News à la carte?

developing recommender systems
for news articles

Pascal Engl & Tim A. Peplies

Motivation

- in the **Age of Information**, recommendations are generally extremely important

- in the context of news articles, **regular challenges for recommender systems are aggravated**: because more classical algorithms rely on previous interactions between users and items (articles), it is **difficult to give recommendations for *new* items and/or users**

- also, the **societal implications of personalized news consumption** have to be taken into account

Goals

- the **implementation and discussion of classical and deep learning** recommendation **methods**

- **technical improvements** for *news* recommendation

- conceptualization of **revisions concerning socio-psychological consequences**

The Data

the MIcrosoft News Dataset
we worked with:

click log of ~ 575,000 **anonymous** Microsoft News **readers**

~ 100,000 individual **news articles**

The Data

the information we worked with:


WEEKLY WORLD
NEWS
40¢
NEW ROMANCE SERIAL
Thrill to the passion of a forbidden love!
Chilling encounter!
I TOUCHED BIGFOOT
Terrified teen recalls a moment of shrieking horror
Patty Hearst tells how terrorists made her their sex slave
How you can land the job of your dreams
Secret reports reveal Soviets have developed army of killer zombies
CENTERFOLD
General Patton's son saved from death by ghost of his father
Experts fear outbreak of giant flu epidemic — from outer space
'One Day at a Time' star: Why all the critics were wrong about my marriage

title

abstract

category

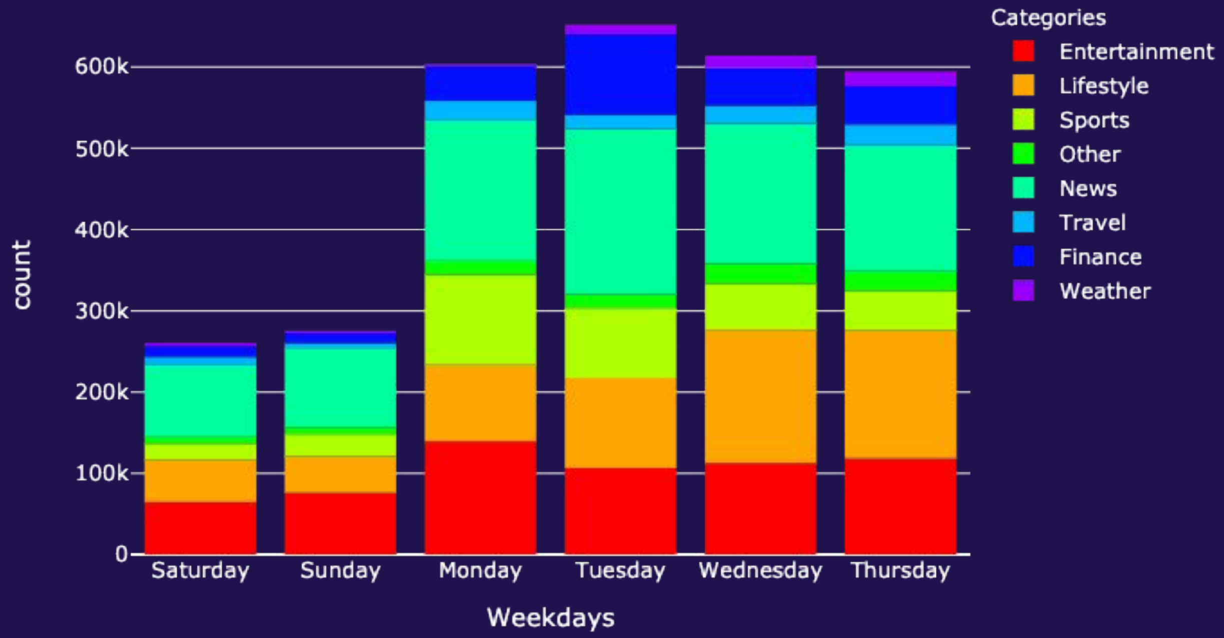
click behavior

reading history

date and daytime

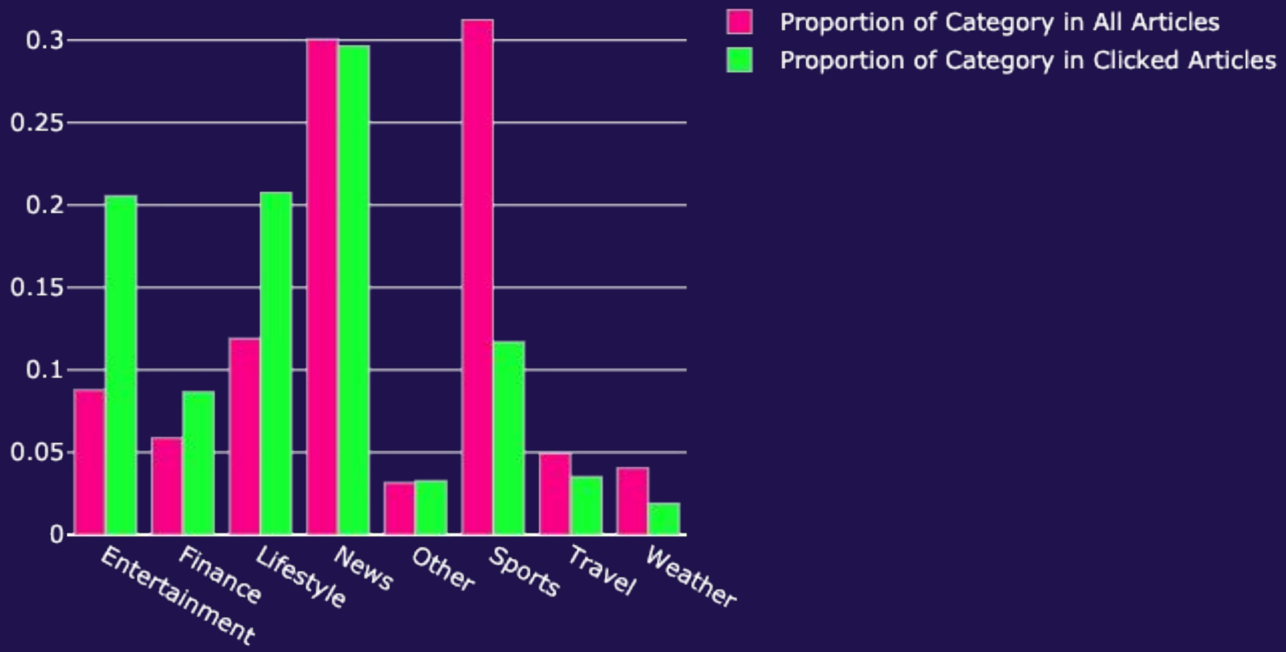
Clicks on Weekdays

there are twice as many clicks on working days and there seems to be a higher interest in sports news during the week as well



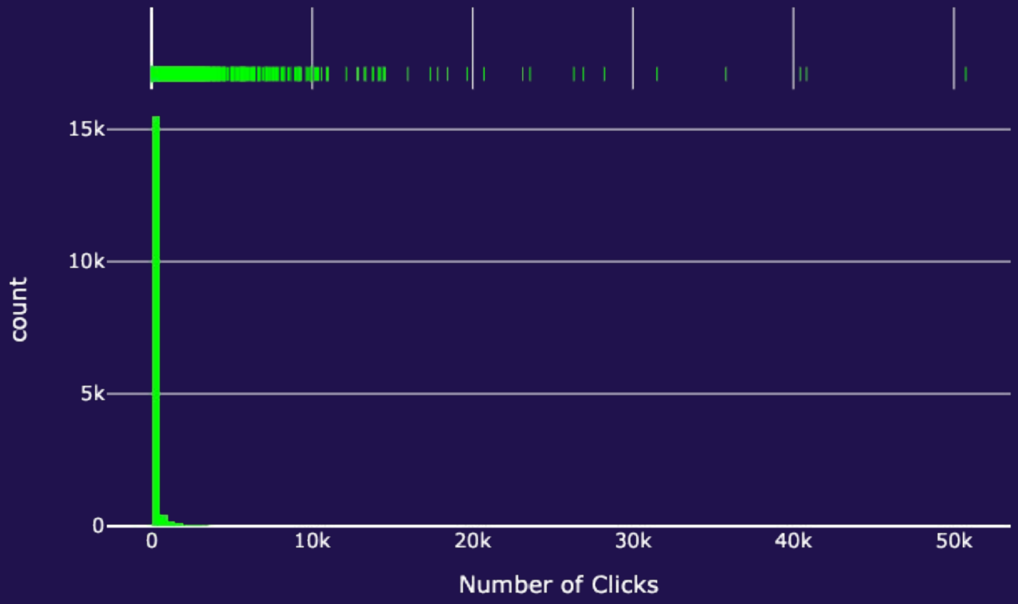
Available News vs. Clicked News

sports articles are clearly over-represented in news supply, whereas lifestyle, entertainment and finance news are proportionally more clicked

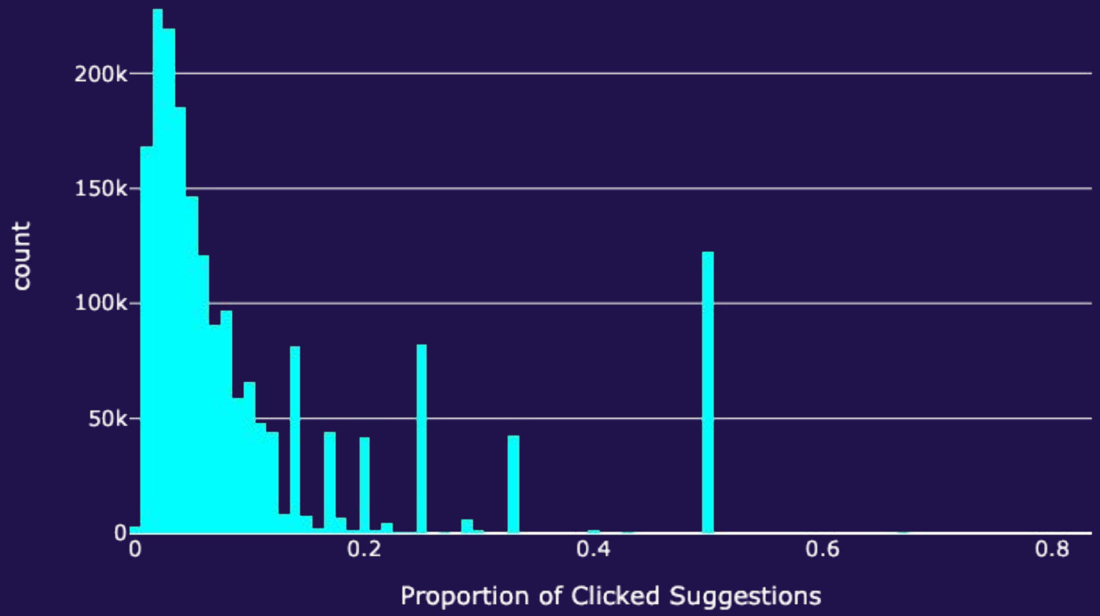


Clicked Articles at Session

there are many articles with fewer than 500 clicks and only very few with tens of thousands of clicks



Click Ratios at Sessions



there are many session where less than 20% of suggestions have been clicked

The Way We Tested

in order to assess our models' performances, we had to **devise a testing procedure**

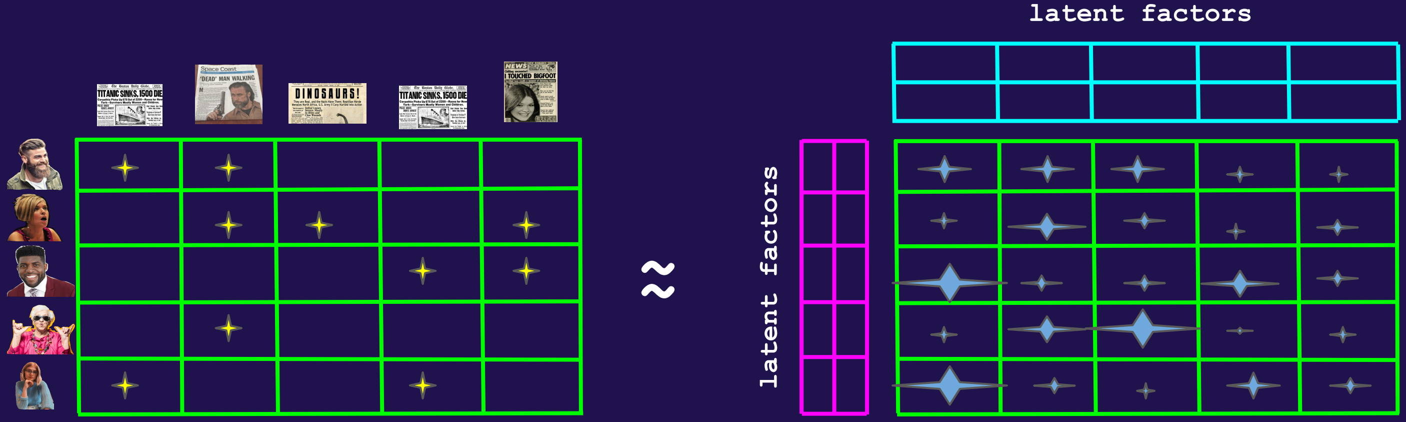
the targeted metric was **Mean Reciprocal Rank**, which indicates how high a known relevant document ranks under assumably irrelevant ones on average

we thus **excluded one reader-article interaction for every user from training** and **ranked** our models prediction for this interaction **under the prediction for 99 known non-interactions** and
we tested on a subset of MIND, containing ~50.000 Articles and ~150.000 Users

Collaborative Filtering

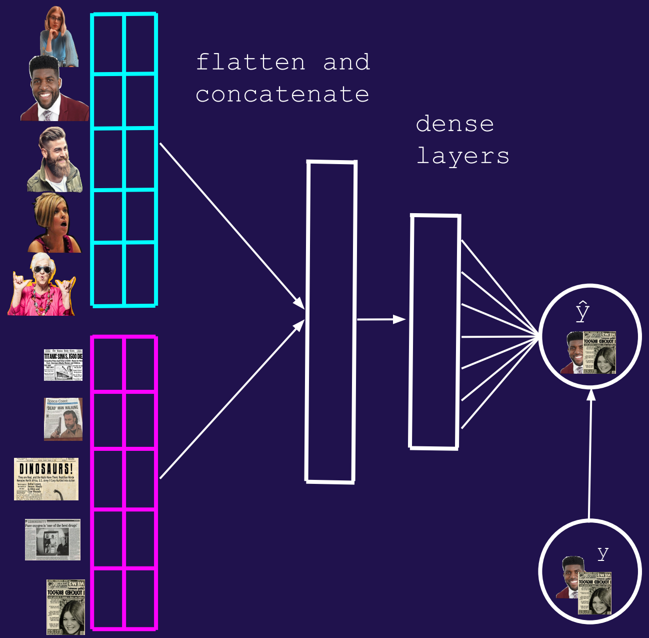
with a **matrix factorization** approach, we generate a densely populated matrix from a sparse user-article-interaction matrix, by **learning two smaller embedding matrices, representing k latent factors**

using **LightFM***, the **best mrr (0,06)** was attained using weighted approximate rank pairwise loss, employing 20 latent factors for 20 epochs



* Kula 2017: *Metadata Embeddings for User and Item Cold-start Recommendations*

Neural Collaborative Filtering



with **neural collaborative filtering***, we still have have 'rigid' embeddings for users and articles, but now, they can have **different numbers of dimensions**, **non linear relationships** can be learned and **latent factors will be weighted**

with an architecture of **64, 32, 16, 8**, a learning rate of **0.001**, **binary crossentropy** as the objective function and **adam** as optimizer, we reached a **mrr of 0.39**

first results on the large dataset showed an improvement of ~ 0.1

* Xiangnan He et al. 2017: *Neural Collaborative Filtering*

Learning Article Trajectories via RNNs

Jeff's Timeline

"Renovations To Make and Skip Before Selling Your Home"

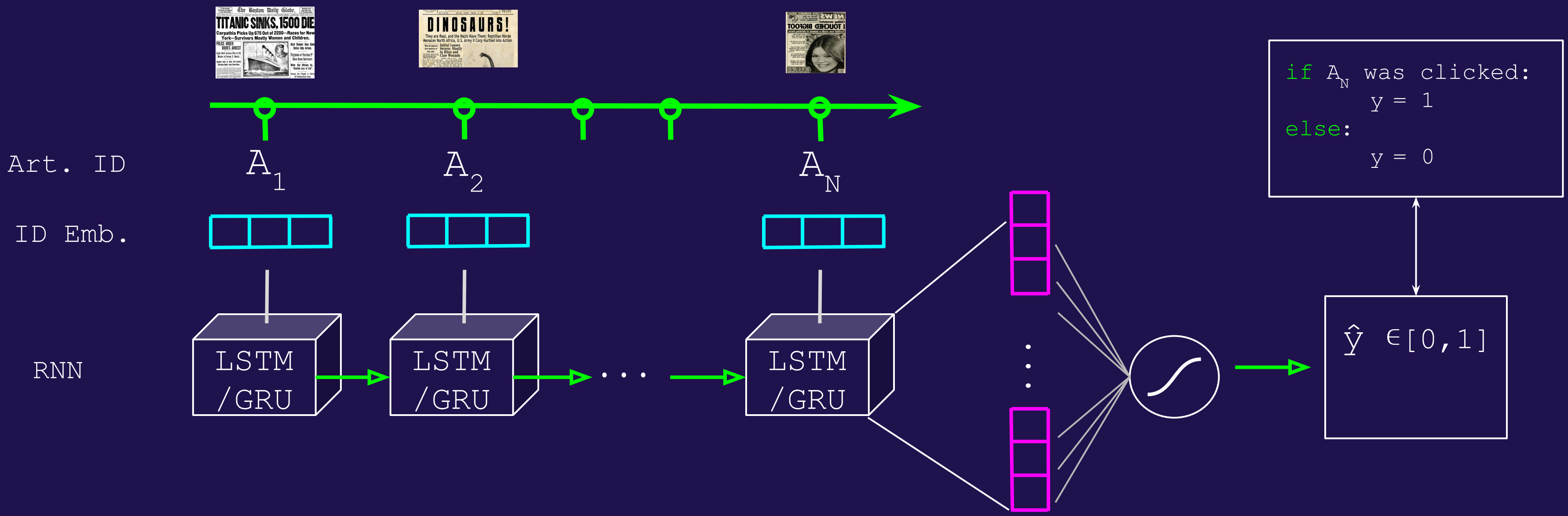
"What Came First, the Chicken or the Dumpling?"

"Dad thought he dressed toddler daughter in a hat, but he was hilariously wrong!"

"Dog Learning to Talk By Using a Custom Soundboard to Speak: 'I'm in Constant Amazement'"

?

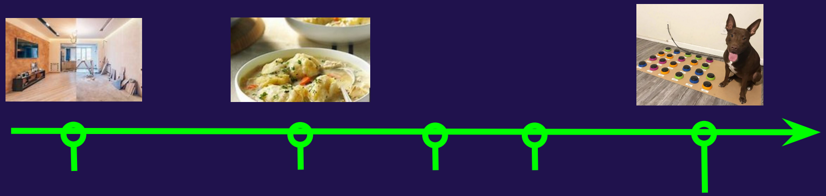
Learning Article Trajectories via RNNs



with learning article trajectories, the goal is to free our models from rigid user embeddings and**represent the users via their reading histories**
here, we took into account the clicked articles at a session and learned them against the not-clicked suggestions

with article emb. dim. of **64, LSTM dim. of 128 and one hidden layer with 10 neurons**:
mrr ~ 0.2

Reading Trajectories with a Title Encoder

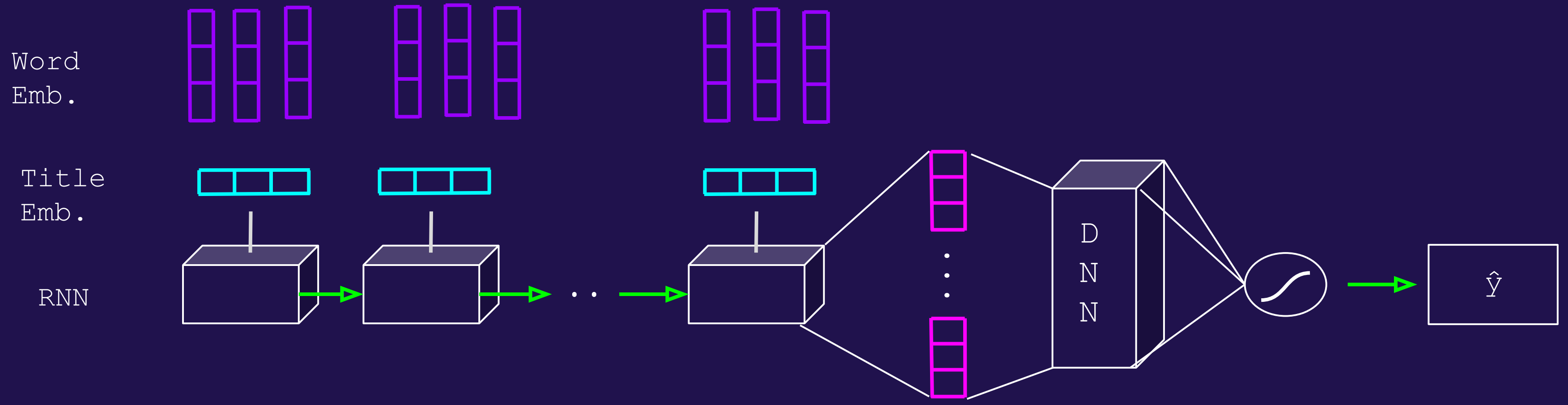


Title

"Renovations To Make and Skip Before Selling Your Home"

"What Came First, the Chicken or the Dumpling?"

"Dog Learning to Talk By Using a Custom Soundboard to Speak: 'I'm in Constant Amazement' "



the utilization of a text encoder even **frees us from fixed article representations**: with this architecture, we can easily now **process new articles and users, needing only a small number of clicked** (and potentially new) **articles**

Outlook

- the RNN approach can be taken much further: full text, entity embeddings, picture and video data (where available) can be utilized to distill **more fine grained and holistic consumption trajectories** (CNNs could be applied here as well)

- when really needed, **post filtering with sentiment analysis and even partial recommendation of relatively distant content** can be undertaken to counter the emergence of extreme filter bubbles

- especially in context of ***news*** recommendation, **information on time of publication should be included**!

Thank You!