ABSTRACT

News recommender systems are aimed to personalize users' experiences and help them discover relevant articles from a large and dynamic search space. Therefore, news domain is a challenging scenario for recommendations, due to its sparse user profiling, fast growing number of items, accelerated item's value decay, and users' preferences dynamic shift.

Deep Learning (DL) has achieved a great success in complex domains, such as computer vision, Natural Language Processing (NLP), machine translation, speech recognition, and reinforcement learning. Therefore, it became a mainstream approach in Recommender Systems research only since 2016.

The main objective of this research is the investigation, design, implementation, and evaluation of a Meta-Architecture for personalized news recommendations using deep neural networks.

As information about users' past interactions is scarce in such cold-start scenario, user context and session information are explicitly modeled, as well as past user sessions, when available. Users' past behaviors and item features are both considered in an hybrid session-aware recommendation approach. The recommendation task addressed in this work is next-item prediction for user sessions: what is the next most likely article a user might read in a session?

This paper presents the research methodology for this Doctoral research, the proposed Meta-Architecture and some preliminary results, as well as the next research challenges.

CCS CONCEPTS

• Information systems → Recommender systems; • Computing methodologies → Neural networks;

KEYWORDS

Recommender Systems; Deep Learning; News Recommendation; Session-Based Recommendation; Context-Based Recommendation; Meta-Architecture; Recurrent Neural Networks

1 INTRODUCTION

Recommender Systems (RS) have been increasingly popular in assisting users with their choices, thus enhancing their engagement and overall satisfaction with online services [32]. They are an important part of information and e-commerce systems, enabling users to filter through large information and product spaces.

Recommender systems have been researched and applied in online services from different domains, like music [10] [60] [64] (e.g., Spotify, Pandora, Last.fm), videos (e.g. YouTube [14]), people [2] (e.g., Facebook), jobs [4] (e.g., LinkedIn [33], Xing [44]), and research papers [62] [5] (e.g., Docear [6]), among others.

1.1 News Recommender Systems

Popular news portals, such as Google News [13], Yahoo! News [58], The New York Times [57], Washington Post [47] [9], among others, have gained increasing attention from a massive amount of online news readers.

Online news recommendations have also been addressed by researchers in the last years, either using Content-Based Filtering [36] [11] [49] [30] [45], Collaborative Filtering [13] [15], and Hybrid approaches [12] [39] [36] [48] [38] [37] [59] [18].

1.2 Deep Learning on Recommender Systems

Deep Learning (DL) [27] [28] [8] [7] is a hot area in machine learning communities. The uptake of deep learning by RS community was relatively slow, as the topic became popular only in 2016, with the first Deep Learning for Recommender Systems workshop at the ACM RecSys 2016 [25].

Early pioneer work applying used neural networks to RS was done in [52], where a two-layer Restricted Boltzmann Machine (RBM) slightly outperformed Matrix Factorization.

After a winter on RS research using neural networks, Deep Collaborative Filtering was addressed by [63] and [66] using denoising auto-encoders [61]. Deep neural networks have also been used to learn item features from unstructured data, like text [3], music [60] [64], and images [43] [21].

Recurrent Neural Networks (RNN) possess several properties that make them attractive for sequence modeling of user sessions. In particular, they are capable of incorporating input from past consumption events, allowing to derive a wide range of sequence-to-sequence mappings [16]. Recent work has explored the usage of RNNs for session-aware-based recommendations [24] [54].

According to [22], the main approaches using Deep Learning in Recommender Systems can be grouped as: (a) Item Embeddings and 2vec Models, (b) Deep Collaborative Filtering, (c) Deep Session-Based/Aware RS, and (d) Deep Feature Extraction from Heterogeneous Data [22]. This research focuses on approaches (c) and (d), due to some challenges of news recommendation and architecture requirements discussed in next sections.
2 RESEARCH

2.1 Problem
The news domain poses some challenges for Recommender Systems, summarized below:
- **Sparse user profiling** – the majority of readers are anonymous, and they actually read only few stories from the entire repository. This results in extreme levels of sparsity in the user-item matrix, and users usually have tracked very few information about past behaviour, if any [36] [38] [15];
- **Fast growing number of items** – hundreds of new stories are added daily in a news portal (e.g., over 300 in New York Times [57]). This intensifies the cold-start problem, as for fresh items you cannot count on lots of interactions before starting to recommend them [15]. For news aggregators, scalability problems may arise, as a high volume of news articles overload the web within limited time span [45];
- **Accelerated item’s value decay** – information value decays over time. This is specially true in the news domain, as most users are interested in fresh information. Thus, each item is expected to have a short shelf life [13]; and
- **Users preferences shift** – users interests on news are not as stable as in the entertainment domain. Some user interests shift over time, while other long-term interests remain stable [15]. User’s current interest in a session may be affected by his context (e.g., location, access time) [15] or by global context (e.g., breaking news or important events) [18].

2.2 Objective
The main objective of this research is to investigate, design, implement, and evaluate a deep learning meta-architecture for news recommendation, in order to improve the accuracy of recommendations provided by news portals, satisfying readers’ dynamic information needs in such a challenging recommendation scenario.

2.3 Research Scope and Methodology
The recommendation task addressed in this work is the next-item prediction for user sessions [55] (e.g., “what is the next most likely news article a user might read in a session?”). Models are trained using past user interactions on news portals to predict future article reads.

The Research Requirements are presented as follows:
- **RRQ1** - To investigate state-of-art methods for news recommendations;
- **RRQ2** - To investigate state-of-art Deep Learning architectures applied for recommender systems;
- **RRQ3** - To elaborate a Deep Learning meta-architecture for news recommendations;
- **RRQ4** - To implement and evaluate instantiations of the proposed meta-architecture; and
- **RRQ5** - To test stated hypotheses based on evaluation results.

2.4 Research Hypotheses
This research aims to evaluate the following hypotheses, related to the proposed meta-architecture for news recommendation:
- **HY1** - Is it possible to improve news recommendation accuracy by usage of Deep Learning architectures to learn articles’ feature representations directly from their textual content, when compared to feature extraction based on using statistical Natural Language Processing (NLP) techniques or other unsupervised learning methods (e.g. Doc2Vec)?
- **HY2** - Can news recommendations accuracy be improved, by modeling with RNNs users’ short-term (session) interests and context, when compared to other session-based recommenders?
- **HY3** - Can news recommendations accuracy be improved, by modeling with RNNs users’ long-term interests (past sessions), when compared to other session-aware recommenders?
- **HY4** - Can Deep Learning architectures automatically learn how to leverage news dynamic attributes (popularity and recency) and users contextual attributes (location, time, device) to improve recommendations accuracy, when compared to other context-aware recommenders?

3 CHAMELEON - A DEEP LEARNING META-ARCHITECTURE FOR NEWS RECOMMENDER SYSTEMS
The main contribution of this research is to propose the first Deep Learning meta-architecture for news recommendation.

For the purpose of this thesis, a meta-architecture is a reference architecture that collects together decisions relating to an architecture strategy. A meta-architecture might be instantiated as different architectures with similar characteristics that fulfill a common task, in this case, news recommendations.

3.1 Conceptual Model of News Relevance Factors
Many factors may influence the relevance of news article for users. In this research, a Conceptual Model of News Relevance Factors was conceptualized and is presented in Figure 1.

![Figure 1: The Conceptual Model of News Relevance Factors](image_url)

News articles have some static properties regarding its content, like title, text, topics, and mentioned entities (e.g., people, places) [36] [51]. The publisher reputation may also add trust or discredit to an article [41] [19].
News articles also have dynamic properties, that change over time, like popularity [12] [42] and recency [59] [19].

Global factors that might affect the article popularity are usually related to breaking events (e.g., natural disasters, or a real family member birth). There are also popular topics, that might be continuously interesting for users (e.g., sports) or may follow some seasonality (e.g., soccer during the World Cup, politics during a presidential election, etc.) [12] [19] [42].

The current user context, like location, device, and time is also important to define his short-term interests [45] [51], as they may differ during and off business hours, for example.

The HTTP referer URL brings some knowledge about the origin of the user request and may also be useful as a predictive contextual information, as shown in the Yahoo! experiments reported on [59].

Users’ long-term interests on news usually remain stable over time [15]. These interests may be specific personal preferences (e.g., chess playing) or influenced by popular global topics (e.g., tech).

The proposed Meta-Architecture leverages those factors to improve the quality of provided recommendations.

### 3.2 Requirements

To conceptualize the CHAMELEON Meta-Architecture, some requirements were first devised, based on the news recommender systems challenges described in Section 2.1 and also on the capabilities provided by Deep Learning. The CHAMELEON Meta-Architecture should be able:

- **RQ1** - to provide personalized news recommendations in extreme cold-start scenarios, as most news are fresh and most users cannot be identified;
- **RQ2** - to use Deep Learning to automatically learn news representations from textual content and news metadata, minimizing the need of manual feature engineering;
- **RQ3** - to leverage the user session information, as the sequence of interacted news may indicate user’s short-term preferences for session-based recommendations;
- **RQ4** - to leverage users’ past sessions information, when available, to model long-term interests for session-aware recommendations;
- **RQ5** - to leverage users’ contextual information as a rich data source, in such information scarcity about the user;
- **RQ6** - to model explicitly contextual news properties – popularity and recency – as those are important factors on news interest life cycle;
- **RQ7** - to support an increasing number of new items and users by incremental model retraining (online learning), without the need to retrain on the whole historical dataset; and
- **RQ8** - to provide a modular structure for news recommendation, allowing its modules to be instantiated by different and increasingly advanced neural network architectures.

### 3.3 The Proposed Meta-Architecture

The CHAMELEON Meta-Architecture presented in Figure 2 was designed based on the Conceptual Model of News Relevance Factors (Section 3.1) and also on requirements described in Section 3.2. It is based on neural networks, which support incremental online learning from mini-batches (RQ7).

The CHAMELEON Meta-Architecture was structured to support changes. It defines inputs, outputs, modules, sub-modules, and their interactions. Modules and sub-modules can be instantiated by different neural architectures, which might be evaluated and compared.

![Figure 2: CHAMELEON - A Deep Learning Meta-Architecture for News Recommender Systems](image)

The CHAMELEON is composed of two modules (RQ8) with independent life cycles for training and inference: the Article Content Representation (ACR) and the Next-Article Recommendation (NAR).

The ACR module is responsible for learning distributed representations (embeddings) for news’ contents. For scalability reasons, those embeddings are not directly trained for recommendation task, but for a side task of news metadata (e.g. category) classification.

The Next-Article Recommendation (NAR) module is responsible to provide news articles recommendations for active user sessions. Traditional CF and Hybrid RS trust in user and item identifications to learn similarities and provide recommendations. In the dynamic scenario of news recommendation (RQ1), the proposed hybrid meta-architecture does not count on item IDs when modeling users preferences on news. Instead, Article Content Embeddings, pre-trained in by ACR module, and stored in a repository, are used to represent the content of items.

User IDs are not directly used by the neural network as well – only users’ contextual information, interactions in the session, and past sessions information, when available.

The inputs for the NAR module are: (1) the pre-trained Article Content Embedding of the last interacted article; (2) the contextual properties of the article (popularity and recency); and (3) the user context (time, location, and device).

### 3.4 Comparisons with Related Work

To the best of the knowledge from this research so far, the only works presenting a deep learning architecture for news recommendation were [56] and [35]. The CHAMELEON is positioned in a higher level of abstraction, for being a Meta-Architecture.

One of the main inspirations for CHAMELEON was the GRU4Rec [24], the seminal work on the usage of Recurrent Neural Networks...
(RNN) on session-based recommendations, and subsequent work [26] [23]. Since then, a research line has emerged on the usage of RNN on session-based [65] [40] [55] and session-aware [16] [46] [50] recommendations.

The aforementioned works on the RNN session-based/aware architectures mostly trust on static user IDs and item IDs to provide recommendations, which does not meet the RQ7. The CHAMELEON uses articles content and context embeddings to represent items and also user context to represent users, dealing smoothly with the incoming stream of new articles and users interactions.

Other main inspirations came from the Multi-View Deep Neural Network (MV-DNN) [17], which adapted Deep Structured Semantic Model (DSSM) [29] for the recommendation task.

The MV-DNN maps users and items to a latent space, where the cosine similarity between users and their preferred items is maximized. That approach makes it possible to keep the neural network architecture static, rather than adding new units into the output layer for each new item (e.g., published article), as required in [24] (softmax loss function).

The MV-DNN was adapted for News recommendation by [56] Temporal DSSM (TDSSM) and [35] Recurrent Attention DSSM (RA-DSSM). Differently from CHAMELEON, TDSSM [56] did not model user sessions explicitly, and items and users representations are not directly learned from news content and user behaviours.

The RA-DSSM [35] ignores past user sessions information, whilst CHAMELEON provides session-aware recommendations. Furthermore, whilst RA-DSSM represents articles content by using Doc2Vec embeddings (unsupervised training), CHAMELEON trains news content embeddings to predict news metadata by using multitask supervised learning. Finally, the RA-DSSM does not use any contextual information about the user and articles, which may limit its accuracy in a extreme cold-start scenario like news RS.

4 INITIAL EXPERIMENTS

Based on the Research Methodology presented in Section 2.3, the student has completed RRQ1, RRQ2, RRQ3, and is currently working on RRQ4, by designing concrete architecture instantiations for ACR and NAR modules and implementing on TensorFlow [1].

For the first evaluation of ACR module, an architecture based on Convolutional Neural Networks (CNN) was proposed, as shown in Figure 3. CNNs have being widely used in Deep NLP research as a fast and powerful method for feature extraction from textual data.

The inputs for the ACR module are (1) article metadata attributes (e.g., publisher), and (2) article textual content, represented as a sequence of word embeddings (e.g., Word2Vec, GloVe). The Article Content Embeddings are trained to a side task – classifying the articles’ categories.

Architecture instantiations of the NAR module are also being designed and implemented, based on RNNs, and evaluated by usage of Top-N accuracy metrics: Recall@N and NDCG@N [31].

The Research Hypotheses stated in Section 2.4 will be tested (RRQ5) by offline evaluation on CHAMELEON instantiations, trained on news portals datasets described in Table 1.

5 CONCLUSIONS

The main contribution of this research is to propose the first Deep Learning meta-architecture for news recommendation - the CHAMELEON. Its instantiations and implementations will be evaluated on real news portals datasets, in order to test stated hypotheses.

This Doctoral student intends to make this research code available open-source as soon as main results are obtained, to allow reproducibility and to foster advances in Deep Recommender Systems.

REFERENCES

[1] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016).

[2] Hernan Badenes, Mateo N Bengualid, Jilin Chen, Liang Gou, Eben Haber, Jalal Mahmud, Jeffrey W Nichols, Aditya Pal, Jerald Schoudt, Barton A Smith, et al. 2014. System U: automatically deriving personality traits from social media for people systems.

[3] The Adressa Dataset (Norwegian) [20]: http://reclab.idi.ntnu.no/dataset

[4] Russian news portal dataset [53]: http://life.ru

Table 1: News portal datasets used by this research

| Language | G1 | GE1 | CLE2 | Adressa | Life.ru |
|----------|----|-----|------|---------|--------|
| Period (months) | 5 | 5 | 1 | 3 | 1 |
| clicks (million) | 124 | 196 | 1.5 | 2.7 | 30 |
| articles (thous.) | 391 | 156 | 6 | 1 | 140 |
| users (thous.) | 1,030 | 957 | 600 | 15 | - |
| user context | Yes | Yes | Yes | Yes | No |
[50] Massimiliano Ruocco, Ole Steinar Lillestøl Skrede, and Helge Langseth. 2017. Inter-Session Modeling for Session-Based Recommendation. In Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems. ACM, 24–31.

[51] Alan Said, Jimmy Lin, Alejandro Bellogín, and Arjen de Vries. 2013. A month in the life of a production news recommender system. In Proceedings of the 2013 workshop on Living labs for information retrieval evaluation. ACM, 7–10.

[52] Ruslan Salakhutdinov, Andryy Mnih, and Geoffrey Hinton. 2007. Restricted Boltzmann machines for collaborative filtering. In Proceedings of the 24th international conference on Machine learning. ACM, 791–798.

[53] Pavel Shashkin and Nikolay Karpov. 2017. Learning to rank for personalized news recommendation. In Proceedings of the International Conference on Web Intelligence. ACM, 1069–1071.

[54] Ole Steinar Lillestøl Skrede. 2017. Inter-/Intra-session Recurrent Neural Network for Session-based Recommender Systems. Master's thesis. NTNU.

[55] Elena Smirnova and Flavian Vasile. 2017. Contextual Sequence Modeling for Recommendation with Recurrent Neural Networks. In Proceedings of the 11th ACM Conference on Recommender Systems.

[56] Yang Song, Ali Mamdouh Elkahky, and Xiaodong He. 2016. Multi-rate deep learning for temporal recommendation. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM, 909–912.

[57] A. Spangher. 2015. Building the Next New York Times Recommendation Engine. https://open.blogs.nytimes.com/2015/08/11/building-the-next-new-york-times-recommendation-engine/. (Aug 2015).

[58] Michele Trevisiol, Luca Maria Aiello, Rossano Schifanella, and Alejandro Jaimes. 2014. Cold-start news recommendation with domain-dependent browse graph. In Proceedings of the ACM Recommender System conference, RecSys, Vol. 14.

[59] Michele Trevisiol, Luca Maria Aiello, Rossano Schifanella, and Alejandro Jaimes. 2014. Cold-start news recommendation with domain-dependent browse graph. In Proceedings of the 8th ACM Conference on Recommender systems. ACM, 81–88.

[60] Aaron Van de Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In Advances in neural information processing systems. 2643–2651.

[61] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning. ACM, 1096–1103.

[62] Chong Wang and David M Blei. 2011. Collaborative topic modeling for recommending scientific articles. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 448–456.

[63] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 1235–1244.

[64] Xinxi Wang and Ye Wang. 2014. Improving content-based and hybrid music recommendation using deep learning. In Proceedings of the 22nd ACM international conference on Multimedia. ACM, 627–636.

[65] Xiaohua Wu, Junwei Wang, Jiantao Liu, and Wenyu Liu. 2016. Recurrent neural network based recommendation for time heterogeneous feedback. Knowledge-Based Systems 109 (2016), 90–103.

[66] Yao Wu, Christopher DuBois, Alice X Zheng, and Martin Ester. 2016. Collaborative denoising auto-encoders for top-n recommender systems. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. ACM, 151–162.

583