이용한 유정배치 최적화

마지막으로 구축한 프록시 모델을 이용하여 유정배치 최적화를 수행하고 샘플링 방법 간의 최적화 성능을 비교하였다. PSO 알고리즘의 입자해는 유정배치 데이터이고 목적함수는 입자해로부터 프록시 모델이 예측한 NPV값이다. 입자해의 개수는 40개를 사용하였고 100세대 동안 최적화를 진행하였다. 그리고 본 연구에서는 PSO 알고리즘의 가중치 값들을 w=0.729, c₁=c₂=1.494로 설정하였다(Clerc와 Kennedy, 2002).

PSO 알고리즘은 속도개념에 기반하여 입자해들의 위치를 갱신하며 해공간을 탐색해 나간다. 하지만 이 탐색과정은 유정배치의 제약조건을 고려하지 않는다. 따라서 입자해들 중에는 제약조건을 어기는 해들이 다수 존재하게 되며 프록시 모델이 학습되지 않은 입력에 대해 예측해야 되기 때문에 최적화 성능이 매우 저하된다. 따라서 PSO 알고리즘에 제약조건을 반영하기 위한 부수적인 기법이 사용된다.

이에 Kim 등(2020)은 제약조건을 어긴 해들에 대한 정보는 교환하지 않고 제거하는 필터기법을 사용하였다. 이 방법은 군집이 전체적으로 제약조건을 만족하면서 높은 목적함수값을 갖는 영역으로 이동하도록 만들 수는 있지만 결과를 저장하지 않을 뿐 탐색은 수행되기 때문에 탐색 효율이 나쁘다고 할 수 있다.

따라서 본 연구에서는 제약조건이 유정 간 혹은 유정과 경계면 사이의 거리가 너무 가까우면 위배된다는 점에 착안하여 거리가 너무 가까우면 반발력이 작용하여 반대 방향으로 유정의 위치를 이동시킴 으로써 제약조건을 만족시키도록 하였다. 모든 입자해들이 제약조건을 만족시킨 채로 탐색을 수행하기 때문에 프록시 모델의 예측성능 및 최적화 성능이 향상되었고 그 결과 Kim 등(2020)이 200세대 동안 최적화를 진행한 반면 본 연구에서는 100세대의 최적화 과정만으로도 입자해들이 안정적으로 수렴하는 것을 확인할 수 있었다. Figure 3.11은 본 연구에서 사용한 프록시 모델을 이용한 유정배치 최적화 과정의 전반적인 순서도를 나타낸다.

41



Figure 3.13 Flow chart of PSO algorithm using proxy model

제 4 장 연구 결과

제 1 절 PUNQ-S3 모델

본 연구의 단계는 크게 sampling, training, optimization으로 나누어지고 본 장에서는 순서대로 각 단계에서의 결과를 정리하였다. 먼저 PUNQ-S3 모델에서 샘플링 방법에 따른 샘플들의 NPV 히스토그램을 그리면 Figure 4.1과 같다. 2-stage sampling과 modified 2-stage sampling 방법의 경우를 보면 두 번째로 취득한 샘플들이 첫 번째로 취득한 샘플들보다 높은 NPV 영역에 분포하고 있음을 알 수 있다.

그리고 원활한 비교를 위해 Figure 4.2처럼 세 방법의 샘플들의 NPV 분포를 박스 그림으로 나타내었다. 그 결과 modified 2-stage sampling이 가장 높은 NPV 샘플을 취득하였으며 샘플들의 NPV 분포도 가장 넓었다. 따라서 품질지도를 활용함으로써 보다 개선된 샘플이 취득되었음을 알 수 있다.



Figure 4.1 NPV histogram of samples (PUNQ-S3)



Figure 4.2 The NPV boxplot of each sampling method (PUNQ-S3)

Figure 4.3은 PUNQ-S3 모델에서 유정이 완결된 층들의 유체투과율 지도와 각 샘플링 방법들에서 취득된 샘플들의 유정 위치 분포를 나타낸다. 유정 위치 분포를 비교해보면 modified 2-stage sampling의 경우 품질지도가 저류층의 특성을 잘 반영하여 PUNQ-S3 모델에서 유체투과율이 높은 구역을 위주로 유정이 많이 설치되었고 따라서 샘플들도 높은 NPV값을 가질 수 있었음을 확인하였다.



(a) The permeability maps of completion layers of PUNQ-S3 model





2 4 6 8

2 4 6 8 10 12 14 16 18 X grid 10 X grid

12 14 16 18

Figure 4.3 The well location distribution in PUNQ-S3 model (left: producer, right: injector)

다음으로 각각의 샘플링 방법들에 대해 구축한 CNN 프록시 모델의 성능은 Figure 4.4와 같다. 세 모델 모두 훈련, 검증, 시험의 결정계수 값이 0.85 이상으로 좋은 성능을 보여주었다. 이는 유정배치와 함께 TOF 지도를 학습에 사용하는 것이 저류층의 생산성을 잘 대표할 수 있음을 시사한다.



Figure 4.4 The performance of CNN proxy models in PUNQ-S3 (left: training, mid: validation, right: test)

마지막으로 프록시 모델을 이용한 유정배치 최적화 결과는 Figure 4.5와 같다. 그림에서 빨간 점은 생산정의 위치를 나타내고 검은 점은 주입정의 위치를 나타낸다. PSO 알고리즘은 추계학적 방법이므로 각 샘플링 방법당 3회씩 반복하여 최적화를 수행하였다. 그리고 Figure 4.6은 각 세대별 최적해의 NPV값을 추적한 결과이다. 빨간색 그래프가 프록시 모델이 예측한 NPV 값이고 파란색 그래프는 저류층 시뮬레이션을 별도로 수행하여 구한 실제 NPV 값이다. 검은색 그래프는 프록시 모델을 사용하지 않고 저류층 시뮬레이션만을 사용한 최적화 결과이다.



Figure 4.5 Optimal well placements in PUNQ-S3 model



Figure 4.6 The NPVs of optimal solutions in each generation (PUNQ-S3)

위 그래프를 보면 random sampling 방법은 세 방법 중 예측값과 실제값 사이의 차이가 가장 크게 나타났고 최적해 부근에서 이러한 경향이 더욱 두드러졌다. 2-stage sampling 방법은 최적해의 예측 성능 자체는 보다 향상되었지만 PSO 결과가 불안정한 모습이 나타났다. 그에 반면 본 연구에서 최종적으로 제안한 modified 2-stage sampling 방법의 경우 최적해들의 예측성능이 최적화 전과정에서 높은 수준을 유지하였고 저류층 시뮬레이션을 사용하여 수행한 최적화 결과(검은색 실선)와도 유사한 양상을 보였다. 따라서 이 방법으로 학습 데이터를 생성한 CNN 프록시 모델이 저류층 시뮬레이션을 대체하기에 가장 적합하였다.

Table 4.1은 각각의 시도에서 CNN 프록시 모델이 찾은 최종해의 예측 오차를 정리한 결과이다. 최종해의 예측 오차의 평균은 random sampling이 11.17%, 2-stage sampling이 4.43%, modified 2-stage sampling이 2.14%가 나왔다. Modified 2-stage sampling의 경우 최종해의 오차가 근소하고 저류층 시뮬레이션을 통해 구한 최적값 (2075MM\$)과도 유사하여 최적화가 효과적으로 수행되었음을 알 수 있다.

Sampling	Trial	Predicted NPV	True NPV	Error	Avg. error
method		(MM\$)	(MM\$)	(%)	(%)
Random	1	2176	2036	6.88	11.17
	2	2047	1892	8.19	
	3	2095	1768	18.46	
2-stage	1	1836	1901	3.42	4.43
	2	1838	1753	4.86	
	3	2002	1906	5.01	
Modified	1	2016	1989	1.36	2.14
2-stage	2	2078	2118	1.89	
	3	2024	2090	3.17	

Table 4.1 The NPVs of final solutions in PUNQ-S3 model

제 2 절 Egg 모델

샘플들의 NPV 히스토그램과 박스그림을 나타낸 Figure 4.7과 Figure 4.8을 보면 PUNQ-S3에서의 경우와 마찬가지로 modified 2stage sampling이 random sampling보다 개선된 샘플이 얻어 졌음을 알 수 있다. 그러나 2-stage sampling은 샘플들의 NPV분포가 random sampling과 비슷하여 개선이 되었다고 보기 어려웠다. 이는 Egg 모델이 채널형태의 저류층이므로 채널의 위치가 생산성에 큰 영향을 미치는데 2-stage sampling의 경우 전체 배치의 NPV값을 가지고 품질지도를 만들었고 따라서 채널 위치에 따른 차이가 제대로 고려되지 못하였기 때문으로 추측된다.



Figure 4.7 NPV histogram of samples (Egg)



Figure 4.8 The NPV boxplot of each sampling method (Egg)

Figure 4.9에 Egg 모델 두 층의 유체투과율 지도와 각 샘플링 방법들에서 취득된 샘플들의 유정 위치 분포를 나타내었다. 유정 위치 분포를 비교해보면 modified 2-stage sampling의 경우 채널 주변으로 유정이 많이 분포하고 random sampling과 2-stage sampling은 유정 위치 분포도로부터 채널의 경향을 확인하기 어려웠다.



(a) The permeability maps of completion layers of Egg model









(d) Modified 2-stage sampling Figure 4.9 The well location distribution in Egg model (left: producer, right: injector)

다음으로 각각의 샘플링 방법들에 대해 구축한 CNN 프록시 모델의 성능은 Figure 4.10과 같다. Egg 모델에서도 샘플링 방법에 상관없이 세 모델 모두 훈련, 검증, 시험의 결정계수 값이 0.9 이상으로 좋은 성능을 보여주었다.



Figure 4.10 The performance of CNN proxy models in Egg (left: training, mid: validation, right: test)

Egg 모델을 대상으로 프록시 모델을 이용하여 유정배치 최적화를 수행한 결과는 Figure 4.11, Figure 4.12와 같다. Egg 모델에서도 각 샘플링 방법당 3회씩 반복하여 최적화를 수행하였다.



Figure 4.11 Optimal well placements in Egg model



Figure 4.12 The NPVs of optimal solutions in each generation (Egg)

Egg 모델에서도 PUNQ-S3 모델에서와 같이 random sampling 방법은 최적화가 진행될수록 최적해에 대한 예측 성능이 떨어지는 결과가 나왔다. 그리고 modified 2-stage sampling으로 구축한 CNN 프록시 모델이 최적화 전과정 동안 최적해의 NPV값을 가장 잘 예측하였다. 또한 저류층 시뮬레이션을 통해 얻은 최적화 곡선과도 잘 부합하여 해당 프록시 모델이 저류층 시뮬레이션을 대체하기에 적합하다고 할 수 있다.

하지만 2-stage sampling의 경우 샘플 데이터를 분석한 결과와 마찬가지로 Egg 모델의 채널성이 샘플링과 프록시 모델 학습에 잘 반영되지 못하였고 예측 성능이 떨어져 입자해들의 탐색이 불안정하게 진행된 결과가 나왔다. 따라서 Egg 모델과 같이 채널성을 가지고 있는 저류층은 각 위치의 개별적인 품질값을 강조해서 품질지도를 만들어야 개선된 샘플을 얻을 수 있고 최적화 성능도 향상될 수 있다. Table 4.2은 각각의 시도에서 CNN 프록시 모델이 찾은 최종해의 예측 오차를 정리한 결과이다. 최종해의 예측 오차의 평균은 random sampling이 6.16%, 2-stage sampling이 11.28%, modified 2-stage sampling이 1.11%가 나왔다. PUNQ-S3 모델의 경우와 달리 Egg 모델에서는 2-stage sampling의 결과가 기존의 random sampling보다 좋지 않았다. Modified 2-stage sampling의 경우 최종해의 오차가 매우 작고 저류층 시뮬레이션을 통해 구한 최적값(1001MM\$)과도 유사하여 최적화가 효과적으로 수행되었음을 알 수 있다.

Sampling	Trial	Predicted NPV	True NPV	Error	Avg. error
method		(MM\$)	(MM\$)	(%)	(%)
Random	1	949.4	873.4	8.69	6.16
	2	1006	951.2	5.77	
	3	1026	985.9	4.02	
2-stage	1	978.5	845.8	15.69	11.28
	2	1046	967.7	8.03	
	3	993.6	902.3	10.12	
Modified	1	963.8	969.6	0.61	1.11
2-stage	2	1005	983.4	2.16	
	3	980.1	974.7	0.55	

Table 4.2 The NPVs of final solutions in Egg model

제 5 장 결 론

본 연구에서는 CNN 프록시 모델의 학습 데이터를 생성함에 있어 샘플링 과정을 두 단계로 나누어 첫 번째 단계에서 얻은 샘플 데이터 들로부터 유정의 배치(개수, 종류, 위치)에 대한 품질 정보를 구하고 그 결과를 이용하여 두 번째 단계에서 개선된 유정배치 샘플들을 얻는 방법을 제안하였다. 그리고 해당 방법으로 구축한 CNN 프록시 모델을 3차원 저류층 모델인 PUNQ-S3 모델과 Egg 모델에 적용하여 유정배치 최적화를 수행하였다. 추가적으로 선행연구에서 사용된 샘플링 기법들과 비교하여 그 효과성을 검증하였다. 본 연구를 통해 도출된 결론들은 다음과 같다.

- 정적 자료인 유정배치 지도와 함께 저류층 내에서의 유체 유동을 정량적으로 나타낸 동적 자료인 TOF 지도를 CNN의 학습에 사용함으로써 NPV 예측성능이 높은 모델을 구축할 수 있었다. 본 연구에서 사용한 세 가지 샘플링 기법으로 구축한 모델들 모두에서 예측값과 실제값 사이의 결정계수 값이 0.93 이상이었다.
- 2. Random sampling은 유정의 개수, 종류, 위치를 임의로 조합하여 샘플을 생성하고 그 경우의 수가 다양하기 때문에 높은 NPV 영역의 샘플을 취득하는 것이 어려웠다. 최적화가 진행되면서 입자해들의 NPV가 높아질수록 학습 자료의 한계로 인해 CNN 프록시 모델의 예측 성능을 저하되었다. 그러나 최적화 과정에서 프록시 모델이 예측한 NPV값과 저류층 시뮬레이션으로 계산한 실제 NPV값이 서로 양의 상관관계를 보여 최적해 탐색이 어느 정도 이루어졌음은 확인할 수 있었다.
- 3. 샘플링을 두 단계로 나누어 실시하는 2-stage sampling과 modified 2-stage sampling의 경우 유정의 개수, 종류, 위치에 대한 품질을 계산하여 샘플링에 반영함으로써 최적화를 수행하지 않고도 초기 단계에서 높은 NPV 영역의 유정배치를 샘플링할 수 있었다.

- 4. 각 기법별로 최적화 성능을 비교하였고 그 기준으로 최적해의 예측오차를 사용하였다. PSO 알고리즘이 추계학적 기법임을 고려하여 최적화를 3회씩 반복수행 하였다. 최적해의 평균 예측 오차는 PUNQ-S3 모델에서 random sampling이 11.17%, 2stage sampling이 4.43%, modified 2-stage sampling이 2.14%였고 Egg 모델에서 random sampling이 6.16%, 2-stage sampling이 11.28%, modified 2-stage sampling이 1.11%였다. 두 저류층 모델 모두에서 modified 2-stage sampling 기법이 가장 최적화 성능이 좋았고 저류층 시뮬레이션을 사용하여 수행한 최적화 결과들과 잘 부합하여 해당 프록시 모델이 저류층 시뮬레이션을 대체하기에 가장 적합하였다.
- 5. 채널 저류층 모델인 Egg 모델에서 2-stage sampling 기법의 예측 성능 및 최적화 성능이 낮았다. 따라서 채널성을 가진 저류층 모델의 경우 각 위치의 개별적인 품질값을 강조해서 품질지도를 만들어야 개선된 샘플을 얻을 수 있고 최적화 성능도 향상될 수 있다.

본 연구에서 최종적으로 제안한 modified 2-stage sampling 기법은 좋은 예측 성능과 최적화 성능을 보여주어 저류층 시뮬레이션을 사용하는 많은 분야에 그 대체재로서 적용될 수 있다. 특히 저류층의 크기가 크고 복잡하거나 불확실성 평가를 위해 앙상블 모델이 계산에 사용될 경우 프록시 모델이 가지고 있는 장점인 계산의 효율성이 더욱 강조될 것으로 기대된다.

마지막으로 본 연구는 가스전을 대상으로 추가적인 연구가 가능할 것으로 기대된다. 유선 시뮬레이션의 경우 비압축성 유체에 대해서는 시간에 따른 유선의 변화가 미미하지만 가스와 같은 압축성의 유체는 유선이 변하기 때문에 유선을 주기적으로 갱신해주어야 한다. 따라서 CNN 프록시 모델의 학습에도 이를 고려해주어야 좋은 학습 성능을 얻을 수 있을 것이라 추측된다. 또한 본 연구에서는 유정을 두 개의 층에서만 완결시켰지만 저류층 모델에 따라 그 이상의 층을 완결시킨 경우에도 제안된 기법을 적용시켜볼 수 있을 것으로 기대된다.

참고 문헌

Afshari, S., Aminshahidy, B., and Pishvaie, M.R. 2011. Application of an improved harmony search algorithm in well placement optimization using streamline simulation. *Journal of Petroleum Science and Engineering*, Vol: 78, Issue: 3, 664–678.

Alenezi, F., and Mohaghegh, S. 2016. A data-driven smart proxy model for a comprehensive reservoir simulation. In: 4th Saudi International Conference on Information Technology (Big Data Analysis) (KACSTIT).

Awotunde, A.A., and Carlos, N. 2014. Well placement optimization constrained to minimum well spacing. In: SPE Latin America and Caribbean Petroleum Engineering Conference, Maracaibo, Venezuela. SPE-169272-MS.

Badru, O. 2003. Well-placement optimization using the quality map approach. M.S. Thesis, Stanford University.

Batycky, R.P., Blunt, M.J., and Thiele, M.R. 1997. A 3D fieldscale streamline-based reservoir simulator. *SPE Reservoir Engineering* 12(4), 246-254.

Bellout, M.C., Ciaurri, D.E., Durlofsky, L.J., Foss, B., and Kleppe, J. 2012. Joint optimization of oil well placement and controls. *Computational Geosciences*, 16(4): 1061–1079.

Bittencourt, A.C., and Roland, N.H. 1997. Reservoir development and design optimization. In: SPE Annual Technical Conference and Exhibition, San Antonio, Texas. SPE-38895-MS.

Bouzarkouna, Z., Ding, D.Y., and Auger, A. 2012. Well placement optimization with the covariance matrix adaptation evolution strategy and meta-models. *Comput Geosci* 16, 75–92.

Carroll, J.A., and Horne, R.N. 1992. Multivariate optimization of production system. *J. Pet. Technol.*, 782–831.

Chen, B., and Reynolds, A.C. 2018. CO2 water-alternating-gas injection for enhanced oil recovery: Optimal well controls and halfcycle lengths. *Computers & Chemical Engineering*, Vol: 113, 44–56. Cheng, G., An, Y., Wang, Z., and Zhu, K. 2012. Oil well placement optimization using niche particle swarm optimization. In: 8th International Conference on Computational Intelligence and Security, GuangZhou, China. 61-64.

Ciaurri, D.E., Isebor, O.J., and Durlofsky, L.J. 2010. Application of derivative-free methodologies to generally constrained oil production optimization problems. *Procedia Computer Science*, Vol: 1, Issue: 1, 1301-1310.

Clerc, M., and Kennedy, J. 2002. The particle swarm explosion, stability and convergence in a multidimensional complex space. *IEEE Trans. on Evolutionary Computation* 6, 58–73.

da Cruz, P.S., Horne, R.N., and Deutsch, C.V. 2004. The quality map: a tool for reservoir uncertainty quantification and decision making. *SPE Res Eval & Eng* 7: 6-14.

Dake, L.P. 2001. The practice of reservoir engineering (revised edition).

Datta-Gupta, A. 2000. Streamline simulation: A technology update. *Journal of Petroleum Technology* 52(12), 68-73.

Denney, D. 2010. Pros and cons of applying a proxy model as a substitute for full reservoir simulations. *J Petrol Technol* 62(07):41–42.

Dutta-Roy, K., and Kattapuram, J. 1997. A new approach to gas lift allocation optimization. In: SPE Western Regional Meeting, Long Beach, California. SPE-38333-MS.

Floris, F.J.T., Bush, M.D., Cuypers, M., Roggero, F., Syversveen, A.R. 2001. Methods for quantifying the uncertainty of production forecasts: a comparative study. *Pet. Geosci.* 7(S), S87-S96.

Forouzanfar, F., Li, G., and Reynolds, A.C. 2010. A two-stage well placement optimization method based on adjoint gradient. In: SPE Annual Technical Conference and Exhibition, Florence, Italy. SPE-135304-MS. Goodwin, N. 2015. Bridging the gap between deterministic and probabilistic uncertainty quantification using advanced proxy based methods. In: SPE Reservoir Simulation Symposium, Houston, Texas. SPE-173301-MS.

Harb, A., Kassem, H., and Ghorayeb, K. 2020. Black hole particle swarm optimization for well placement optimization. *Comput Geosci* 24, 1979–2000.

Isebor, O.J., Ciaurri, D.E., and Durlofsky, L.J. 2014a. Generalized field-development optimization with derivative-free procedures. *SPE J.*19: 891-908.

Isebor, O.J., Durlofsky, L.J., and Ciaurri, D.E. 2014b. A derivative-free methodology with local and global search for the constrained joint optimization of well locations and controls. *Computational Geosciences*, 18(3-4): 463-482.

Jaber, A.K., Al-Jawad, S.N., and Alhuraishawy, A.K. 2019. A review of proxy modeling applications in numerical reservoir simulation. *Arab J Geosci* 12, 701.

Jansen, J.D., Fonseca, R.M., Kahrobaei, S.S., Siraj, M.M., van Essen, G.M., and Van Den Hof, P.M.J. 2014. The egg model: a geological ensemble for reservoir simulation. *Geoscience Data Journal* 1:192-195.

Jansen, J.D., and Durlofsky, L.J. 2017. Use of reduced-order models in well control optimization. *Optimization and Engineering*, 18(1): 105-132.

Kennedy, J., and Eberhart, R. 1995. Particle swarm optimization. In: International Conference on Neural Networks, Perth, Australia. vol.4, 1942-1948.

Kim, J., Kang, B., Jung, H., and Choe, J. 2018. Closed-loop field development optimization with well type conversion. In: 80th EAGE Conference & Exhibition, Copenhagen, Denmark. Vol: 2018, 1-5.

Kim, J., Kang, B., Jeong, H., and Choe, J. 2019. Field development optimization using a cooperative micro-particle swarm optimization with parameter integration schemes. *Journal of Petroleum Science and Engineering*, 183: 106416. Kim, J., Yang, H., and Choe, J. 2020. Robust optimization of the locations and types of multiple wells using CNN based proxy models. *Journal of Petroleum Science and Engineering*, 193.

Kingma, D.P., and Ba, L.J. 2015. Adam: A method for stochastic optimization. In: 3th International Conference on Learning Representations, San Diego.

Kosmidis, V.D., Perkins, J.D., and Pistikopoulos, E.N. 2004. Optimization of well oil rate allocations in petroleum fields. *Industrial & Engineering Chemistry Research* 43(14), 3513–3527.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. 1998. Gradientbased learning applied to document recognition. *Institute of Electrical and Electronics Engineers* 86(11): 2278–2324.

Lee, S.H. 2021. Well placement optimization using a CNN-based proxy models. M.S. Thesis, Seoul National University.

Litvak, M.L., Gane, B.R., Williams, G., Mansfield, M., Angert, P.F., Macdonald, C.J., McMurray, L.S., Skinner, R.C., and Gregory, J.W. 2007. Field development optimization technology. In: SPE Reservoir Simulation Symposium, Houston, Texas. SPE-106426-MS.

Nwachukwu, A., Jeong, H., Pyrcz, M., and Lake, L.W. 2018. Fast evaluation of well placements in heterogeneous reservoir models using machine learning. *Journal of Petroleum Science and Engineering*, 163: 463–475.

Onwunalu, J.E., and Durlofsky, L.J. 2010. Application of a particle swarm optimization algorithm for determining optimum well location and type. *Computational Geosciences*, 14(1): 183–198.

Pollock, D.W. 1988. Semi-analytical computation of path lines for finite difference models. *Groundwater* 26(5): 743-750.

Sarma, P., and Wen, H.C. 2008. Efficient well placement optimization with gradient-based algorithms and adjoint models. In: Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands. SPE-112257-MS.

Taware, S., Park, H.Y., Datta-Gupta, A., Bhattacharya, S., Tomar, A.K., Kumar, M., and Rao, H.S. 2012. Well placement optimization in a mature carbonate waterflood using streamline-based quality maps. In: SPE Oil and Gas India Conference and Exhibition, Mumbai, India. SPE-155055-MS.

Tukur, A.D., Nzerem, P., Nsan, N., Okafor, I.S., Gimba, A., Ogolo, O., Oluwaseun, A., and Okafor, A. 2019. Well placement optimization using simulated annealing and genetic algorithm. In: SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria. SPE-198858-MS.

Udy, J., Hansen, B., Maddux, S., Petersen, D., Heilner, S., Stevens, K., Lignell, D., and Hedengren, J.D. 2017. Review of field development optimization of waterflooding, EOR, and well placement focusing on history matching and optimization algorithms. *Processes* 5(3):34.

van Essen, G.M., Zandvliet, M.J., Van Den Hof, P.M.J., Bosgra, O.H., and Jansen, J.D. 2009. Robust waterflooding optimization of multiple geological scenarios. *SPE Journal*, 14(1): 202-210.

Welge, H.J. 1952. A simplified method for computing oil recovery by gas or water drive. *J. Pet. Technol.* 4(04): 91–98.

Yang, H., Kim, J., and Choe, J. 2017. Field development optimization in mature oil reservoirs using a hybrid algorithm. *Journal of Petroleum Science and Engineering*, 156: 41–50.

Yeten, B., Durlofsky, L.J., and Aziz, K. 2003. Optimization of nonconventional well type, location, and trajectory. *SPE J.*8: 200–210.

Zhang, Y., Bansal, N., Fujita, Y., Datta-Gupta, A., King, M.J., and Sankaran, S. 2016. From streamlines to fast marching: Rapid simulation and performance assessment of shale-gas reservoirs by use of diffusive time of flight as a spatial coordinate. *SPE Journal*, 21(5): 1883-1898.
Abstract

Initial sampling methods to improve CNN-based well placement optimization

Changgyun Son Department of Energy Systems Engineering The Graduate School Seoul National University

Oilfield development planning is essential for the hydrocarbon recovery factor and the economic feasibility of projects. The number, type, location, and operating condition of each well are the main design variables for the development. These variables are highly related to reservoir parameters, which are heterogeneously distributed and have nonlinear correlations. Thus, it is difficult to grasp complex fluid behavior only by geological models and intuition. Reservoir simulation and mathematical optimization algorithms are generally utilized to perform optimization. Since they require a lot of computation, diverse proxy models have been developed to replace them.

Deep learning-based proxy models are heavily influenced by the quality of training data. And there is a problem that predictive performance is degraded as well placement optimization proceeds due to the limitation of training data. To handle this problem, retraining the proxy model in the middle of the optimization process was used. But this method has a limitation in that the timing and criteria of retraining are unclear.

In this study, initial sampling methods of training data are presented to ensure the stable predictive performance of the proxy model. The methods perform sampling in two stages. Potential productivity (quality) information on the well placement (number, type, location) is obtained from the samples generated in the first stage. After that, the additional samples are generated by reflecting this quality information as a probability in the second stage. NPV(Net Present Value), WOPT(Well Oil Production Total), and WWIT(Well Water Injection Total) are used as quality information here.

To verify the effectiveness of the proposed methods, CNN proxy models are constructed using several sampling methods and PSO(Particle Swarm Optimization) algorithm is performed three times with each proxy model to compare the results. Well placement optimizations are performed in PUNQ-S3 and Egg models, which are 3D reservoir models. The random sampling method that arbitrarily determines well placement showed average predictive errors of 11.17% in the PUNQ-S3 model and 6.16% in the Egg model, while the proposed method showed improved predictive performance of 2.14% and 1.11% respectively.

Keywords : Well placement optimization, Proxy model, CNN, 2-stage sampling, Quality map Student Number : 2020-22103

66