BIGBASE Big Data Management and Application Laboratory

# *Various Training Regimes and Strategies for Continual Deep Learning*

*Min-Seon Kim*
*Dept. of Industrial Engineering*
*Seoul National University of Science and Technology*

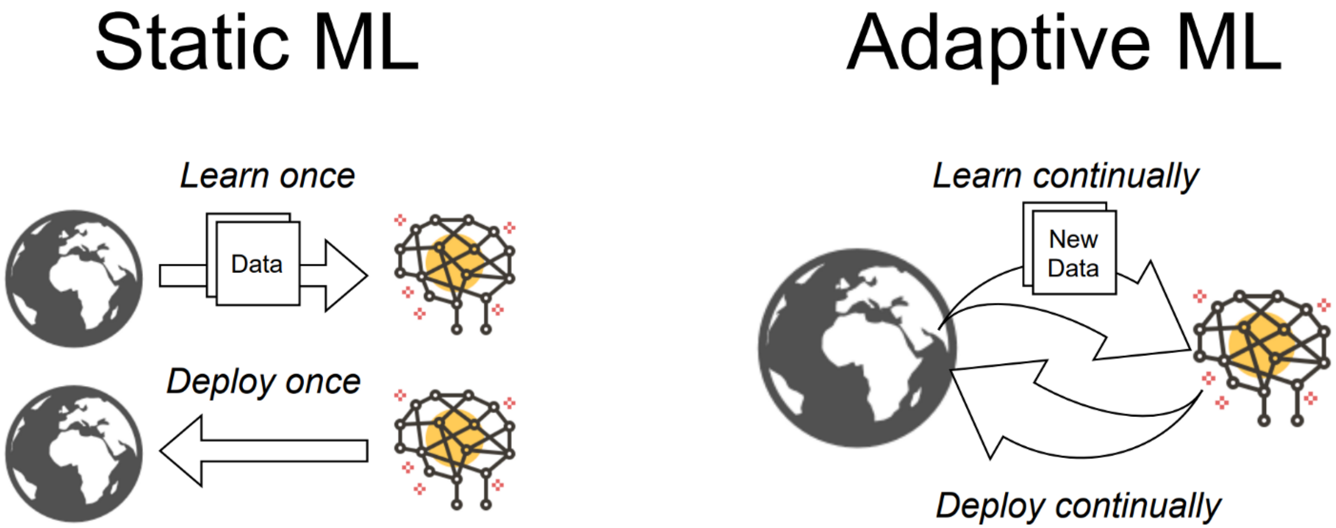
# Background

- Continual Learning 이란?

  : 시간에 따라서 학습 데이터가 주어지는 상황에서 Neural Network가 계속적으로 학습을 수행해야되는 상황

- Problems to be addressed

  1) Catastropic Forgetting 2) Limited Computing Resource (memory)



# Motivation

## 기존의 Time varying 데이터를 처리하기 위한 methodologies

1) Just In Time Learning
2) Moving Window
3) ***Continual Learning***

## 기존의 Catastrophic forgetting을 방지하기 위한 strategies

1) Deep Learning model modification
2) Appropriate training regimes (Dropout, Batch size, Learing rate)
3) Partitioning Reservoir Sampling (Replay-based approach)
4) Fairness of the data distribtuion per slice
5) Two-side loss function (Standard classification loss & distillation loss)

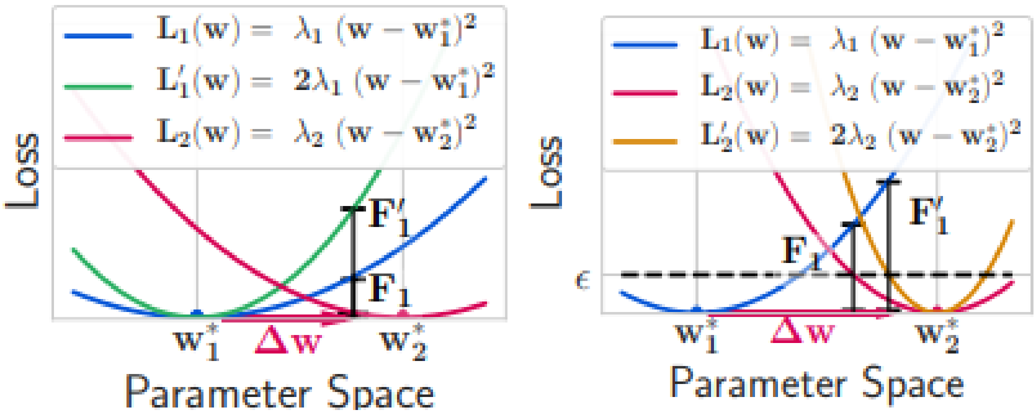
SEOULTECH

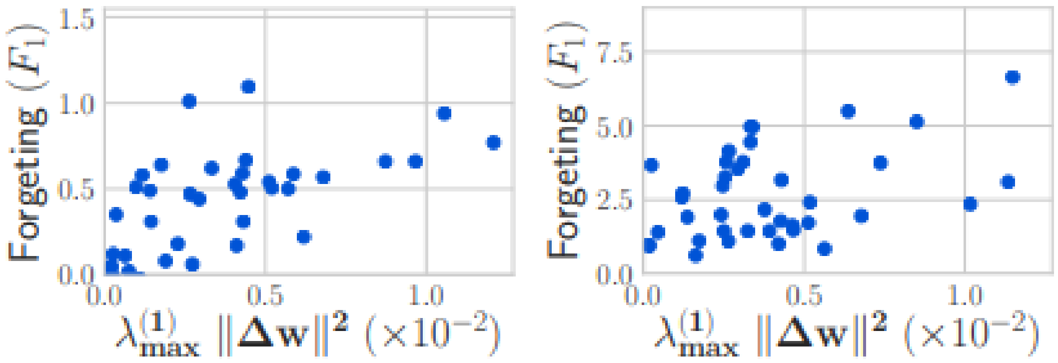
# Related Work

## Understanding the Role of Training Regimes in Continual Learning

- Dropout, Batch size, Learning  rate를 적절히 활용하는것 만으로도 catastrophic forgetting을 개선할 수 있음
- Wide local minima in loss function
  1) High Learning rate with small batch size
  2) Drop out strategy

$$F_1 = L_1(w_2^*) - L_1(w_1^*) \approx \frac{1}{2}\Delta w^\top \nabla^2 L_1(w_1^*)\Delta w \leq \frac{1}{2}\lambda_1^{max}\left\|\Delta w\right\|^2$$

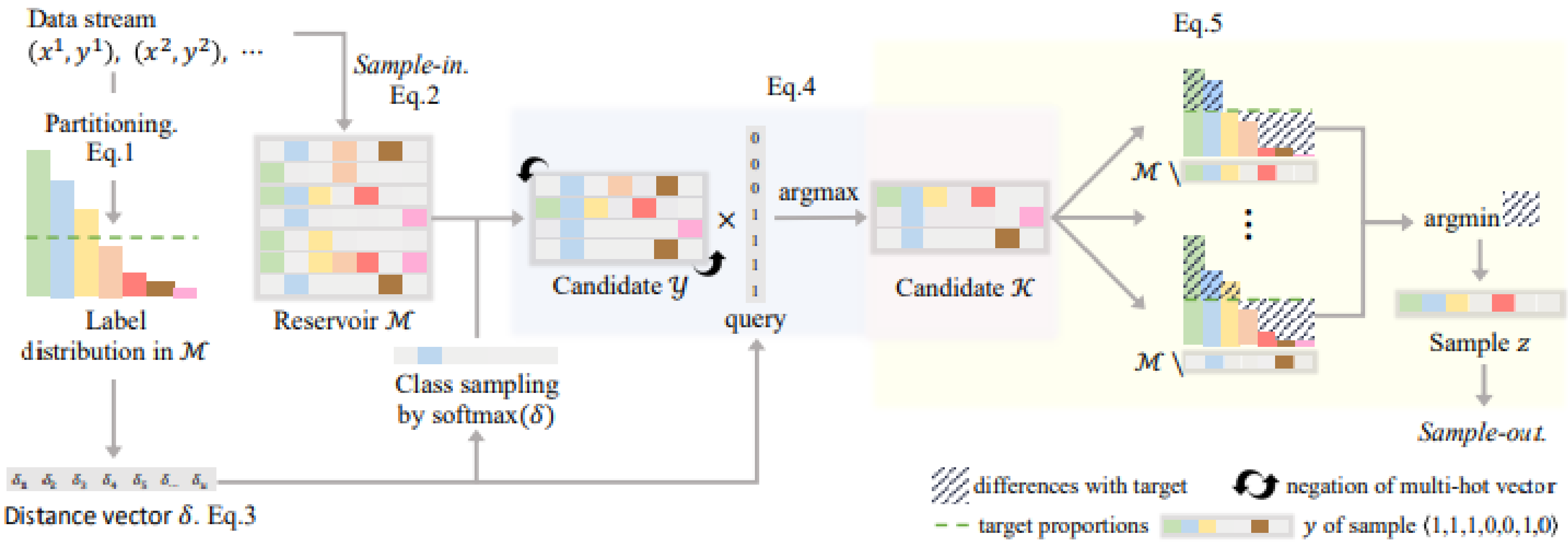


# Related Work

## Imbalanced Continual Learning with Partitioning Reservoir Sampling

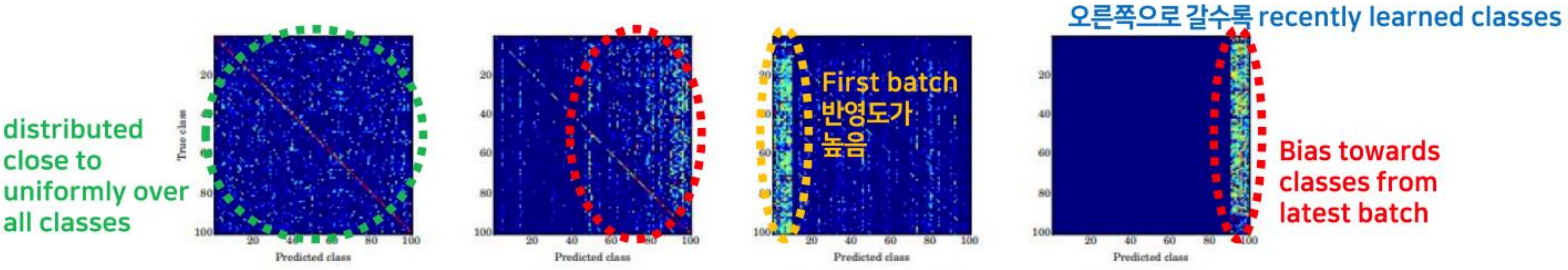
- 클래스마다 과거 데이터를를 균형있게 저장하는 Partitioning Reservoir Sampling(PRS)
  1. 데이터가 들어올때 마다 목표할당량을 바꾸어 준다
  2. 새로운 데이터를 Sample Set에 넣을지 말지를 확률적으로 결정



# Related Work

## iCaRL: Incremental Classifier and Representation Learning

- Classifier와 동시에 Feature Representation의 학습
  1) Finetuning: 최근 학습된 batch의 반영도가 높음
  2) Fixed Representation: First batch의 반영도가 높음
  3) LwF.MC(Distillation loss 고려): 비교적 최근 클래스 반영
  4) iCaRL (**Distillation loss + Prioritized exemplar set + Mean-of-exemplars classification rule**)




(a) iCaRL (b) LwF.MC (c) fixed representation (d) finetuning

# Related Work

## Enhancing incremental deep learning for FCCU end-point quality prediction

- Online learning에서 이벤트의 변화가 일어난 경우, model 구성의 modification을 통한 adaptive strategy 채택
- 과거 hidden layer에 뉴런을 추가하여 새로운 feature에 대한 학습 후 보다 compact한 representation을 얻고자 하는 시도가 있었다.

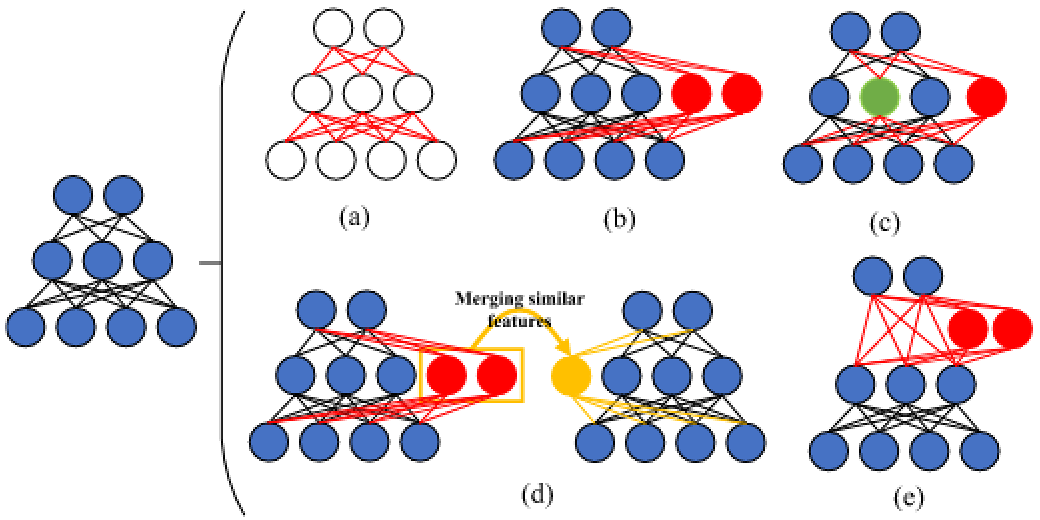
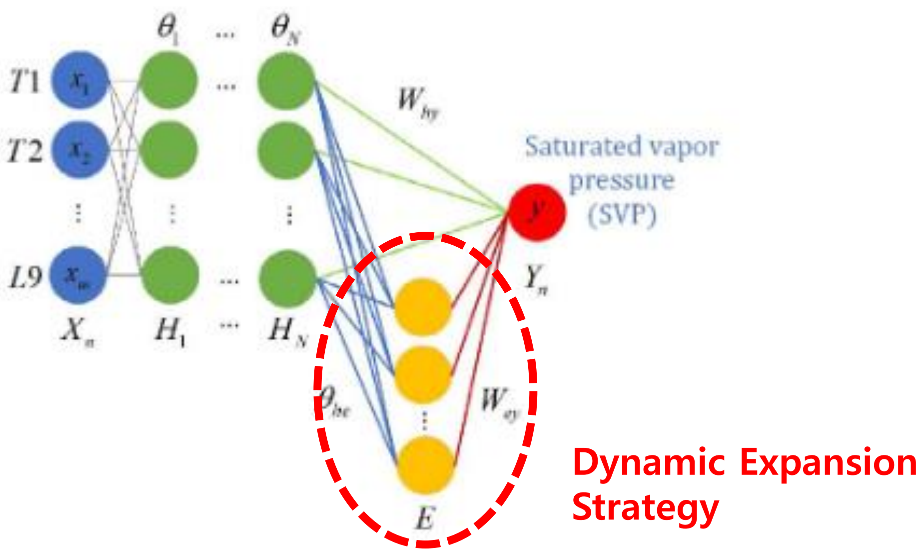


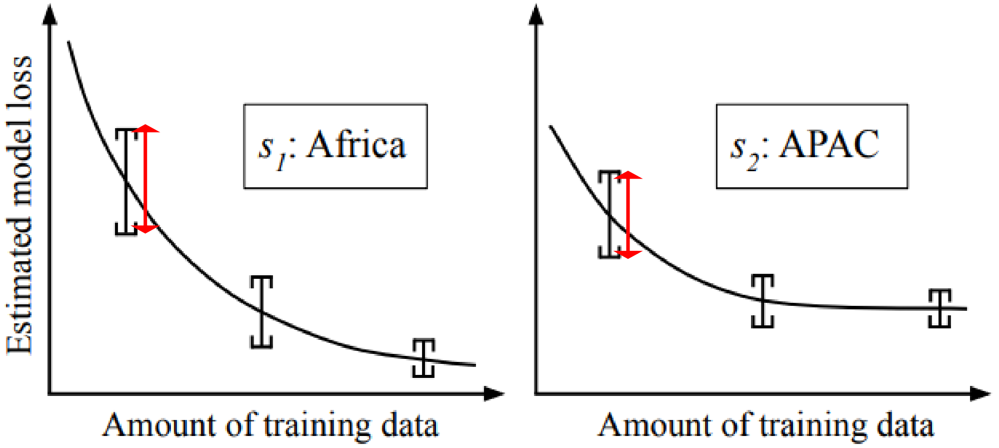

Fig. 2. Main ideas of incremental learning methods for neural network.

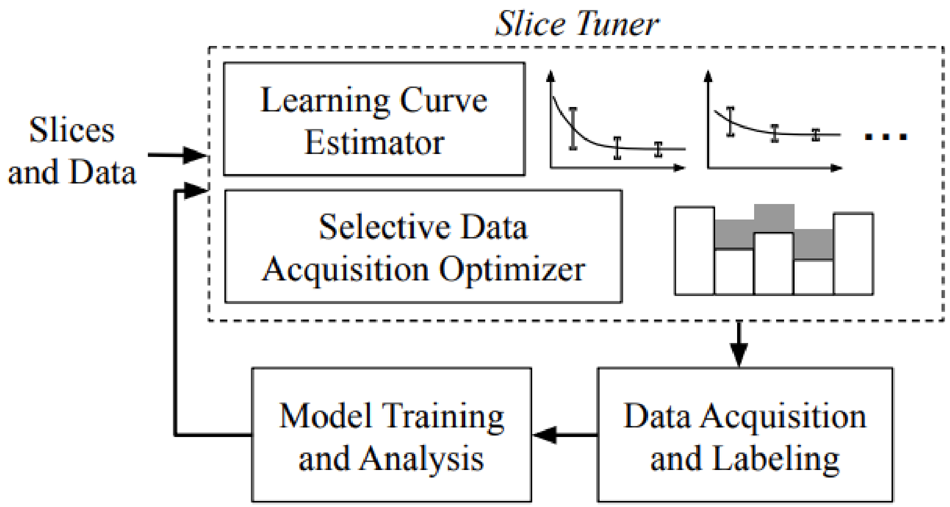


# Related Work

## Slice Tuner: A Selective Data Acquisition Framework for Accurate and Fair Machine Learning Models

- Learning Curve Estimator 와 Selective Data Acquisition Optimizer 활용하여 각 slice마다 얼마나 데이터를 수집해야 균형 있는 학습이 되는지를 결정




Africa(S1)에 대한 정보를 더 많이 얻는 것이 S2에 대한 정보를 얻는것 보다 benefit이 크다 (그래프가 더 가파르기 때문)

# Related Work

## An Incremental Deep Convolutional Computation Model for Feature Learning on Industrial Big Data

- Parameter-incremental Learning
  1) 초기 parameter조정을 위해서 새로운 loss function 도입(post prediction + deformation)
  2) 새로운 Drop-out strategy 도입하여 idle한 뉴런에 대한 새로운 feature 학습

- Structure-incremental Learning
  1) Tensor convolutional, pooling and FCN layers가 과거 정보를 transfer하도록 설계
  2) 네트워크 topology의 업데이트 이후, 새로운 커널들은 정규분포를 따르는 0에 근사한 수로 초기화 된다.

# Related Work

**In-Situ AI: Towards Autonomous and Incremental Deep Learning for IoT Systems**

: Deep learning based IoT system with autonomous IoT data diagnosis (minimize data movement), and incremental and unsupervised training method (tackle the big raw IoT data generated in ever-changing in-situ environments).

**An Incremental Iterative Acceleration Architecture in Distributed Heterogeneous E nvironments With GPUs for Deep Learning**

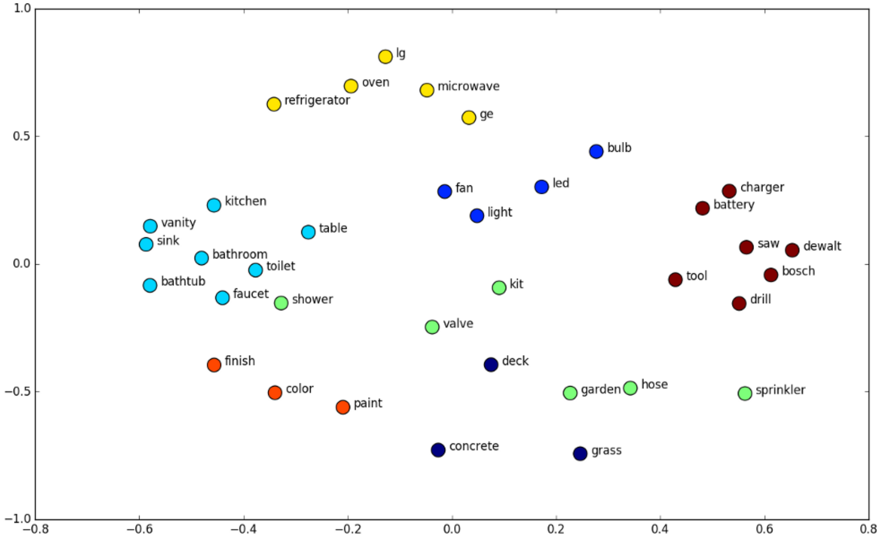
: Redundant iterative calculations을 위한 GPU-based distributed incremental iterative computing architecture.

**실제 Framework구축시 incremental learning의 장점을 이용하기 위한 여러가지 연구도 진행되었다.**

# Idea Adaptation & Future Contribution

## Online Learning에서의 word embedding의 변화 반영

- Keyword in/out logic의 구현을 통한 top N개의 prioritized keyword set 유지
- Continual word2vec model의 fine-tuning과정을 통한 keyword set 변화 반영




```
print(base_model.most_similar("hacker"))
```

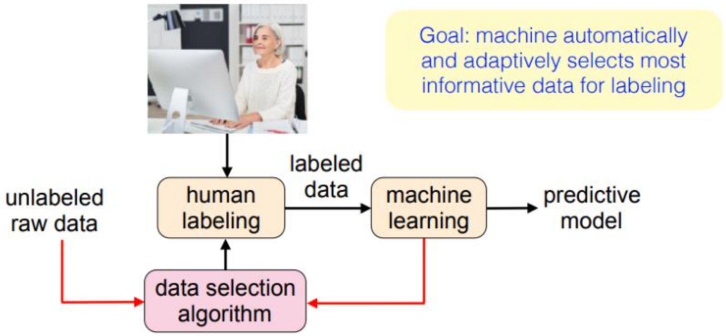
```
[('marvel', 0.37642768025398254), ('spreading', 0.35681360960006714), ('nicolasbrulez', 0.324
90628957748413), ('ni', 0.32277053594589233), ('pquotes', 0.3173757791519165), ('soho', 0.314
8268461227417), ('footage', 0.31234291195869446), ('deployments', 0.3036871552467346), ('pres
ident', 0.30267661809921265), ('runasand', 0.3025776743888855)]
```

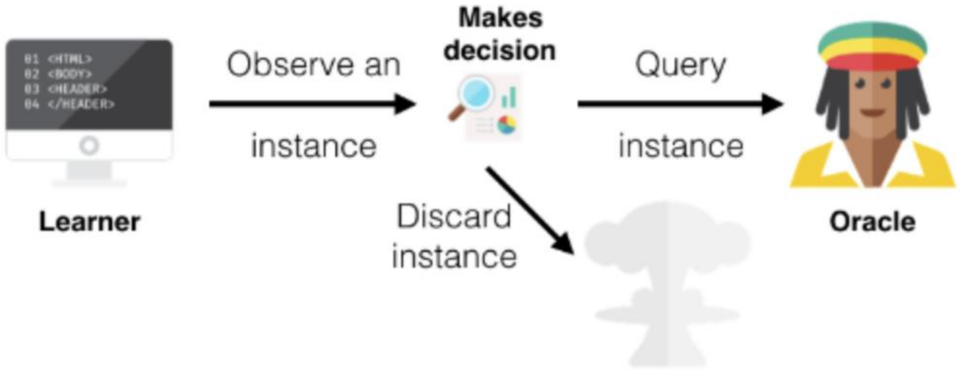
```
print(base_model.most_similar("hacker"))
```

```
[('cyberwarbooks', 0.6760294437408447), ('reviews', 0.6101169586181641), ('guide', 0.58915281
29577637), ('cwa', 0.5561687350273132), ('book', 0.5283864140510559), ('pioneers', 0.52824318
40896606), ('symanteccomconnectblogs', 0.5169950127601624), ('basics', 0.515626847743988),
('techniques', 0.5116923451423645), ('tobemcomcyberwarcyber', 0.5109221339225769)]
```

# Idea Adaptation & Future Contribution

- Valuable(event representative)한 트윗의 재 학습
  1) Background dataset의 활용
  2) Active learning의 개념 차용 (unlabeled data는 쉽게 얻을 수 있다는 특징에 기반)




- Dynamic model structure modification
  1) Feature extractor의 layer의 확장
  2) Classification layer의 output의 재활용