

Ranking

czyli kto pierwszy, ten lepszy

Lidia Wojciechowska

allegro

Plan

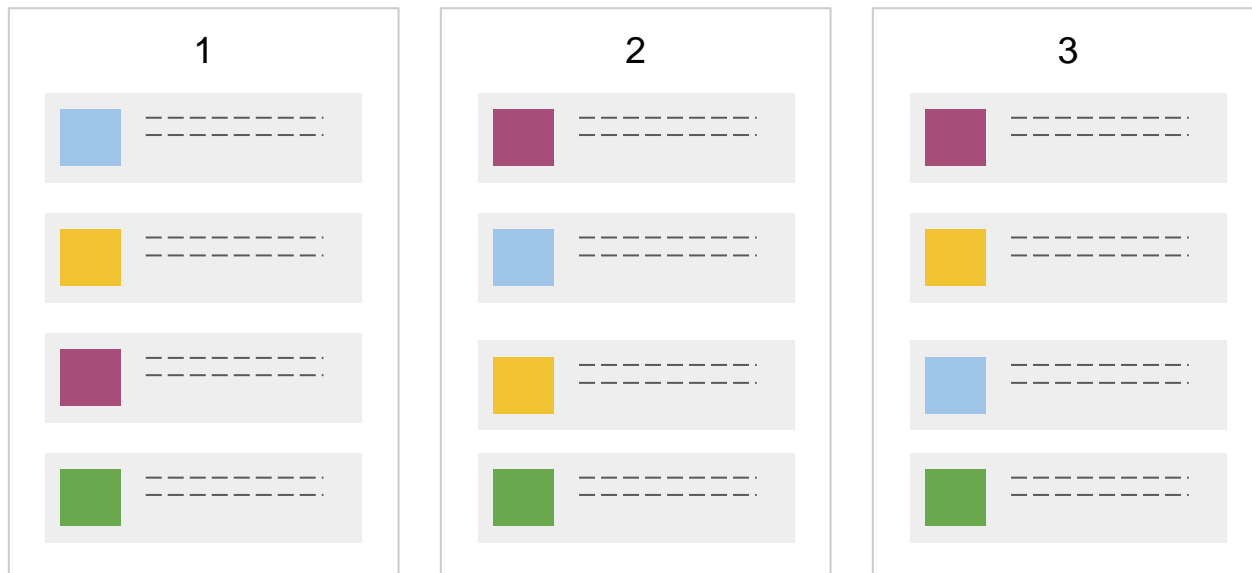
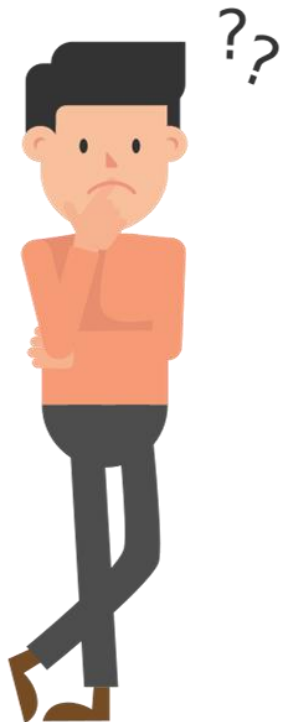
1. Eksperyment
2. Ranking - co to jest?
3. Dane
4. Metryka
5. Model
6. Problemy danych i modelu

Eksperyment



Ranking

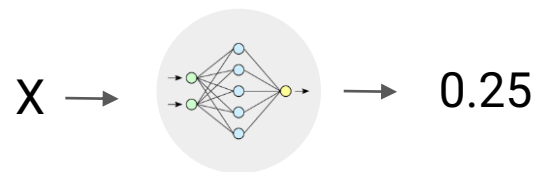
czyli kolejność gra rolę



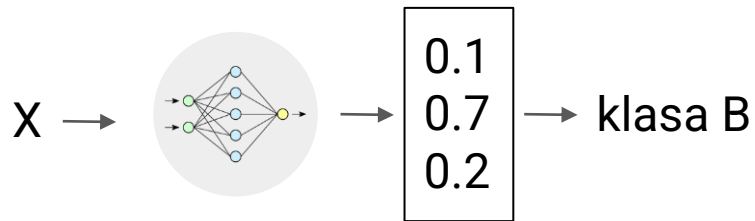
Wszędzie tam, gdzie ważna jest kolejność, czyli:

- wyszukiwarki (Google, Yahoo, Allegro, AirBnb, Spotify, Facebook, Gmail)
- rekomendacje (j.w. :))
- analiza ryzyka
- systemy question answering
- ...

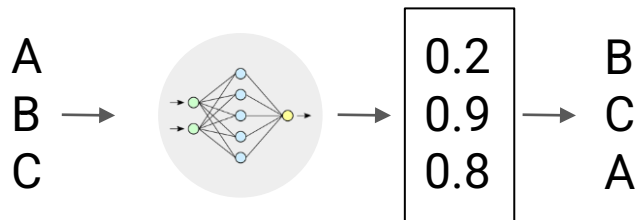
regresja



klasyfikacja



ranking



Dane

czyli jeden rabin powie tak,
drugi rabin powie nie

Wyszukiwarka stron internetowych



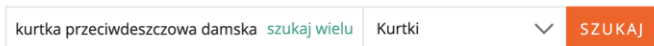
- treść zapytania
- treść strony
- popularność strony

Wyszukiwarka stron internetowych



- treść zapytania
- treść strony
- popularność strony

Wyszukiwarka ofert



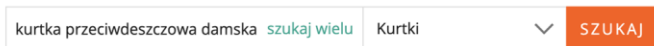
- treść zapytania
- tytuł oferty
- ocena sprzedającego
- cena oferty

Wyszukiwarka stron internetowych



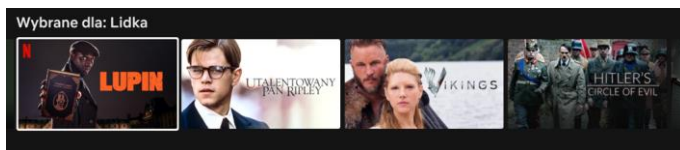
- treść zapytania
- treść strony
- popularność strony

Wyszukiwarka ofert



- treść zapytania
- tytuł oferty
- ocena sprzedającego
- cena oferty

Rekomendacje filmów



- historia użytkownika
- popularność filmu
- lista filmów "do obejrzenia" użytkownika

Skąd bierzemy etykiety do treningu?

Experts annotation	Explicit feedback	Implicit feedback
Małe zbiory danych	Średniej wielkości zbiory danych	Bardzo dużo danych
Drogi (wykorzystujemy płatnych "ekspertów")	Tani	Bardzo tani
Dane dobrej jakości	Dane średniej jakości	Dane niskiej jakości
Wymaga udziału ekspertów	Wymaga dodatkowej interakcji użytkownika	Nie wymaga dodatkowej interakcji użytkownika



Metryka

czyli jak porównać rankingi?

NDCG

Normalized Discounted Cumulative Gain

Trafne dokumenty powinny być wyżej w rankingu.

Bardzo trafne dokumenty są bardziej użyteczne niż **trochę trafne** dokumenty, a te z kolei są bardziej użyteczne niż **nietrafne** dokumenty.

NDCG

Normalized Discounted **Cumulative Gain**

$$A = [1, 0, 1, 0, 0]$$

$$B = [0, 1, 0, 1, 1]$$

$$\text{Cumulative Gain}(CG) = \sum_{i=1}^n \text{relevance}_i$$

$$CG(A) = 1 + 0 + 1 + 0 + 0 = 2$$

$$CG(B) = 0 + 1 + 0 + 1 + 1 = 3$$

NDCG

Normalized **Discounted Cumulative Gain**

$$A = [1, 0, 1, 0, 0]$$

$$B = [0, 1, 0, 1, 1]$$

$$DCG = \sum_{i=1}^n \frac{relevance_i}{\log_2(i+1)}$$

lub

$$DCG = \sum_{i=1}^n \frac{2^{relevance_i} - 1}{\log_2(i+1)}$$

$$DCG(A) = \frac{1}{\log_2(2)} + \frac{1}{\log_2(4)} = 1 + 0.5 = 1.5$$

$$DCG(B) = \frac{1}{\log_2(3)} + \frac{1}{\log_2(5)} + \frac{1}{\log_2(6)} = 0.63 + 0.43 + 0.39 = 1.45$$

NDCG

Normalized Discounted Cumulative Gain

$$NDCG = \frac{DCG}{iDCG}$$

$$iDCG(A) = \frac{1}{\log_2(2)} + \frac{1}{\log_2(3)} = 1 + 0.63 = 1.63$$

$$iDCG(B) = \frac{1}{\log_2(2)} + \frac{1}{\log_2(3)} + \frac{1}{\log_2(4)} = 2.13$$

$$NDCG(A) = \frac{1.5}{1.63} = 0.92$$

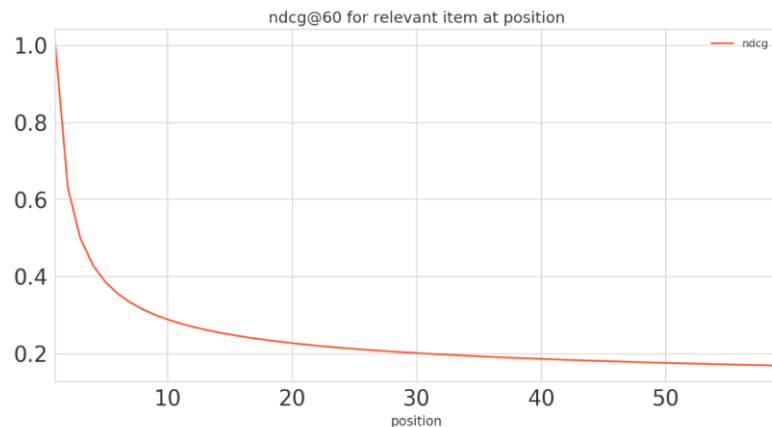
$$NDCG(B) = \frac{1.45}{2.13} = 0.68$$

$$A = [1, 0, 1, 0, 0]$$

$$B = [0, 1, 0, 1, 1]$$

$$ideal\ A = [1, 1, 0, 0, 0]$$

$$ideal\ B = [1, 1, 1, 0, 0]$$



- NDCG@K
- Średnie NDCG@K

A	0
B	1
C	0

A	0.5
B	0.9
C	0.1



B	0.9	1
A	0.5	0
C	0.1	0

model 1

$$\text{NDCG}@3 = 1$$

A	0.7
B	0.4
C	0.8



C	0.8	0
A	0.7	0
B	0.4	1

model 2

$$\text{NDCG}@3 = 0.5$$

$$\mathbf{X} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$$

$$\mathbf{R}(\mathbf{X}) = [2, 1, 0]$$

$$iDCG(X) = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)} = 2.63$$

$$\mathbf{X} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$$

$$\mathbf{R}(\mathbf{X}) = [2, 1, 0]$$

$$\mathbf{R}'(\mathbf{X}) = [2, 1, x]$$

$$iDCG(X) = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)} = 2.63$$

$$X = [A, B, C]$$

$$R(X) = [2, 1, 0]$$

$$R'(X) = [2, 1, x]$$

$$iDCG(X) = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)} = 2.63$$

$$1. \quad x \in (0, 1)$$

$$NDCG(X') = 1$$

$$X' = [A, B, C]$$

$$X = [A, B, C]$$

$$R(X) = [2, 1, 0]$$

$$R'(X) = [2, 1, x]$$

$$iDCG(X) = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)} = 2.63$$

$$1. \quad x \in \langle 0, 1 \rangle$$

$$X' = [A, B, C]$$

$$NDCG(X') = 1$$

$$1. \quad x \in \langle 1, 2 \rangle$$

$$X' = [A, C, B]$$

$$DCG(X') = \frac{2}{\log_2(2)} + \frac{1}{\log_2(4)} = 2.5$$

$$NDCG(X') = 2.5/2.63 = 0.95$$

$$X = [A, B, C]$$

$$R(X) = [2, 1, 0]$$

$$R'(X) = [2, 1, x]$$

$$iDCG(X) = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)} = 2.63$$

$$1. \quad x \in \langle 0, 1 \rangle$$

$$X' = [A, B, C]$$

$$NDCG(X') = 1$$

$$1. \quad x \in \langle 1, 2 \rangle$$

$$X' = [A, C, B]$$

$$DCG(X') = \frac{2}{\log_2(2)} + \frac{1}{\log_2(4)} = 2.5$$

$$NDCG(X') = 2.5/2.63 = 0.95$$

$$1. \quad x \in \langle 2, +\infty \rangle$$

$$X' = [C, A, B]$$

$$DCG(X') = \frac{2}{\log_2(3)} + \frac{1}{\log_2(4)} = 1.76$$

$$NDCG(X') = 1.76/2.63 = 0.67$$

$$X = [A, B, C]$$

$$R(X) = [2, 1, 0]$$

$$R'(X) = [2, 1,$$

$$iDCG(X) = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)} = 2.63$$

$$1. \quad x \in \langle 0, 1 \rangle$$

$$NDCG(X') = 1$$

$$X' = [A, B, C]$$

$$1. \quad x \in \langle 1, 2 \rangle$$

$$X' = [A, C, B]$$

$$DCG(X') = \frac{2}{\log_2(2)} + \frac{1}{\log_2(4)} = 2.5$$

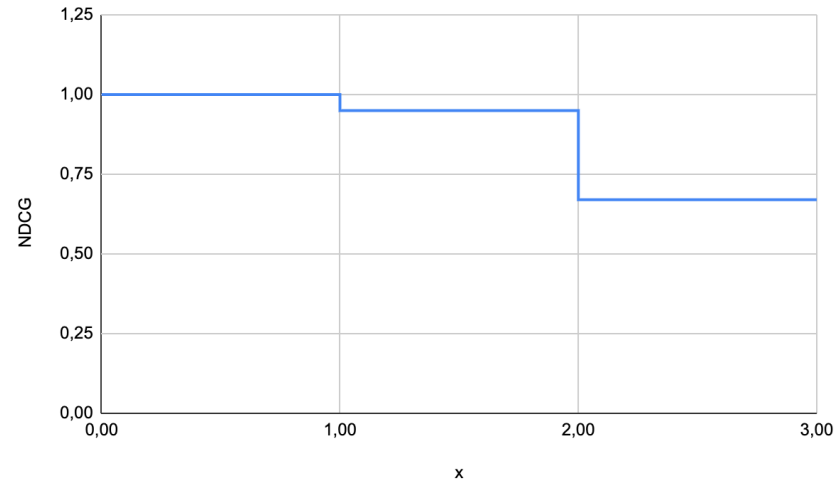
$$NDCG(X') = 2.5/2.63 = 0.95$$

$$1. \quad x \in \langle 2, +\infty \rangle$$

$$X' = [C, A, B]$$

$$DCG(X') = \frac{2}{\log_2(3)} + \frac{1}{\log_2(4)} = 1.76$$

$$NDCG(X') = 1.76/2.63 = 0.67$$



$X = [A, B, C]$

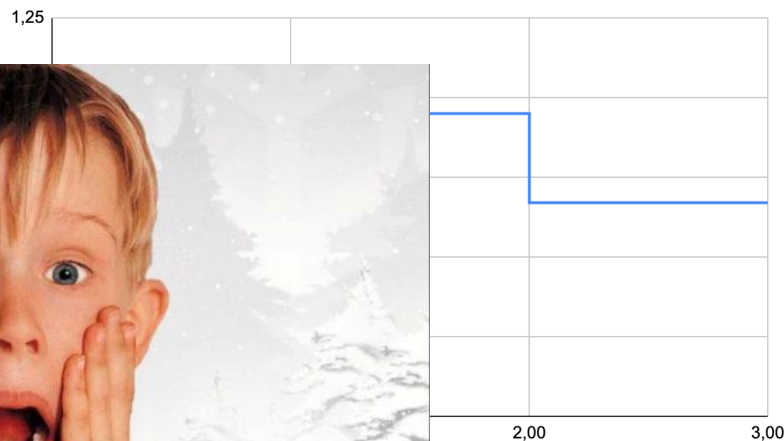
$R(X) = [2, 1, 0]$

$$iDCG(X) = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)}$$

1. $x \in \langle 0, 1 \rangle$
 $NDCG(X') = 1$

1. $x \in \langle 1, 2 \rangle$
 $DCG(X') = \frac{2}{\log_2(2)} + \frac{1}{\log_2(3)} = 2.5$
 $NDCG(X') = 2.5/2.63 = 0.95$

1. $x \in \langle 2, +\infty \rangle$
 $DCG(X') = \frac{2}{\log_2(3)} + \frac{1}{\log_2(4)} = 1.76$
 $NDCG(X') = 1.76/2.63 = 0.67$



NDCG jest nieróżniczkowalne!
Co teraz?

Podójście pointwise

- najprostsze i najstarsze
- loss obliczany dla kaźdego dokumentu osobno
- transformacja problemu rankingu w regresję lub klasyfikację (predykcja trafności dokumentu)
- przykładowy loss: MSE

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Wada: nie interesuje nas dokładny score, a tylko kolejność dokumentów. Często nie znamy nawet dokładnej trafności kaźdego dokumentu (implicit feedback!)

Podójście pairwise

- loss obliczany dla kaźdej pary dokumentów
- transformacja problemu rankingu w klasyfikację binarną (predykcja “czy dokument A jest bardziej trafny niż dokument B?”)
- model minimalizuje liczbę przypadków, w których para dokumentów jest w innej kolejności, niż by na to wskazywały etykiety
- przykładowy loss: RankNet

$$P_{ij} \equiv P(U_i \triangleright U_j) \equiv \frac{1}{1 + e^{-\sigma(s_i - s_j)}}$$

Prawdopodobieństwo tego, że jeden dokument powinien być wyżej niż drugi

$$C = -\bar{P}_{ij} \log P_{ij} - (1 - \bar{P}_{ij}) \log(1 - P_{ij})$$

Cross-entropy między rozkładami prawdopodobieństwa: predykowanym i prawdziwym

Podójście listwise

- loss obliczany dla całej listy dokumentów naraz
- bardziej bezpośrednia optymalizacja metryk rankingowych (np. NDCG) - ale nie całkiem bezpośrednia, bo operacja sortowania nie jest różniczkowalna
- możliwe opcje:
 - iteracyjne wykorzystanie pairwise lossów
 - aproksymacja NDCG funkcją ciągłą
 - zdefiniowanie lossa na przestrzeni permutacji
- przykładowy loss: ApproxNDCG

ApproxNDCG

We wzorze na NDCG możemy zamienić sumowanie po pozycjach na sumowanie po dokumentach

$$\text{NDCG}@k = N_k^{-1} \sum_{x \in \mathcal{X}} \frac{2^{r(x)} - 1}{\log_2(1 + \pi(x))} \mathbf{1}\{\pi(x) \leq k\}.$$

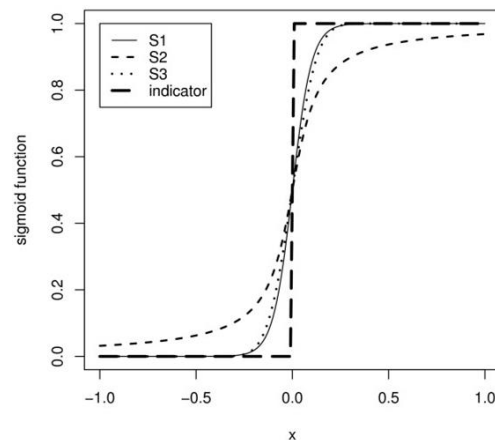
ApproxNDCG

We wzorze na NDCG możemy zamienić sumowanie po pozycjach na sumowanie po dokumentach

$$\text{NDCG}@k = N_k^{-1} \sum_{x \in \mathcal{X}} \frac{2^{r(x)} - 1}{\log_2(1 + \pi(x))} \mathbf{1}\{\pi(x) \leq k\}.$$

Funkcję pozycji (nieróżniczkowalną) można przedstawić następująco:

$$\pi(x) = 1 + \sum_{y \in \mathcal{X}, y \neq x} \mathbf{1}\{s_{x,y} < 0\} \quad s_{x,y} = s_x - s_y.$$



ApproxNDCG

We wzorze na NDCG możemy zamienić sumowanie po pozycjach na sumowanie po dokumentach

$$\text{NDCG}@k = N_k^{-1} \sum_{x \in \mathcal{X}} \frac{2^{r(x)} - 1}{\log_2(1 + \pi(x))} \mathbf{1}\{\pi(x) \leq k\}.$$

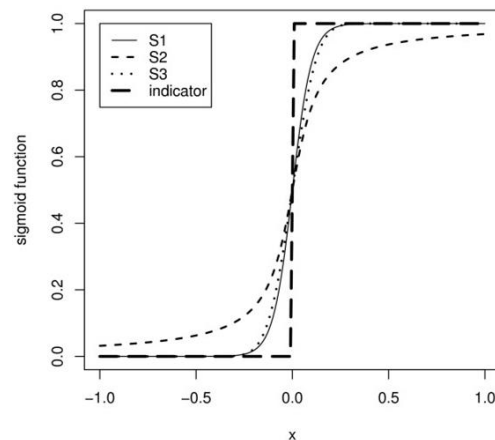
Funkcję pozycji (nieróżniczkowalną) można przedstawić następująco:

$$\pi(x) = 1 + \sum_{y \in \mathcal{X}, y \neq x} \mathbf{1}\{s_{x,y} < 0\} \quad s_{x,y} = s_x - s_y.$$

I aproksymować ją, zastępując indykator sigmoidą (różniczkowalną):

$$\hat{\pi}(x) = 1 + \sum_{y \in \mathcal{X}, y \neq x} \frac{\exp(-\alpha s_{x,y})}{1 + \exp(-\alpha s_{x,y})}$$

$$\widehat{\text{NDCG}} = N_n^{-1} \sum_{x \in \mathcal{X}} \frac{2^{r(x)} - 1}{\log_2(1 + \hat{\pi}(x))}.$$




Model





czyli mięsko

Przykładowe cechy, których używamy w rankingu ofert w wyszukiwarce Allegro:


- zapytanie użytkownika (puste, jeżeli użytkownik tylko przegląda kategorie, bez korzystania z wyszukiwarki tekstowej)
- tytuł oferty
- cena
- stan przedmiotu
- średnia ocena zgodności przedmiotu z opisem
- ...



Cechy możemy podzielić na **tekstowe**
i **numeryczne**




Lekka damska kurtka przeciwdeszczowa Regatta Pack
od  Super Sprzedawcy
Stan: Nowy Marka: Regatta Rozmiar: 42 Kolor: niebieski Rodzaj: Z membraną
109,99 zł 
zapłać później z 
 1 Moneta
117,98 zł z dostawą




dostępne warianty dostawa pojutrze 8 osób kupiło



Lekka damska kurtka przeciwdeszczowa Regatta Pack
Stan: Nowy Marka: Regatta Rozmiar: 36 Kolor: różowy Rodzaj: Z membraną
109,99 zł 
zapłać później z 
118,98 zł z dostawą

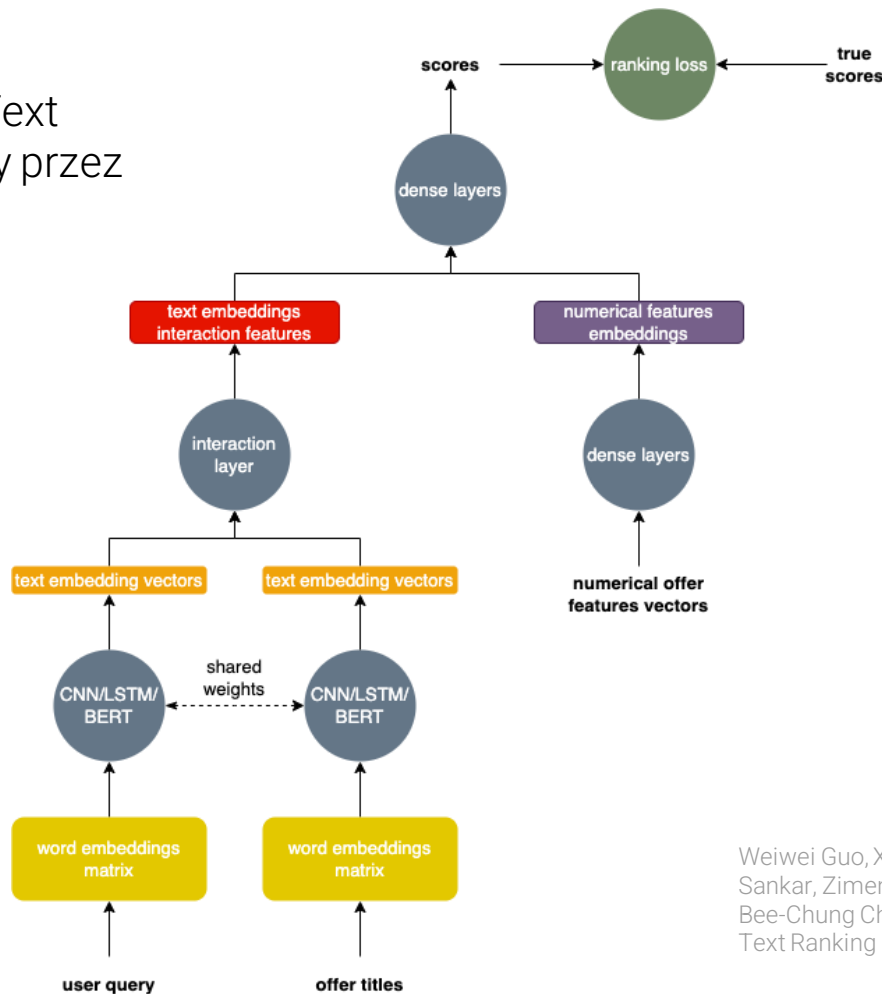
dostępne warianty 1 osoba kupiła



REGATTA Kurtka Damska Przeciwdeszczowa CORINNE IV
od  Super Sprzedawcy
Stan: Nowy Marka: Regatta Rozmiar: XS Kolor: różowy Rodzaj: Z membraną
160,00 zł 
zapłać później z 
168,99 zł z dostawą

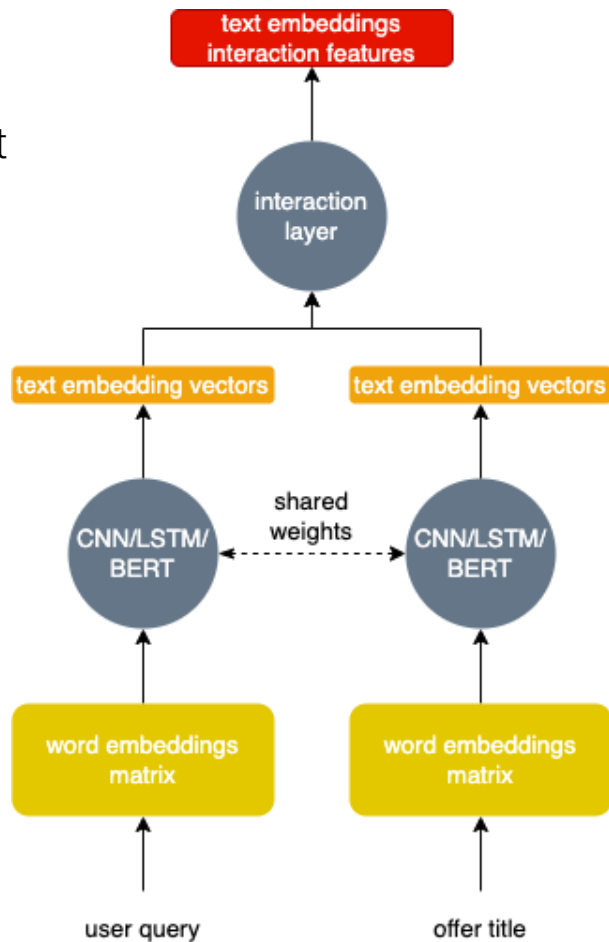
dostawa pojutrze

Cały model: DeText
(zaproponowany przez
LinkedIn)



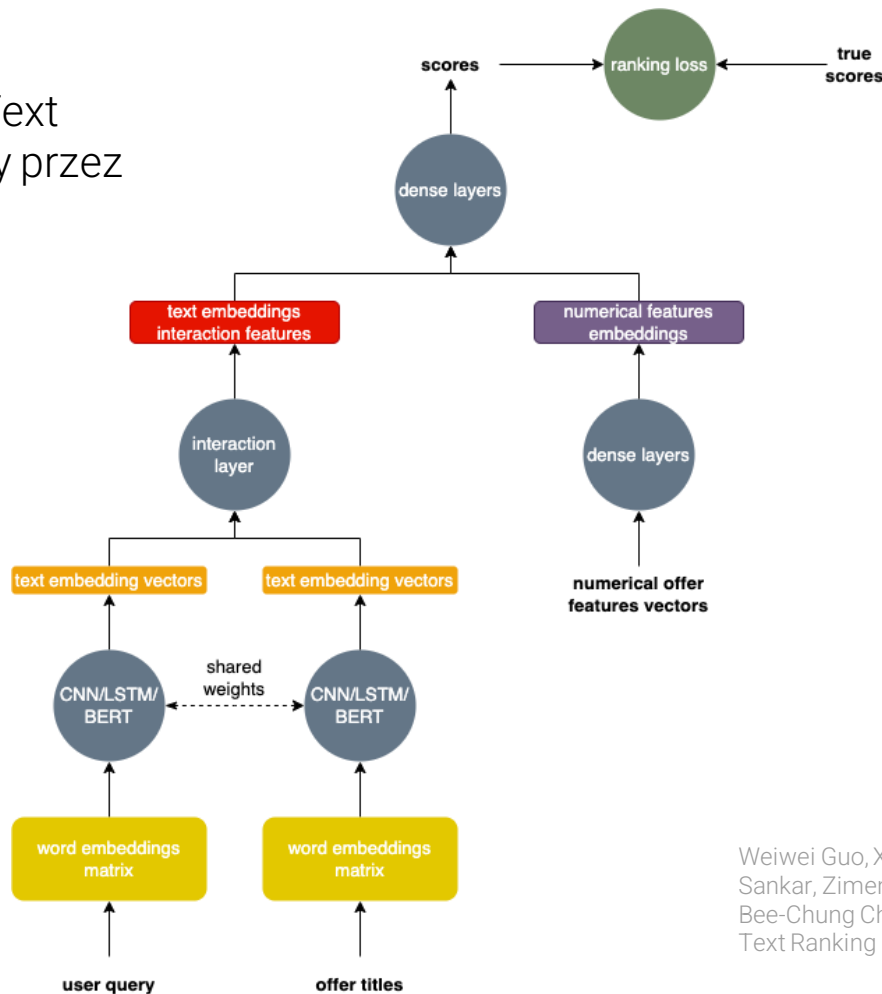
Weiwei Guo, Xiaowei Liu, Sida Wang, Huiji Gao, Ananth Sankar, Zimeng Yang, Qi Guo, Liang Zhang, Bo Long, Bee-Chung Chen and Deepak Agarwal. "DeText: A Deep Text Ranking Framework with BERT". In CIKM 2020.

Część modelu
przetwarzająca tekst



Weiwei Guo, Xiaowei Liu, Sida Wang, Huiji Gao, Ananth Sankar, Zimeng Yang, Qi Guo, Liang Zhang, Bo Long, Bee-Chung Chen and Deepak Agarwal. "DeText: A Deep Text Ranking Framework with BERT". In CIKM 2020.

Cały model: DeText
(zaproponowany przez
LinkedIn)



Weiwei Guo, Xiaowei Liu, Sida Wang, Huiji Gao, Ananth Sankar, Zimeng Yang, Qi Guo, Liang Zhang, Bo Long, Bee-Chung Chen and Deepak Agarwal. "DeText: A Deep Text Ranking Framework with BERT". In CIKM 2020.

Problemy

żeby nie było za łatwo

Szum w danych

- różne intencje potencjalnych kupujących podczas przeglądania ofert - mogą chcieć np.:
 - zainspirować się
 - zorientować się w cenach danego produktu
 - kupić wybrany uprzednio przedmiot
 - porównać kilka ofert danego przedmiotu i kupić go
- clickbaity
- presentation bias



Presentation bias

To wpływ graficznej prezentacji listy wyników na jej odbiór przez użytkowników

- **graficzne wyróżniki i ozdobniki** ←
- kolejność
- umieszczenie w określonym kontekście (np. **tańszy** przedmiot w otoczeniu **drogich**)



Position bias - jeden z możliwych presentation biasów

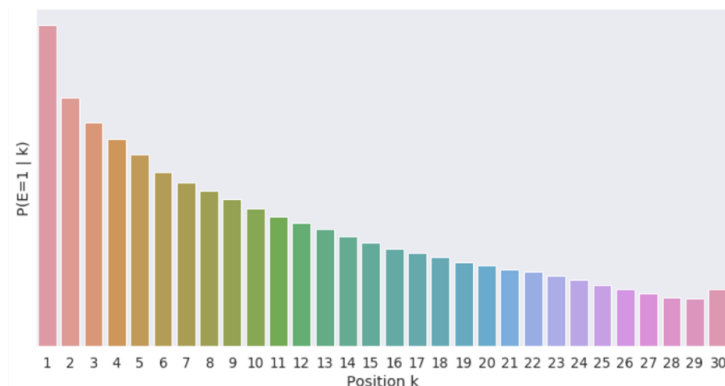
Użytkownicy łatwiej zauważają to, co jest na szczycie listy wyników, więc są bardziej skłonni klikać oferty będące wyżej

Jak sobie z tym radzić?

- eksperymenty randomizacyjne (trudne do zastosowania w praktyce)

$$P(C = 1|Hotel, k) = P(E = 1|k) \cdot P(R = 1|Hotel)$$

- click propensity, czyli skłonność do klikania, można też estymować bez eksperymentów randomizacyjnych
- Inverse Propensity Weighing (IPW)

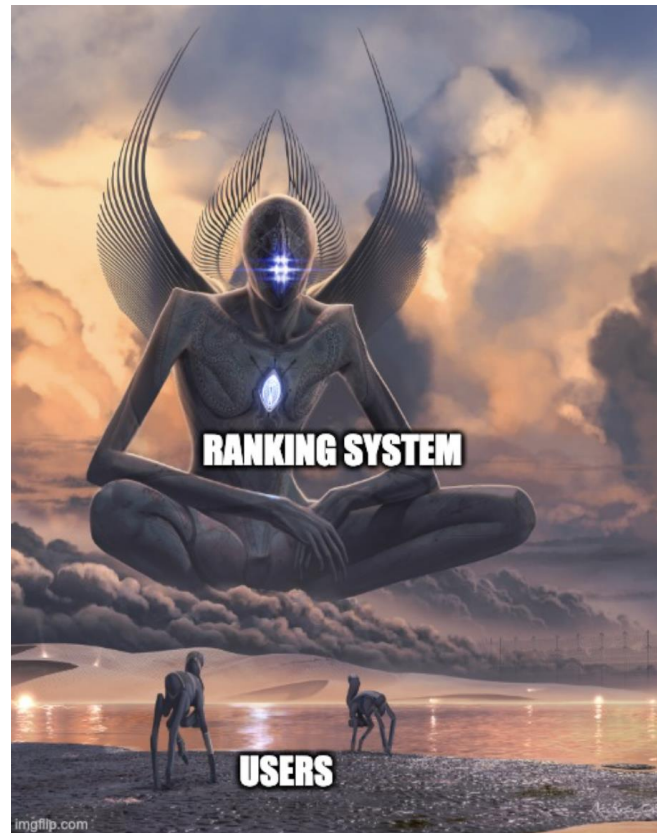


Trust bias

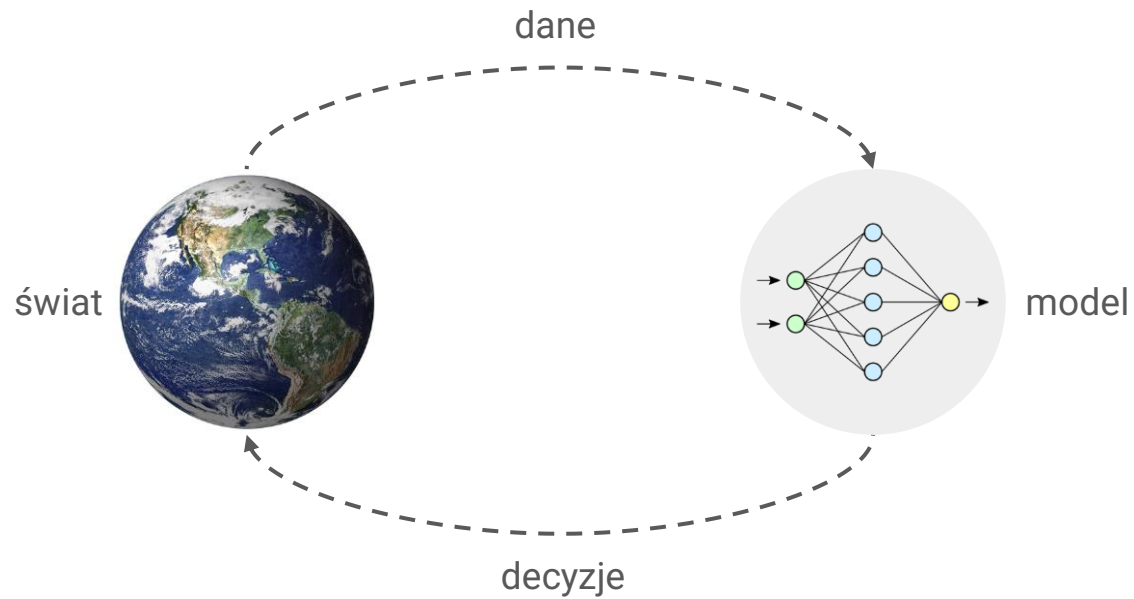
Użytkownik wierzy w mądrość systemu rankingowego i jest skory do kliknięcia w wynik o (zauważonej przez niego) niskiej trafności, jeżeli jest on wysoko na liście.

Innymi słowy: jeżeli wynik słabo pasuje do zapytania, to im będzie wyżej, tym większa szansa, że użytkownik kliknie w niego z powodu zaufania do systemu rankingowego.

Aman Agarwal, Xuanhui Wang, Cheng Li, Michael Bendersky, and Marc Najork. 2019. Addressing Trust Bias for Unbiased Learning-to-Rank. In The World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 4–14.
<https://doi.org/10.1145/3308558.3313697>



Sprzężenie zwrotne



Jak niwelować niekorzystny wpływ sprzężenia zwrotnego?

- uczyć się na nieobciążonych danych (czyli na wiedzy, która nie pochodzi ze świata, na który ma wpływ model) - to może być produkcyjnie ciężkie do zrobienia
- być świadomym biasów w danych i reagować na nie odpowiednio (np. IPW w biasie pozycyjnym)
- w szczególności w przypadku rankingu: nie polegać wyłącznie na popularności sortowanych dokumentów

Dodatkowa bibliografia:

1. Valizadegan, Hamed & Jin, Rong & Zhang, Ruofei & Mao, Jianchang. (2009). Learning to Rank by Optimizing NDCG Measure. 1883-1891
(<https://papers.nips.cc/paper/2009/file/b3967a0e938dc2a6340e258630febd5a-Paper.pdf>)
2. Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. Accurately interpreting clickthrough data as implicit feedback. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05). Association for Computing Machinery, New York, NY, USA, 154–161.
(https://www.cs.cornell.edu/people/tj/publications/joachims_etal_05a.pdf)
3. <https://towardsdatascience.com/evaluate-your-recommendation-engine-using-ndcg-759a851452d1>
4. <https://towardsdatascience.com/learning-to-rank-a-complete-guide-to-ranking-using-machine-learning-4c9688d370d4>
5. <https://medium.com/@nikhilbd/intuitive-explanation-of-learning-to-rank-and-ranknet-lambdarank-and-lambdamart-fe1e17fac418>
6. <https://levelup.gitconnected.com/how-positive-feedback-loops-are-hurting-ai-applications-6eae0304521c>

