Predict Hacker News Upvotes

Feature Fusion
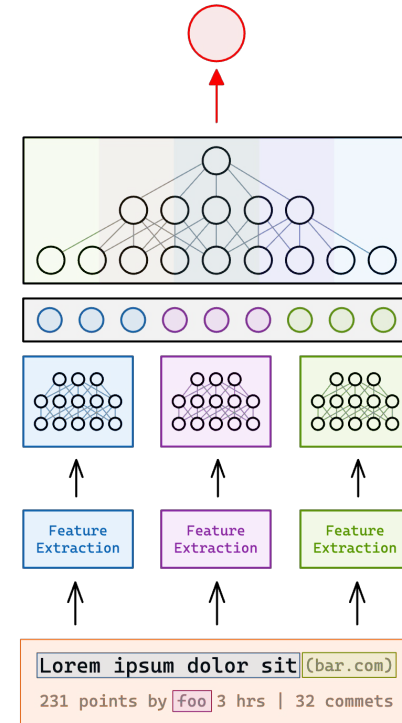
Predict Hacker News Upvotes

Predict Hacker News Upvotes

Ubiquitous supervised learning task:

- Content Recommendation Systems
- Email & Message Prioritization
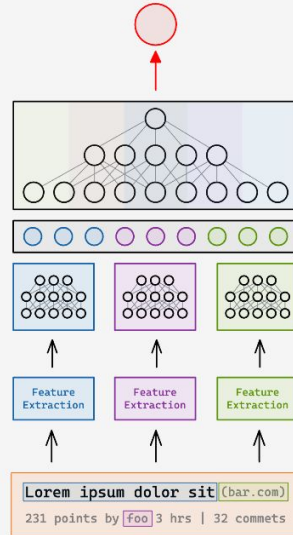- Homework Assignment Score
- Spam Detection Systems
- …

But also a Reward Model where on top we can train a Policy Model in a RLHF setup

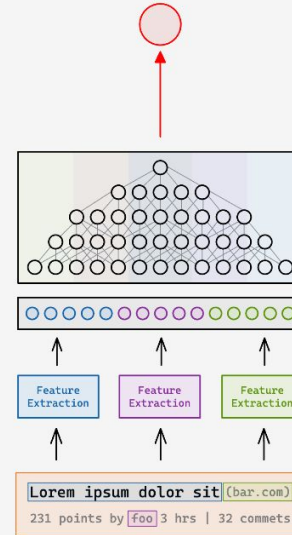# Predict Hacker News Upvotes

Late Fusion

- Pro: Specific modality processing
- Pro: Independent specialised training
- Con: Requires separate data
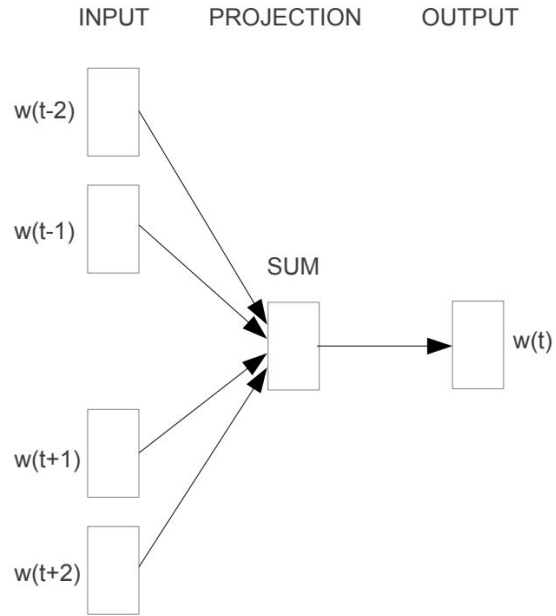- Con: May miss modalities interactions

Early Fusion

- Pro: Simple single train dataset
- Pro: Captures modality interactions
- Pro: Easy with SOTA models
- Con: Requires very complex models

Feature Extraction

Lorem ipsum dolor sit (bar.com)
231 points by foo 3 hrs | 32 commets

Predict Hacker News Upvotes

INPUT   PROJECTION   OUTPUT

w(t-2)

w(t-1)

SUM

w(t)

w(t+1)

w(t+2)

CBOW

INPUT   PROJECTION   OUTPUT

w(t-2)

w(t-1)

w(t)

w(t+1)

w(t+2)
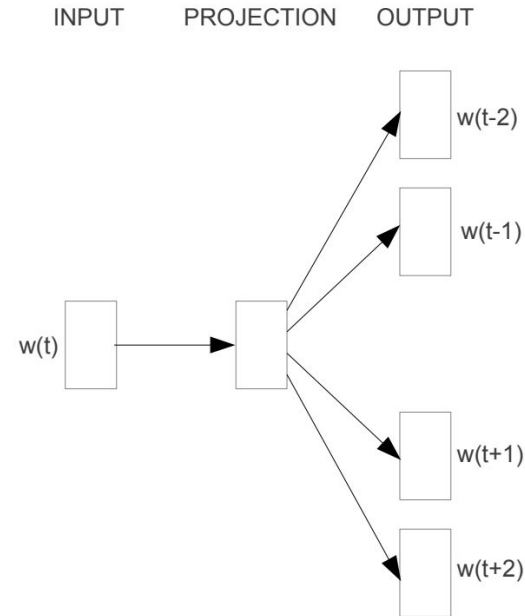
Skip-gram

Predict Hacker News Upvotes

Order of execution:

1. Exploratory Data Analysis (EDA)
   a. look for cues
   b. understand the distributions
   c. data quality issues

2. Build prediction models
   a. pre-train word2vec
   b. finetune embeddings
   c. regression task

3. Deploy a predictor service
   a. grab a server on Computa
   b. run Docker container
   c. respond to live requests

Predict Hacker News Upvotes

Tools you can use:

- PyTorch, Python, FastAPI
- Bash, Docker
- Word2Vec

You can find the data:

- postgres://sy91dhb:g5t49ao@178.156.142.230:5432/hd64m1ki
  - SELECT * FROM "hacker_news"."items" LIMIT 1;
  - "hacker_news"."items" → 28GB of data, ~42 million rows
  - "hacker_news"."users" → ~15MB, eventually ~1 million

Predict Hacker News Upvotes

Word2Vec and Hacker
News Upvote
Prediction

4 Lessons | 40 hours
with **Ardavan Afshar**

https://cortex.mlx.institute/

Good luck!

Recurrent Rebels
Anton Dergunov

Nikolas Kuhn

Tyrone Nicholas

Rasched Haidari

Perceptron Party
Jingyan Chen

Melanie Wong

Joao Esteves

James Carter

Kernel Kittens
Peter O'Keeffe

Ben Williams

Miguel Parracho

Oleh Khomiak

Gradient Gigglers
Helen Zhou

Prima Gouse

Andrei Zhirnov

Hikaru Tsujimura

Backprop Bunch
Kadriye Turkcan

Yali Pan

David Edey

Ben Liong

Bayesian Buccaneers
Tomas Krajcoviech

Ben Bethell

Umut Sagir

Arjuna James

Overfitting Overlords
Esperanza Shi

Maria Sharif

Adam Beedell

Andrew

Dropout Disco
Ethan Edwards

Dan Goss

Charles Cai

Ewan Beattie

Feature Fiestas
Tao Zamorano

Jacob Jenner

Ayman Abbas

Halil Serkan Uz

Hyperparameter Hippies
James Yan

Aparna Pillai

Clement Ha

Marcin Tolysz