Leveraging Emotion Features in News Recommendations

Nastaran Babanejad[1], Ameeta Agrawal[1], Heidar Davoudi[2], Aijun An[1], Manos Papagelis[1]

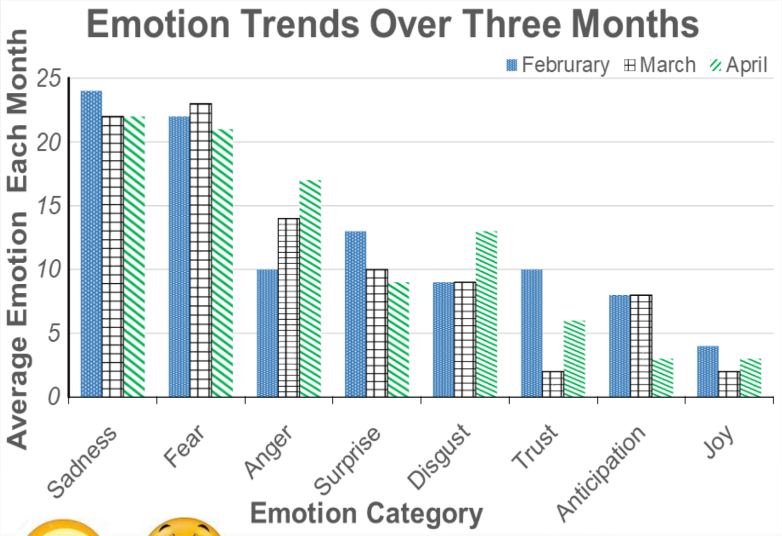
{nasba,ameeta,aan,papaggel}@eecs.yorku.ca, heidar.davoudi@uoit.ca
[1]Department of Electrical Engineering and Computer Science, York University Canada; [2]Ontario Tech University, Canada

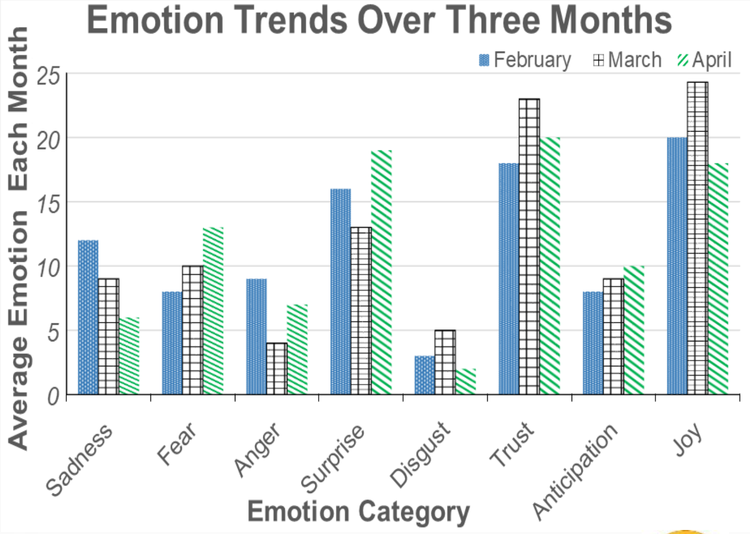
Outline

- **Motivation**
- Objective
- Challenges
- Contributions
- Framework
- Proposed Model (EmoRec)
- Experiments and Results
- Conclusion and Future Works

Motivation

Emotional profiles of two users, **User 1** and **User 2**, based on **eight basic emotions (Plutchick Emotion Model)**, expressed in articles read by them over a period of **three months**.






User 1

User 2


Outline

- Motivation
- **Objective**
- Challenges
- Contributions
- Framework
- Proposed Model (EmoRec)
- Experiments and Results
- Conclusion and Future Works

Objective

Objective: the goal is to investigate whether, how and to what extent emotion features can improve the accuracy of recommendations. By assuming that each user $\boldsymbol{u_i}$ has already interacted with a set of items $\boldsymbol{Iu_i \subseteq I}$, then the problem is to accurately predict the probability of recommending an unread article by:

$$\mathcal{P}u_a,i_j$$

a user	$u_a \in U$
an item	$i_j \in \boldsymbol{I} \setminus Iu_a$
set of m users	$U = \{u_1, u_2, \ldots, u_m\}$
set of n items	$\boldsymbol{I} = \{i_1, i_2, \ldots, i_n\}$

Outline

- Motivation
- Objective
- **Challenges**
- Contributions
- Framework
- Proposed Model (EmoRec)
- Experiments and Results
- Conclusion and Future Works

Challenges

1) In order to evaluate the importance of the emotional context to recommendations, we have to incorporate emotional features to state-of-the-art recommendation algorithms and evaluate their accuracy performance.

2) We have to generate a number of features attributed to both users and items.

Outline

- Motivation
- Objective
- Challenges
- **Contributions**
- Framework
- Proposed Model (EmoRec)
- Experiments and Results
- Conclusion and Future Works

Contributions

1) We define the new problem of emotion-aware recommendation model in digital news media. To the best of our knowledge, it has not been studied before.

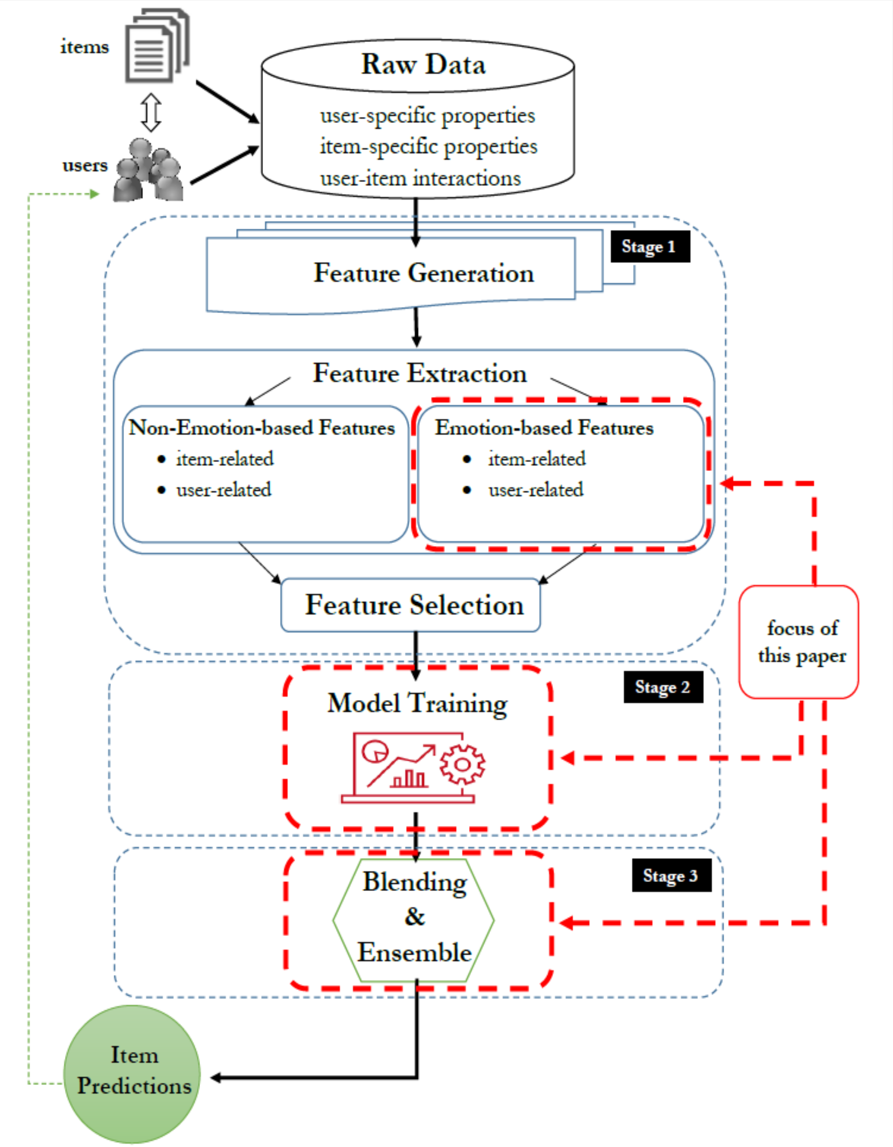
2) We systematically identify, extract and select the most relevant emotion-based features for use in news recommendation models. These features are associated with both items (e.g., news articles) and users (e.g., readers).

3) The proposed framework is applied to a real dataset obtained from The Globe and Mail and we propose EmoRec, an emotion-aware recommendation model.

Outline

- Motivation
- Objective
- Challenges
- Contributions
- **Framework**
- Proposed Model (EmoRec)
- Experiments and Results
- Conclusion and Future Works

Framework


items
users
Raw Data
user-specific properties
item-specific properties
user-item interactions
Stage 1
Feature Generation
Feature Extraction
Non-Emotion-based Features
item-related
user-related
Emotion-based Features
item-related
user-related
Feature Selection
focus of this paper
Stage 2
Model Training
Stage 3
Blending & Ensemble
Item Predictions


NEWS

Emotions in News Articles

An example of textual content of items (i.e., an article) in news domains.

excerpt of news article

"It was [**gratifying**] **yay**… Who wouldn't want it to keep going?" Walker told NBA.com. "**Oh**, I know teams will be gearing up on me and double-teaming me. But I just want to **win**, man. I want to get back to the playoffs any way possible. I don't care what I average the rest of the year."

- There are several words such as win and gratifying, expressing the emotion of happiness.
- Moreover, interjections such as yay and oh can be indicators of different emotions.

Item Emotion-based Features

- **Number of Emotion Words:** This feature represents the number of words in an emotion lexicon (i.e., WordNet-Affect) that occur in the item (i.e., news article) more than once.

- **Interjections:** This feature counts the number of interjections in a document. A short sound, word or phrase spoken suddenly to express an emotion, e.g., oh, look out!, ah, are called interjections.

- **Capitalized Words:** This feature counts the number of words in a document with all uppercase characters. (e.g., I said I am FINE).

- **Punctuation:** Two features are included to model the occurrence of question marks and exclamation marks repeated more than two times in a document.

Item Emotion-based Features

Plutchik Emotion Scores: First, we measure the semantic relatedness score between a word W_i in the text and an emotion category C_j in the NRC lexicon (see Table 1) as follows [1]:

$$PMI(W_i, C_j) = \sqrt[n]{\prod_{k=1}^{n} PMI(W_i, C_j^k)} \qquad (1)$$

where C_j^k $(k = 1 \ldots n)$ is the k^{th} word of emotion category C_j. PMI is the Pointwise Mutual Information calculated as follows:

$$PMI(W_i, C_j^k) = \log \frac{P(W_i, C_j^k)}{P(W_i)P(C_j^k)} \qquad (2)$$

where $P(W_i)$ and $P(C_j^k)$ are the probabilities that W_i and C_j^k occur in a text corpus, respectively, and $P(W_i, C_j^k)$ is the probability that W_i and C_j^k co-occur within a sliding window in the corpus. Finally, we calculate the average, maximum and minimum of score for all words in the text for each emotion category and consider each as an individual feature.

User Emotion-based Features

- **User Emotions Across Items:** We determine the emotion score (i.e., Plutchik's emotion scores) for the last accessed item before subscription as well as for the last 20 items accessed by the user. Then, we pick the top 3 frequent emotions.

- **User Emotions Across Categories:** We determine the emotion of categories of items (e.g., sports in news domain) accessed by a user by counting the number of items assigned to an emotion in a specific category, with the most frequent emotion considered as the emotion of the category. The feature is calculated for the whole history of the user.

Item Non-Emotion-based Features

- **Potential to Trigger Subscription:** This feature represents the total number of times the item was requested right before a **paywall** was presented to a user who subsequently made a subscription.

- **Item Topic:** We extract topics in the article using Latent Dirichlet Allocation (LDA). In LDA, each topic is a distribution over words, and each document is a mixture of topics. The number of topics for the news articles are 112 , which were chosen empirically to minimize the perplexity score of the LDA result. Thus, the item topic is represented by a **vector of length 112**.

User Non-emotion-based Features

- **Visit Count:** We calculate the average number of items (articles) accessed by a user per visit. A visit is terminated if a user is inactive for more than 30 minutes.

- **User Spent Time:** Two features are represented. One is the average time the user spent per item, and the other is the average time the user spent per visit.

- **User Interest in Subcategory:** This feature represents the empirical probability of subcategory s given a user u and a category c denoted as P(s |u, c). For example, P(election|u, politics) can be determined by the total number of articles the user read on election over the total number of articles that the user read on politics.

Feature Importance (XGBoost)

Emotion Features	Gain Score
Plutchik emotion scores	3200.86
User emotions across items	1985.36
User emotions across categories	1850.33
Ekman's emotion label	1101.38
Punctuation	910.55
Grammatical markers and extended words	860.13
Interjections	773.12
Capitalized words	640.21
Mixed emotions	526.97
Sentiment features	360.68
Non-emotion Features	**Gain Score**
User latent vector	3640.87
Potential to trigger subscription	2974.46
User interest in subcategory	1530.28
Topic labeling	1421.19
User spent time	1110.57
Visit count	920.53
Item topic	867.12
Coherence	685.23
TF-IDF	410.29

Outline

- Motivation
- Objective
- Challenges
- Contributions
- Framework
- **Proposed Model (EmoRec)**
- Experiments and Results
- Conclusion and Future Works

Proposed Model (EmoRec)

The final model EmoRec is the weighted average of the three models' predictions. We use Nelder-Mead Method to find the optimum weights of each models.

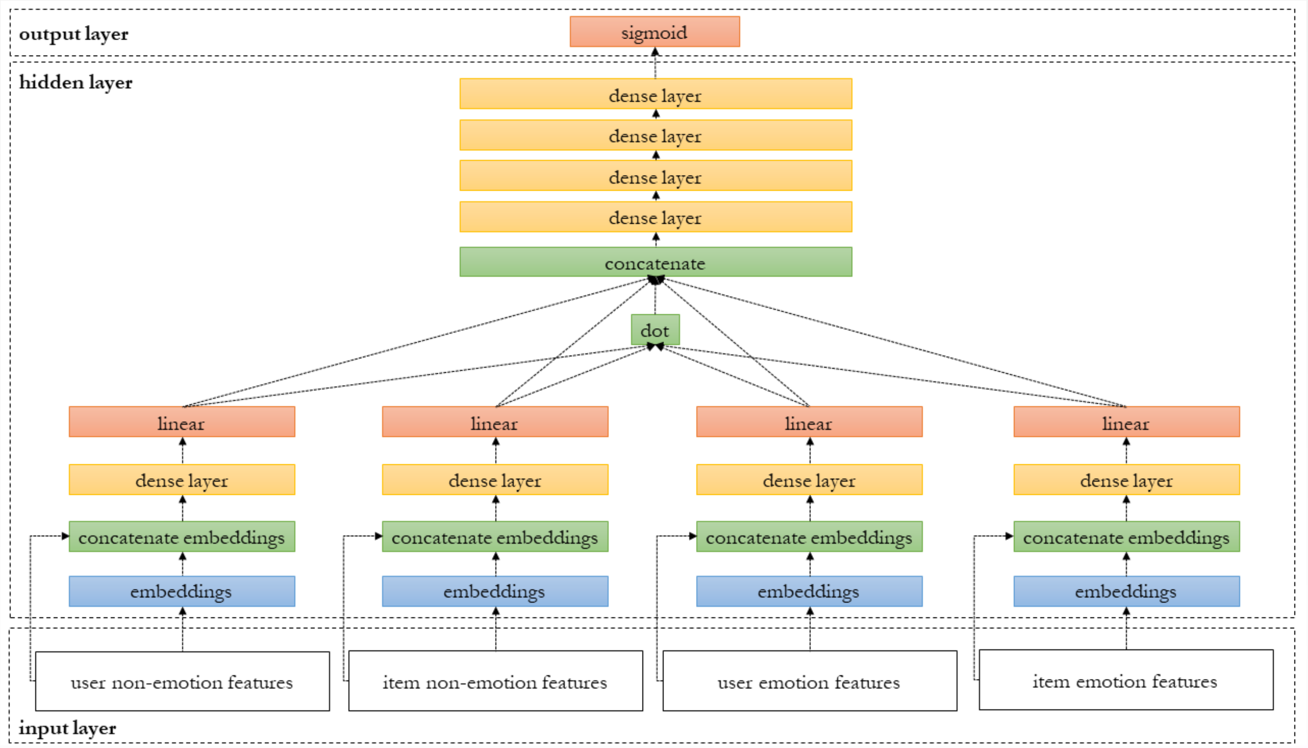
Model 1 (Boost Model):

- We train both **XGBoost** and **Catboost** individually (three base models for each). The final model output (i.e., probability that a user is interested in an item) is the combination of all base models outcomes:

$$\sum_{i}^{6} \alpha_i p_i$$

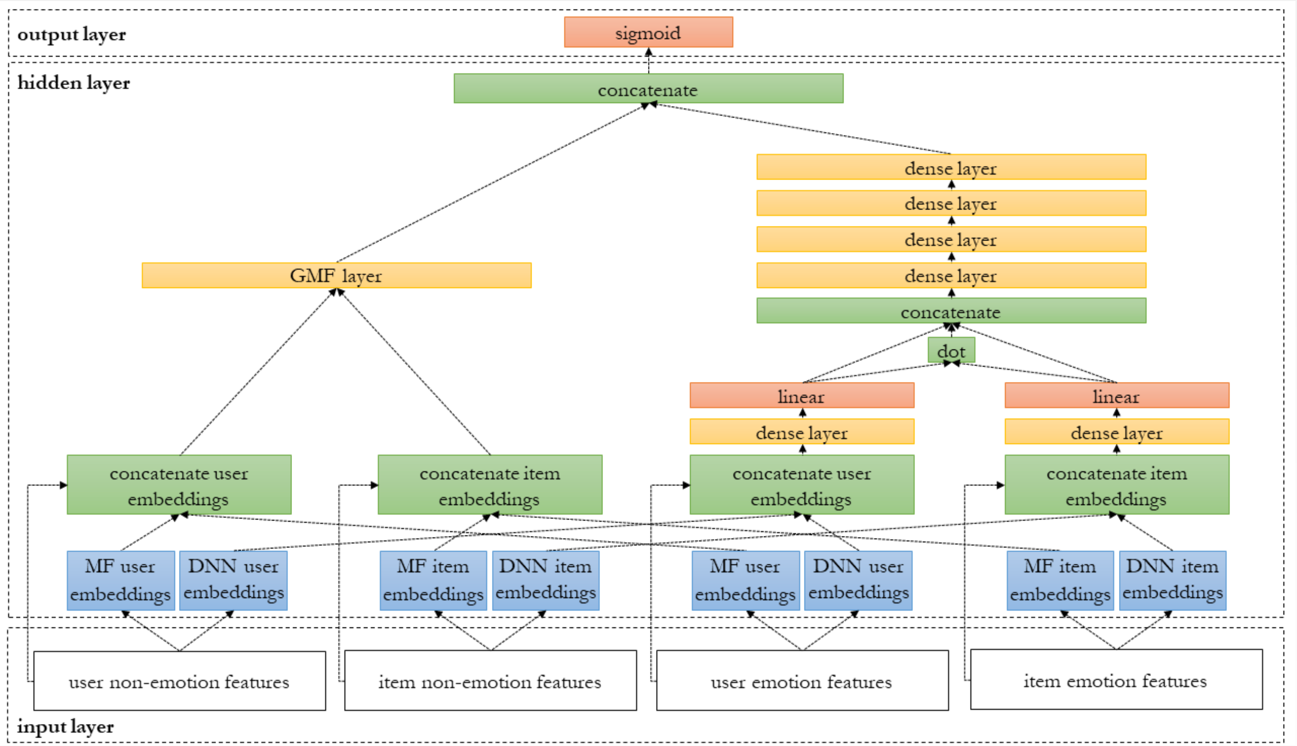
Proposed Model (EmoRec)

Model 2 (Deep Neural Network Model):



Proposed Model (EmoRec)

Model 3 (Deep Matrix Factorization Model):



Outline

- Motivation
- Objective
- Challenges
- Contributions
- Framework
- Proposed Model (EmoRec)
- **Experiments and Results**
- Conclusion and Future Works

Dataset

- **Globe and Mail** is one of the major newspapers in Canada. We use the data spanning from **January to July 2014** (a 6-month period) in our experiments where the data in the **first 4 months** were used for **training**, and the **last 2 months** for **testing**.

- **359,145** articles in total and **88,648** users in total, out of which **17,009** became subscribers during this period, and **71,639** were non-subscribers.

Experiments and Results

Results of EmoRec on News Dataset

Model	Non-Emo	All
Single Boost Model	70.19	70.86
Boost Blend	70.69	71.50
Deep MF	72.93	73.29
Single DNN Model	70.88	73.00
DNN Ensemble	73.62	74.30
Boost Blend + Deep MF	73.07	74.98
Boost Blend + DNN Ensemble	74.00	74.23
Deep MF + DNN Ensemble	74.61	75.10
EmoRec (Boost Blend + Deep MF + DNN Ensemble)	78.20	**80.30**

Experiments and Results

Comparison of EmoRec with State-of-the-art Baselines

Model	Non-Emo	All
Basic MF	69.10	71.23
FDEN and GBDT	72.02	73.28
Truncated SVD-based Feature Engineering	73.12	74.01
EmoRec	78.20	**80.30**

Experiments and Results

Effect of Individual Emotion Features on EmoRec

Emotion Features	(F-score)
ALL emotion features	80.30
- Sentiment features	78.15
- Mixed emotions	76.90
- Capitalized words	76.21
- Interjections	75.84
- Grammatical markers and extended words	75.23
- Ekman's emotion label	74.98
- Punctuation	75.17
- User emotions across categories	74.15
- User emotions across items	73.23
- Plutchik emotion scores	72.10

Evaluation Metrics

We use F-score to measure the predictive performance of a recommender system. For each user in the test data set, we use the original set of read articles in the test period as the ground truth, denoted as T_g. Assuming the set of recommended news articles for the user is T_r, precision, recall, and F-measure are defined as follows:

$$Precision = \frac{|T_g \cap T_r|}{|T_r|}, \quad Recall = \frac{|T_g \cap T_r|}{|T_g|}$$

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The F-score on a test data set is the average over all the users in the test data set.

Experiments and Results

Effect of Top Three Emotion Features

Model	No Emotion	Top Three Emotion
Basic MF	69.10	70.38
Boost Blend	70.69	71.00
FDEN and GBDT	72.02	72.77
Deep MF	72.93	73.01
Truncated SVD-based	73.12	73.60
DNN Ensemble	73.62	73.98

Outline

- Motivation
- Objective
- Challenges
- Contributions
- Framework
- Proposed Model (EmoRec)
- Experiments and Results
- **Conclusion and Future Works**

Conclusion

- We considered the problem of leveraging emotion features to improve recommendations.
- We derived a large number of **emotion features** that can be attributed to both **items** and **users** in news domain and can provide an emotional context.
- We proposed **EmoRec (**an ensemble model combining three models (Boost Blend + Deep MF + DNN Ensemble)), an emotion-aware recommendation model, which demonstrates the best predictive performance in news recommendation task.
- Our results indicate that emotion-aware recommendation models consistently outperform state-of-the-art **non-emotion-based** recommendation models.
- To the best of our knowledge, this is the first attempt to systematically and broadly evaluate the utility of a number of emotion features for the recommendation task.

Future Works

- We are working on how to predict the popularity of news headline using emotion features based on its content and other information.

- While the scope of our current study is limited to emotions extracted by textual information, there is evidence that emotions can be extracted through other means of communication, such as audio and video, or other cues.


Thank You
For Your Attention


It's just the beginning