**Preference Inference with Machine Learning: A Longitudinal Analysis (2019-2023)- 25\***

A longitudinal analysis of research on preference inference using machine learning over the past 5 years (2019-2023). It covers key trends, prominent methods and provides examples of research papers.

**Key Trends:**

* **Shift from Explicit to Implicit Preference Elicitation:** Early work often relied on explicit feedback (ratings, rankings). Recent research emphasizes inferring preferences from implicit signals like clicks, dwell time, purchase history and even physiological data.
* **Deep Learning Dominance:** Deep learning models, particularly neural networks, have become the dominant approach, offering greater capacity to model complex, non-linear preference relationships.
* **Contextualization:** Incorporating contextual information (time, location, user demographics) has become crucial for accurate preference inference, leading to personalized recommendations and experiences.1
* **Reinforcement Learning for Interactive Preference Learning:** Reinforcement learning (RL) is increasingly used to actively learn user preferences through interactive dialogues or by observing user interactions with a system.2
* **Explainable Preference Inference:** Research is focusing on making preference inference models more transparent and explainable, addressing concerns about bias and building user trust.

**Methods and Research**

**1. Collaborative Filtering with Deep Learning:**

* **Concept:** Extends traditional collaborative filtering by using deep learning to model user-item interactions. It suggests content based on user behavior, eg, Netflix.
* **Example:** "Neural Collaborative Filtering" (He et al., 2017). Many subsequent papers have built upon this, using various neural architectures.3
* **Focus:** Improved accuracy and ability to handle sparse data.

**2. Content-Based Filtering with Deep Learning:**

* **Concept:** Uses deep learning to extract features from item content (text, images, etc.) and match them to users preferences.
* **Example:** Research on recommendation systems for specific domains like music or movies often utilizes content-based filtering with deep learning. Extracting of feature embeddings from the movies for example.
* **Focus:** Leveraging rich item metadata for personalized recommendations.

**3. Hybrid Approaches:**

* **Concept:** Combines collaborative and content-based filtering using deep learning.
* **Example:** Weighted hybrid, combining scores from both methods. For example, Netflix. Switching hybrids like AWS, which starts cold.
* **Focus:** Capturing both user-item interaction patterns and item characteristics.

**4. Preference Inference from Implicit Feedback:**

* **Concept:** Inferring preferences from user behavior like clicks, dwell time, or purchase history.
* **Example:** Research on click-through rate (CTR) prediction often falls into this category. Deep learning models are used to predict the probability of a user clicking on an item.6
* **Focus:** Learning from noisy and sparse implicit data.

**5. Reinforcement Learning for Preference Elicitation:**

* **Concept:** Using RL to actively learn user preferences through interactive dialogues or by observing user interactions.7
* **Example:** Conversational recommendation systems or interactive preference learning.
* **Focus:** Efficiently exploring the preference space and adapting to individual users.

**Longitudinal Analysis (2019-2023):**

* **2019-2020:** Continued exploration of deep learning for collaborative and content-based filtering. Growing interest in implicit feedback and contextualization.
* **2021-2022:** Increased focus on RL for interactive preference learning and explainable AI for recommendations. More research on personalized recommendations in specific domains.
* **2023:** Emerging trends include graph neural networks for preference inference, federated learning for privacy-preserving preference learning, and the use of large language models (LLMs) for understanding user preferences from natural language input.

**Challenges and Future Directions:**

* **Data Sparsity:** Dealing with limited user interaction data.
* **Cold Start Problem:** Recommending items to new users or recommending new items.8
* **Bias and Fairness:** Ensuring that preference inference models are fair and unbiased.
* **Privacy:** Protecting user privacy when collecting and using preference data.
* **Scalability:** Developing scalable methods for large datasets and user bases.

**1. Preference Learning from Implicit Feedback:**

* **Challenge:** Implicit feedback (clicks, dwell time, etc.) is abundant but noisy and doesn't directly indicate preference strength.1
* **Recent Advances:**
  + **Deep Learning for Implicit Feedback Recommendation (WWW 2017)**: This paper by He et al. (not the same He as in Neural Collaborative Filtering) explores using deep neural networks to learn user preferences from implicit feedback data. It demonstrates the effectiveness of deep learning in capturing complex user-item relationships
  + **RecSys Challenge 2019: Real-Time Personalized Recommendations with Implicit Feedback (RecSys 2019)**: This challenge focused on real-time recommendation systems using implicit feedback, pushing the boundaries of efficient preference inference from large-scale data.
* **Key Ideas:**
  + **Pointwise vs. Pairwise Learning:** Pointwise methods predict the absolute preference score for an item, while pairwise methods learn to rank items based on relative preferences.2
  + **Negative Sampling:** Since only positive interactions are observed, negative samples are crucial to train models effectively.

**2. Contextualized Preference Inference:**

* **Challenge:** User preferences are not static; they change based on context (time, location, device, etc.).
* **Recent Advances:**
  + **Session-based Recommendations with Recurrent Neural Networks (ICLR 2016)**: This work by Hidasi et al.3 shows how recurrent neural networks (RNNs) can capture sequential patterns in user behavior and improve session-based recommendations.
  + **"Attention-aware Personalized Recommendation with Contextual Information" (KDD 2018)**: This paper by Chen et al. introduces an attention mechanism to learn the importance of different contextual features for personalized recommendations.
* **Key Ideas:**
  + **Contextual Features:** Incorporating features like time, location, demographics, and device information.4
  + **Attention Mechanisms:** Learning to focus on relevant contextual information for each user.

**3. Reinforcement Learning for Interactive Preference Elicitation:**

* **Challenge:** Actively learning user preferences through interactions is more efficient than passively observing behavior.
* **Recent Advances:**
  + **Deep Reinforcement Learning for Interactive Recommendation (RecSys 2018)**: This paper by Zhao et al. proposes using deep reinforcement learning to train an agent that can interact with users and learn their preferences through trial and error.
  + **Conversational Recommendation Systems (SIGIR 2020)**: This survey paper by Zhang et al. provides a comprehensive overview of conversational recommendation systems, which use natural language to interact with users and elicit their preferences.
* **Key Ideas:**
  + **Exploration-Exploitation Dilemma:** Balancing exploration of the preference space with exploitation of known preferences.
  + **Reward Function Design:** Defining a reward function that accurately reflects user satisfaction.

**4. Explainable Preference Inference( Explore or not?)**

* **Challenge:** Understanding why a model makes a particular recommendation is crucial for building trust and addressing bias.
* **Recent Advances:**
  + **Attention is not Explanation (NAACL 2019)**: This paper by Jain and Wallace challenges the common assumption that attention weights provide meaningful explanations.
  + **Explainable Recommendation: A Survey (Foundations and Trends in Information Retrieval 2020)**: This survey paper by Zhang and Chen provides a comprehensive overview of explainable recommendation techniques.7
* **Key Ideas:**
  + **Post-hoc Explanations:** Generating explanations after the model has made a prediction.
  + **Ante-hoc Explanations:** Building explainability directly into the model architecture.

**Sources**

* **Google Scholar**
* **Journals**
  + NeurIPS (Advances in Neural Information Processing Systems)
  + ICML (International Conference on Machine Learning)8
  + ICLR (International Conference on Learning Representations)
  + KDD (Knowledge Discovery and Data9 Mining)
  + RecSys (ACM Conference on Recommender Systems)
  + SIGIR (ACM Conference on Research and Development in Information Retrieval)
* **Code Repositories**

**Enhanced Architecture**

Enhanced preference inference pipeline with an architecture diagram. The focus is to incorporate deep learning for feature engineering, balancing improvement and complexity.

* + 1. User Data
    2. Hierarchical Clustering
    3. Deep learning
    4. Random Forest
    5. Recommendations

**Explanation**

1. **User Data:** Represents the input data, which includes user-item interaction data (e.g., which items users have interacted with, ratings, etc.) and contextual information (e.g., time, location, demographics).
2. **Hierarchical Clustering:** This stage uses hierarchical clustering with the Jaccard distance (or another appropriate metric) to group users based on their interaction patterns. The output is a hierarchical structure of user clusters.
3. **Deep Learning Feature Extractor (Autoencoder):** For each cluster, a deep learning model (autoencoder) is trained on the interaction data of the users within that cluster. The autoencoder learns a compressed representation of the cluster, which is then used as a feature vector. Eg. a simple feedforward network, or a recurrent network if sequence matters.
4. **Random Forest Classifier:** The feature vectors generated by the deep learning models, along with the contextual features, are fed into a random forest classifier. The random forest learns to predict user preferences or make recommendations based on these features.
5. **Preferences/Recommendations:** The output of the random forest is a set of inferred user preferences or personalized recommendations.