国际人工智能会议
AAAI 2021 论文北京预讲会

# FairRec: Fairness-aware News Recommendation with Decomposed Adversarial Learning

Chuhan Wu[1], Fangzhao Wu[2], Xiting Wang[2], Yongfeng Huang[1] and Xing Xie[2]

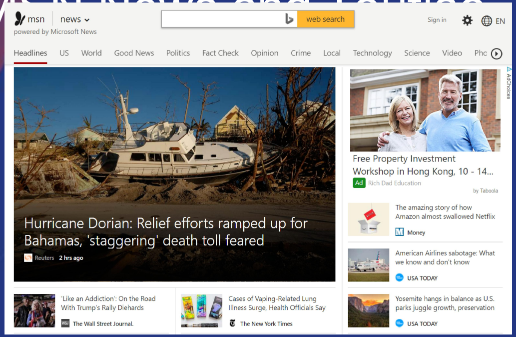
[1]Tsinghua University, Beijing 100084, China

[2]Microsoft Research Asia, Beijing 100080, China

*wuchuhan15@gmail.com

# News Recommendation

- Online news websites provide convenient access to news information
  - E.g., Google News, MSN News and Toutiao

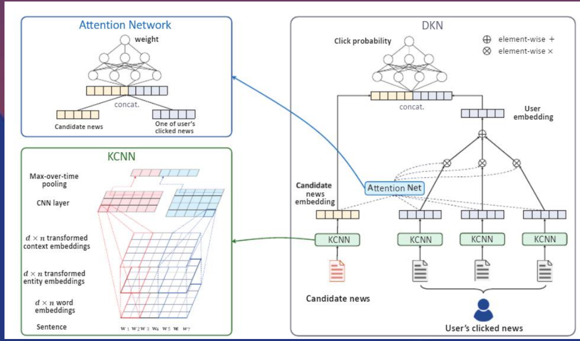





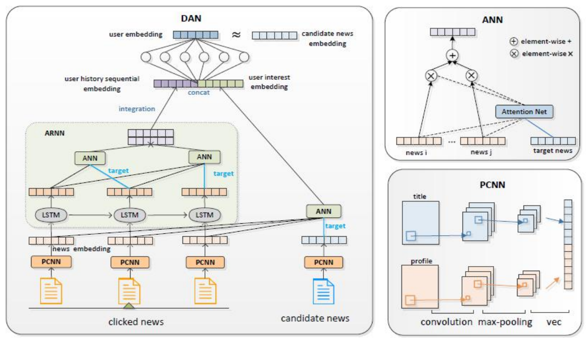
- Thousands of news generated everyday will overwhelm users
- Personalized news recommender systems are very important
  - Alleviate information overload
  - Improve reading experience

# News Recommendation

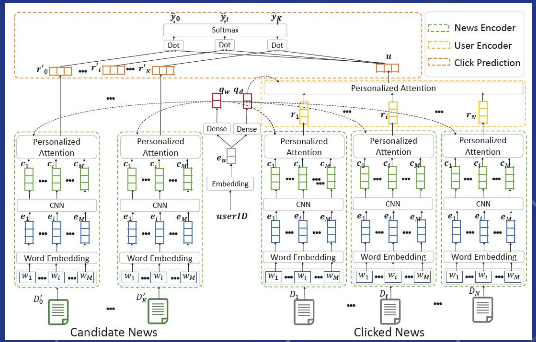
- Existing methods make personalized news recommendation based on users’ news browsing behaviors
  - E.g., DKN[1], DAN[2] and NPA[3]




DKN




DAN

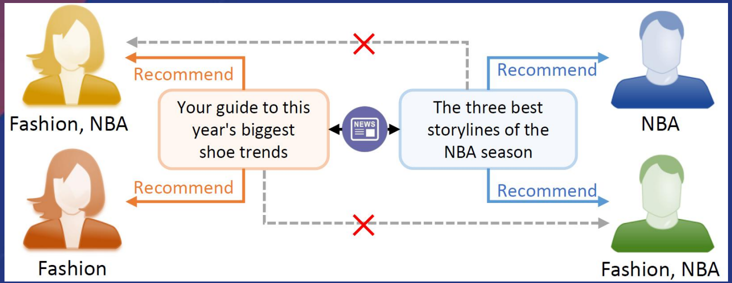



NPA

[1] Wang et al. Dkn: Deep knowledge-aware network for news recommendation. WWW 2018: 1835-1844.
[2] Zhu et al. Dan: Deep attention neural network for news recommendation. AAAI. 2019, 33: 5973-5980.
[3] Wu et al. Npa: Neural news recommendation with personalized attention. KDD 2019: 2576-2584.

# Unfairness in News Recommendation

- Users with the same sensitive attributes may have similarities in news browsing behaviors
  - E.g., many males may prefer sports news, many females may prefer fashion news




- The model can easily inherit the biases related to sensitive user attributes
  - The recommendation results are heavily influenced by sensitive user attributes
- Unfairness: the users interested in both NBA and fashion news

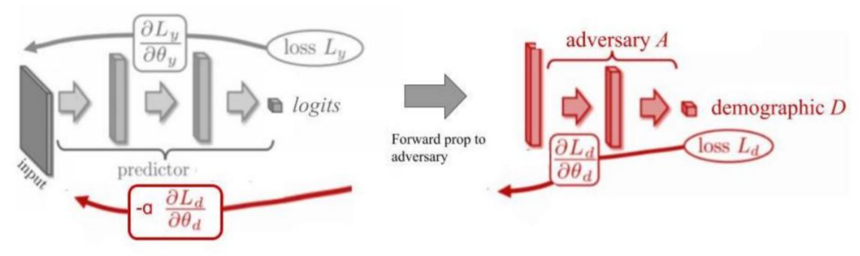
# Fairness-aware Recommendation

- The problems studied in fairness-aware recommendation methods:
  - Item fairness
    - E.g., items from different providers have a fair chance of being recommended
  - User fairness
    - E.g., provide same rankings to both protected and unprotected user groups
- Methods to achieve user fairness in recommendation
  - Rules[1], removing subspace[2], model regularization[3]
- Most of these methods focus on e-commerce scenarios
  - Rely on the predicted ratings to derive fairness metrics
  - Focus on recommendation accuracy rather than recommendation results

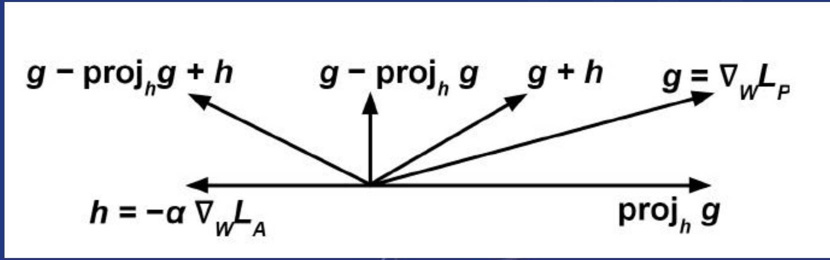
[1] Farnadi et al. A fairness-aware hybrid recommender system. FATREC 2018.
[2] Zhu et al. Fairness-aware tensor-based recommendation. CIKM 2018.
[3] Yao et al. Beyond parity: Fairness objectives for collaborative filtering. NIPS 2017.

# Fairness-aware Deep Learning

- Adversarial learning has been used for fairness-aware machine learning to remove sensitive attributes from representations
  - E.g., [1] and [2]




Adversarial learning




Adversarial learning with gradients projection

- The discriminator decision space may have shifts with the attribute space
  - The sensitive attributes may not be fully removed


[1] Wadsworth et al. Achieving fairness through adversarial learning: an application to recidivism prediction. FAT/ML 2018.

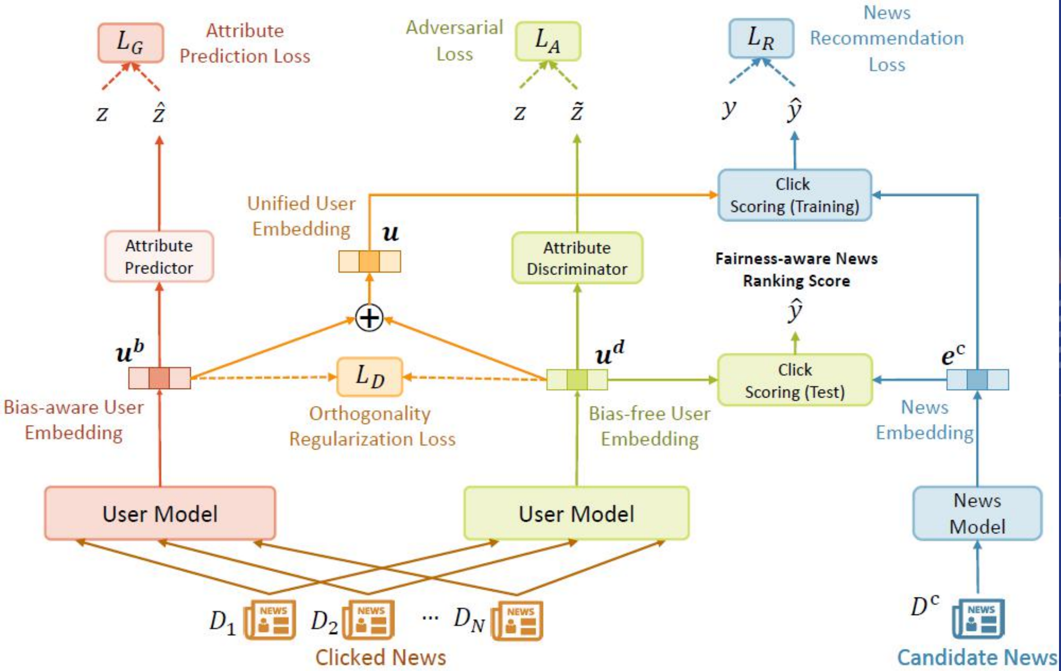
## Problem Definition

- News Recommendation
  - A target user u with a sensitive attribute z
  - A set of clicked news articles
  - A set of candidate news
  - Predict the click scores of candidate news for ranking

- Unfairness
  - If the sensitive user attribute z can be predicted from the top K ranking result more accurately, the recommendation result is more unfair

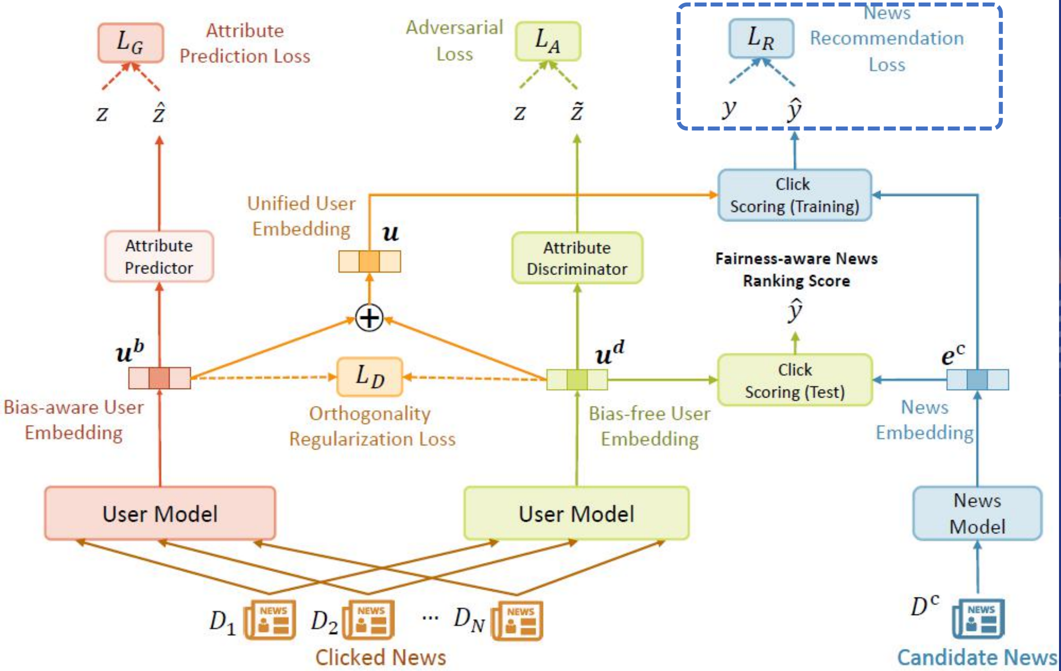
# Fairness-aware news recommendation (FairRec)

- Decomposed adversarial learning



# Fairness-aware news recommendation (FairRec)

- News Recommendation Loss

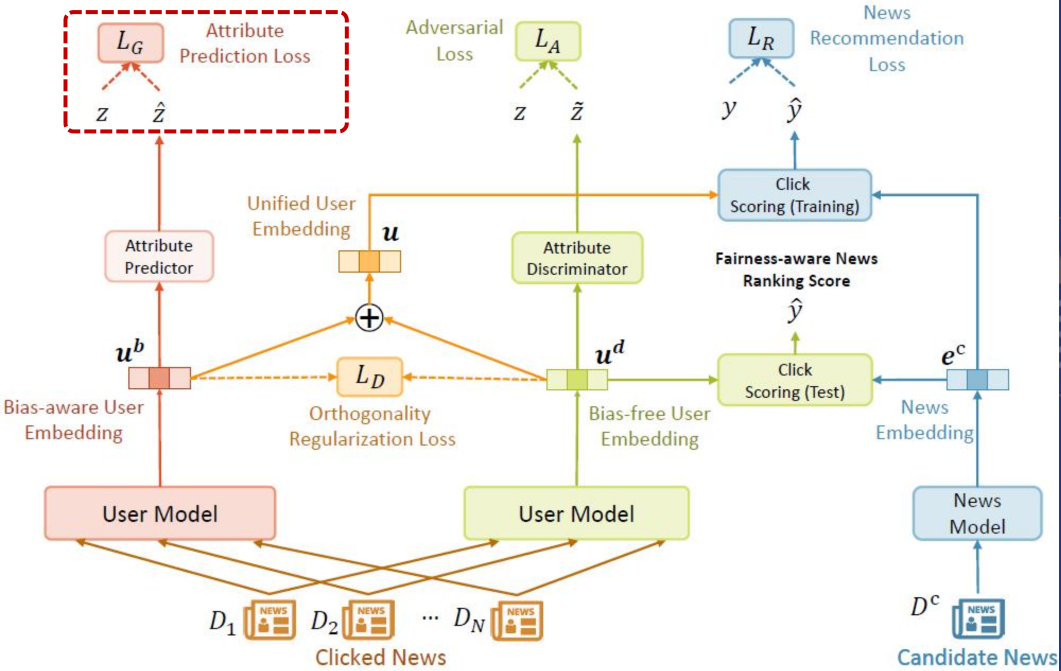



$$L_R = -\sum_{y \in Y} \sum_i y_i \log(\hat{y}_i)$$

# Fairness-aware news recommendation (FairRec)

- Sensitive Attribute Prediction Loss

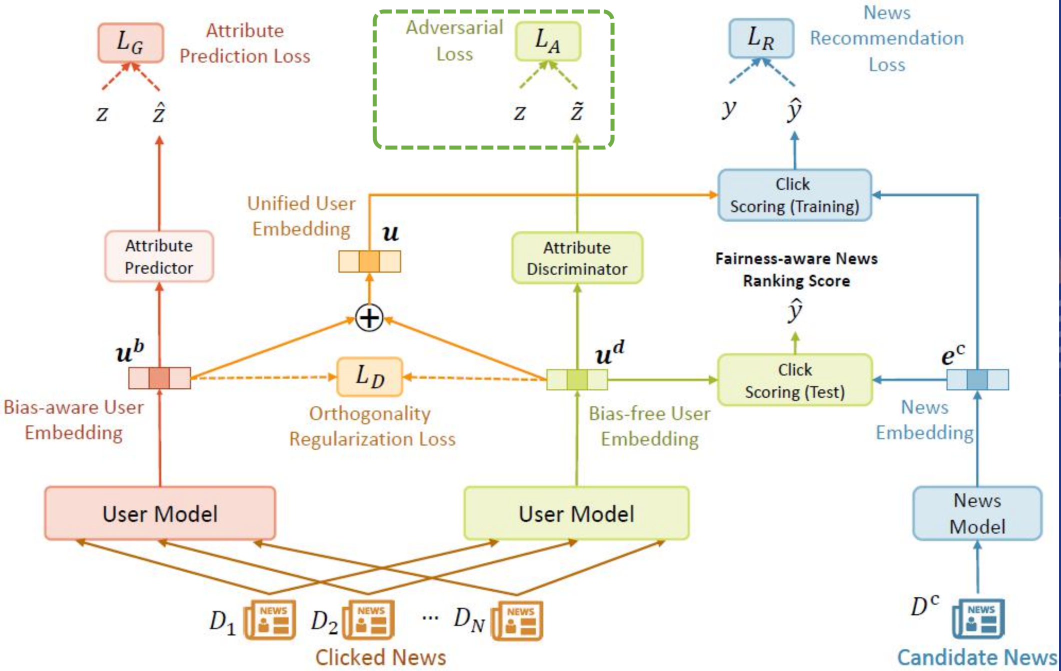
$$L_G = -\sum_{z \in G} [z \log(\hat{z}) + (1 - z) \log(1 - \hat{z})]$$



# Fairness-aware news recommendation (FairRec)

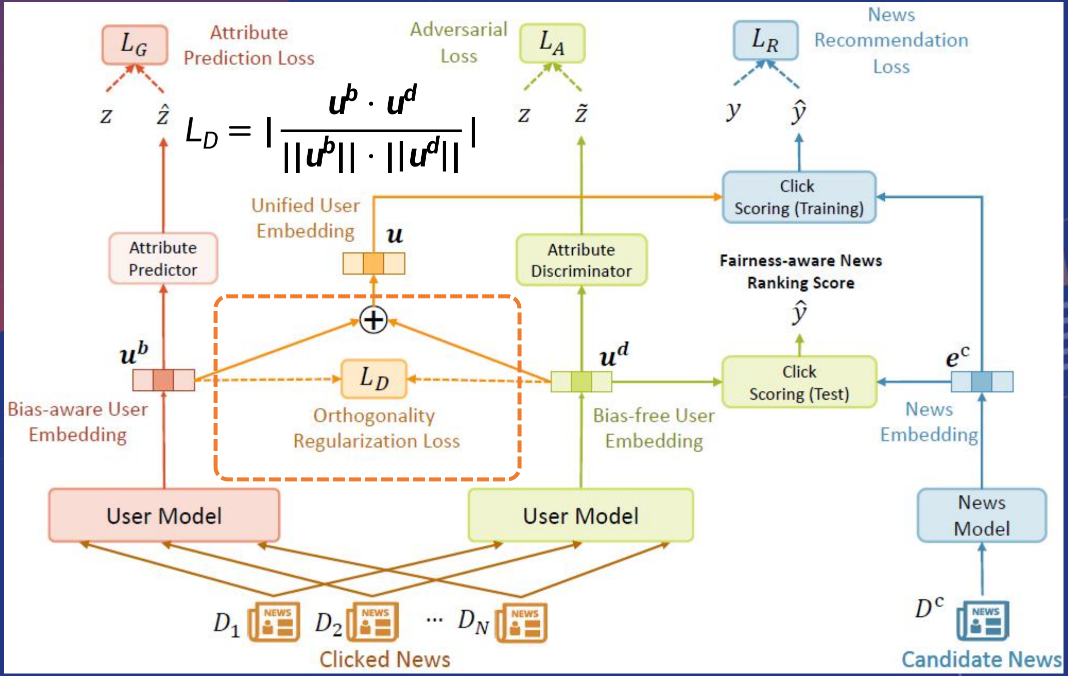
- Adversarial Loss

$$L_A = -\sum_{z \in G} [z \log(\tilde{z}) + (1 - z) \log(1 - \tilde{z})]$$



# Fairness-aware news recommendation (FairRec)

- Orthogonality Regularization Loss



# Experiments

- Dataset
  - 10,000 users and their news browsing behaviors (from 12/13/2019 to 01/12/2020)
  - Gender as the sensitive attribute
  - 4,228 users provide their gender label (M: 2,484, F:1,744)

| | | | |
|---|---|---|---|
| #users | 10,000 | avg. #words per news title | 11.29 |
| #news | 42,255 | #clicked news logs | 503,698 |
| #impressions | 360,428 | #non-clicked news logs | 9,970,795 |

- Settings
  - Loss coefficients: $L_G = L_A = L_D = 0.5$
- Metrics:
  - News recommendation: AUC, MRR, nDCG@5, nDCG@10 scores
  - Fairness: using the attribute prediction performance (accuracy and Macro-F

# Experiments

- Fairness performance evaluation

| Methods | Top 1 | | Top 3 | | Top 5 | | Top 10 | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Macro-F | Accuracy | Macro-F | Accuracy | Macro-F | Accuracy | Macro-F |
| LibFM | 62.96±0.95 | 53.73±0.89 | 65.13±0.81 | 60.07±0.80 | 66.99±0.76 | 61.69±0.78 | 68.37±0.69 | 65.41±0.66 |
| EBNR | 63.64±0.83 | 54.21±0.82 | 65.51±0.76 | 60.46±0.77 | 67.49±0.75 | 62.06±0.74 | 68.73±0.69 | 65.75±0.68 |
| DKN | 63.66±0.78 | 54.30±0.80 | 65.58±0.79 | 60.52±0.80 | 67.53±0.73 | 62.17±0.73 | 68.99±0.71 | 65.80±0.72 |
| DAN | 63.71±0.81 | 54.26±0.79 | 65.59±0.75 | 60.54±0.74 | 67.51±0.74 | 62.19±0.75 | 69.01±0.70 | 65.83±0.72 |
| NPA | 63.88±0.82 | 54.34±0.84 | 65.72±0.77 | 60.75±0.75 | 67.59±0.71 | 62.32±0.73 | 69.14±0.65 | 65.89±0.62 |
| NRMS | 63.89±0.86 | 54.40±0.83 | 65.78±0.75 | 60.79±0.76 | 67.64±0.72 | 62.35±0.70 | 69.19±0.63 | 66.01±0.68 |
| MR | 62.96±0.91 | 53.48±0.83 | 64.57±0.82 | 58.83±0.81 | 66.19±0.73 | 60.82±0.70 | 68.36±0.65 | 65.12±0.67 |
| AL | 62.55±0.85 | 52.80±0.83 | 63.31±0.74 | 57.62±0.75 | 65.43±0.68 | 59.88±0.66 | 66.86±0.62 | 63.55±0.61 |
| ALGP | 62.48±0.86 | 52.72±0.82 | 63.09±0.75 | 57.31±0.73 | 65.21±0.66 | 59.43±0.67 | 66.16±0.61 | 63.28±0.63 |
| FAN | **62.10**±0.80 | **52.41**±0.76 | **62.61**±0.69 | **54.36**±0.68 | **62.95**±0.62 | **55.98**±0.63 | **63.39**±0.59 | **57.13**±0.58 |
| Random | 62.08±0.91 | 52.39±0.90 | 62.57±0.79 | 54.27±0.79 | 62.86±0.78 | 55.91±0.76 | 63.12±0.68 | 56.97±0.67 |

Performance of different methods in terms of fairness. Lower scores indicate better fairness.

# Experiments

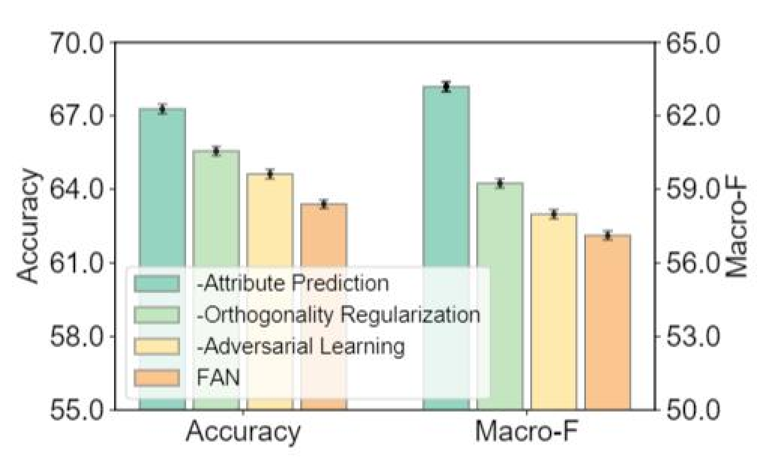
- Recommendation performance evaluation

| Methods | AUC | MRR | nDCG@5 | nDCG@10 |
|---|---|---|---|---|
| LibFM | 56.83±0.51 | 24.20±0.53 | 26.95±0.49 | 35.64±0.52 |
| EBNR | 60.94±0.24 | 28.22±0.25 | 30.31±0.23 | 39.60±0.24 |
| DKN | 60.34±0.33 | 27.51±0.29 | 29.75±0.31 | 38.79±0.30 |
| DAN | 61.43±0.31 | 28.62±0.30 | 30.66±0.32 | 39.81±0.33 |
| NPA | 62.33±0.25 | 29.46±0.23 | 31.57±0.22 | 40.71±0.23 |
| NRMS | 62.89±0.22 | 29.93±0.20 | 32.19±0.18 | 41.28±0.18 |
| FAN | 61.95±0.22 | 29.01±0.21 | 31.25±0.18 | 40.24±0.21 |

News recommendation performance of different methods. Higher scores indicate better results.

## Experiments

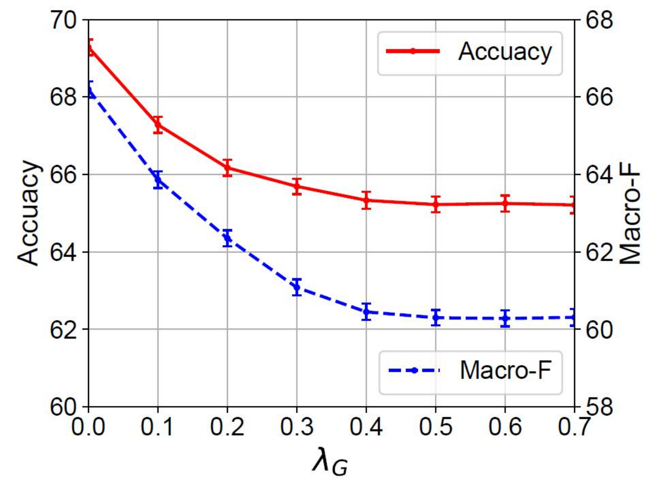
- Effectiveness of decomposed adversarial learning



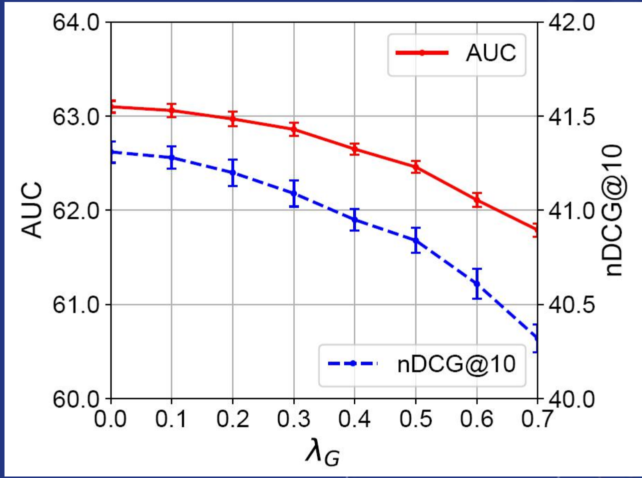

Recommendation fairness. Lower scores are better.

# Experiments

- Hyperparameter Analysis
  - Select $\lambda_G$ under $\lambda_D = \lambda_A = 0$



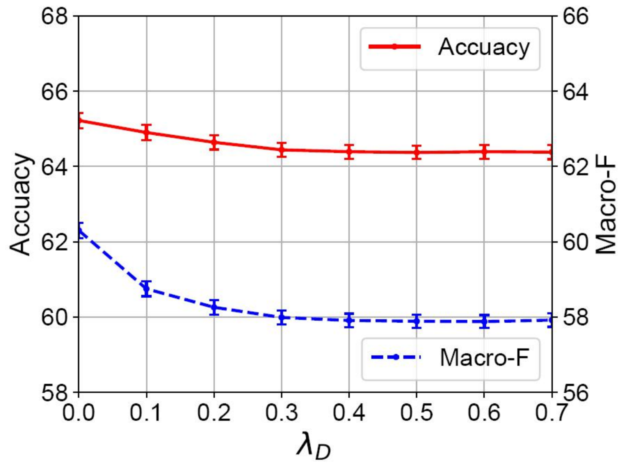

(a) Performance in terms of fairness.



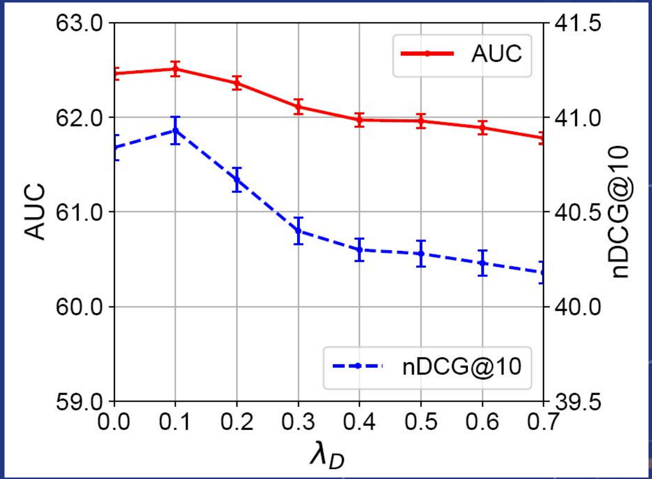

(b) Performance of news recommendation.

# Experiments

- Hyperparameter Analysis
  - Select $\lambda_D$ under $\lambda_G = 0.5,\ \lambda_A = 0$



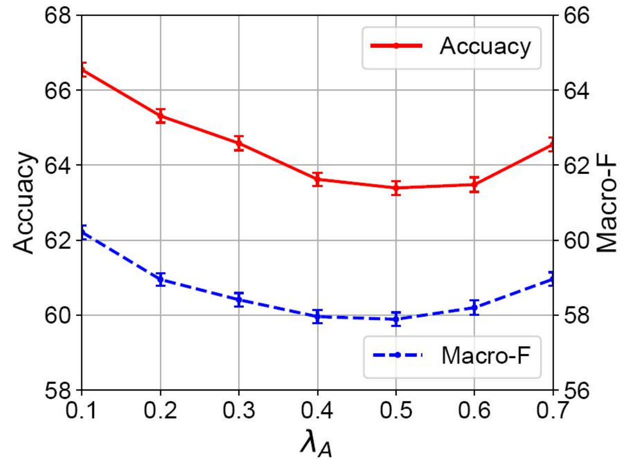

(a) Performance in terms of fairness.



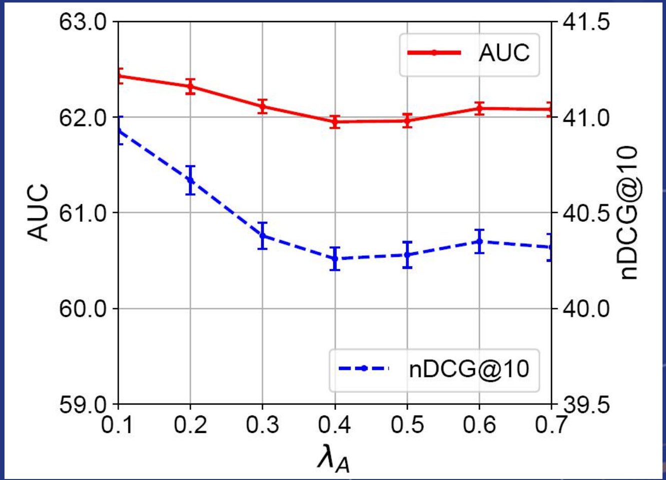

(b) Performance of news recommendation.

# Experiments

- Hyperparameter Analysis
  - Select $\lambda_A$ under $\lambda_G = 0.5,\ \lambda_D = 0.5$




(a) Performance in terms of fairness.




(b) Performance of news recommendation.

# Experiments

- Case study


Male User

| Clicked News |
|---|
| NFL playoff picture: Saints close to Clinching; Patriots fall behind Texans |
| Tom Brady had a classy reason for running right up to the ref after Sunday's win |
| 2019 Golden Globes Best Actress |

| Candidate News | Score (NRMS) | Score (FairRec) |
|---|---|---|
| Cowboys WR Allen Hurns gets encouraging news after injury | **0.92** | **0.90** |
| The Biggest Fashion Trends of 2019 Are Here — Can You Handle It? | 0.24 | **0.84** |
| 8 things making the rich even richer | 0.36 | 0.23 |
| Chefs reveal the 20 items they never make from scratch | 0.30 | 0.19 |
| Best Mexican Restaurant in Every State | 0.22 | 0.17 |


Female User

| Clicked News |
|---|
| Chris Duncan, former St. Louis Cardinals outfielder, battling brain cancer |
| Oscars fumble host test in wake of Kevin Hart's exit |
| These 5 countries have produced the most Miss Universe winners |

| Candidate News | Score (NRMS) | Score (FairRec) |
|---|---|---|
| 2019 Golden Globes Best Actress | **0.87** | **0.90** |
| Report: Mike Mccarthy only pursuing Jets coaching vacancy | 0.24 | **0.81** |
| 9 Ravens who could be potential salary cap casualties this offseason | 0.20 | **0.75** |
| 10 Myths About Frozen Foods You Need to Stop Believing | 0.30 | 0.22 |
| Here's Why Saunas Are So Good For You | 0.22 | 0.11 |

Comparison between the recommendation results of NRMS and FairRec for a male and a female user. The clicked candidate news are in blue.

国际人工智能会议
AAAI 2021 论文北京预讲会

# THANKS

2020.12.19