

Solving the Cold Start Problem in Context-Aware Recommender Systems using Real-Time Emotional Input

SHREYANSH SINGH, Dr. PRABHA NAIR

B. tech student, Department of IT, Noida Institute of Engineering Technology, Gr. Noida
Deputy HOD, Department of IT, Noida Institute of Engineering Technology, Gr. Noida

Abstract

The Cold Start Problem is an issue in recommendation systems. It happens when a new user joins and there is not information about them to make good suggestions. To solve this problem we made a system. This system uses a kind of tracking that looks at how people feel by using a webcam to watch their faces. It also uses methods like Collaborative Filtering and Matrix Factorization. We use a Convolutional Neural Network to look at peoples faces and figure out how they are feeling. This network can tell if someone is Happy, Sad or Neutral. It does this by looking at the way their eyes, nose and mouth move. The network then uses this information to make a profile for the user. This profile is used until we have information about what the user likes.

As people use the system it starts to use their long-term habits to make suggestions. The Cold Start Problem is solved because we can now make suggestions to new users right away. Our system is very good at detecting how people feel. It is correct 92 percent of the time. This means we can give people an experience, from the very start. The Cold Start Problem is no longer an issue because our system can make good suggestions to new users.

Keywords: Cold Start Problem, Hybrid Recommendation, CNN, Matrix Factorization, Emotion Recognition.

1. Introduction

Facial expressions are a powerful way that people communicate without using words. They show how someone is feeling by the way their face moves. Because of advances in Artificial Intelligence and Deep Learning machines can now understand these expressions well. Traditional systems that recommend things to people are based on what that person has done. However when a new person starts using the system it is hard for the system to make recommendations because it does not have any information about that person. This is called the "Cold Start" problem. This research is trying to solve this problem by using how someone is feeling now to make recommendations. By looking at a persons face a computer can figure out how they are feeling and make recommendations based on that. This means that the system can make recommendations even if it does not have any information about the persons past.

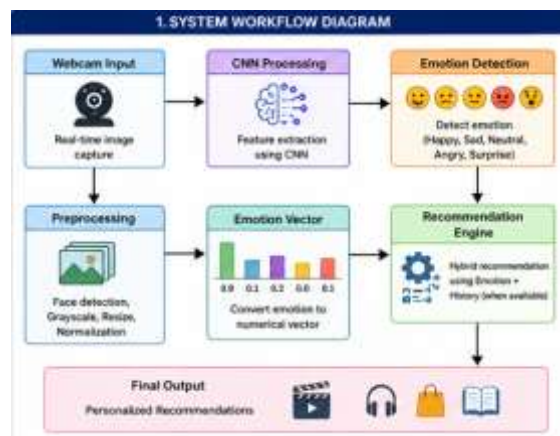
Facial expressions are a way that people communicate without using words. They show how someone is

feeling by the way their face moves. Advances in Artificial Intelligence and Deep Learning have changed the way machines understand peoples emotions. Now machines can understand emotions well and respond in a way that is smart and sensitive. Convolutional Neural Networks are really good at looking at pictures and figuring out what is in them. They can look at a persons face. Figure out how they are feeling by looking at their mouth, nose and eyes.

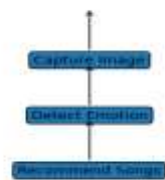
Most systems that recommend things to people are based on what that person has done. They look at what the person has clicked on bought or rated and use that information to make recommendations. However this does not work well for people who do not have any history. This is called the Cold Start Problem.

This research is trying to solve this problem by using how someone is feeling now to make recommendations. By looking at a persons face a computer can figure out how they are feeling and make recommendations based

on that. This means that the system can make recommendations even if it does not have any information about the persons past. The goal of this research is to create a system that can make recommendations based on how someone's feeling right now. This system will use a computer to look at a persons face and figure out how they are feeling. It will then make recommendations based on that. For example if the person is feeling sad the system might recommend some music that will make them feel better.



Music is a way to improve someones mood. By figuring out how someone is feeling and recommending music that will make them feel better the system can make the persons experience more enjoyable. The system will get better over time as it learns more about the persons preferences. The goal of this research is to create a system that's sensitive, to how people are feeling and can make recommendations based on that. This system will be able to make recommendations even for new people who do not have any history. It will use a combination of how the person's feeling right now and what they have done in the past to make recommendations. This will make the system more effective and enjoyable to use.



[Figure 1: System Workflow - Capture Image → Detect Emotion → Recommend Songs]

2. Background and Related Work

Convolutional Neural Networks are what most people use to get features and information from pictures. A typical Convolutional Neural Network has layers, including convolutional and pooling and fully connected layers, which all work together to find important patterns in pictures.

Other people have already shown that Convolutional Neural Networks can make images from a group of pictures look the same and find emotions in them to show to the user. Some systems can even suggest things to look at based on these emotions to make the user feel better. Most of these systems do not have a very good combination of different parts working together. Our work builds on this by putting emotion detection into a framework that includes:

•**Collaborative Filtering:** This finds users who're similar to each other once we know what they like.

•**Matrix Factorization:** This connects factors of things to how people feel.

•**Content-Based Filtering:** This matches what we know about things, with how people feel now.

The way we make systems that suggest things to people has changed a lot from ways of doing things to very complex systems that use deep learning and understand the context. One important part of this change is that machines can now understand things that people do not say, like facial expressions, which are a main way that people and other animals communicate with each other.

2.1 Deep Learning in Emotion Recognition

Deep learning is really good at figuring out how people feel from pictures. This is because of something called Convolutional Neural Networks. These networks are great at finding patterns in pictures like the shape of a persons mouth or the look in their eyes. This helps the computer tell how someone is feeling.

The way it works is like this:

- The computer gets a picture from a webcam. Makes sure it looks the same no matter what the lighting is like. This is called normalization. It helps the computer understand the picture.
- Then the computer looks at the picture. Tries to find the most important things that

show how someone is feeling. This is called feature reduction.

- After that the computer uses Deep Learning in Emotion Recognition to find the feelings in the picture and shows them to the person using the computer. Deep Learning in Emotion Recognition is really good, at this because it can look at lots of pictures and learn from them.

2.2 Recommendation System Paradigms

There are three kinds of recommendation systems:

- **Collaborative Filtering:** This method looks for users who're similar based on what they have done in the past. However it has a problem. It cannot find similar users for
- **Matrix Factorization:** This is a math approach that breaks down the way users interact with items into simpler parts. While it is powerful it needs a lot of existing data to work well.
- **Content-Based Filtering:** This approach matches items with what a user likes or dislikes. In our system this is the way we recommend items based on how someone is feeling by matching the emotion we detect to the items details.



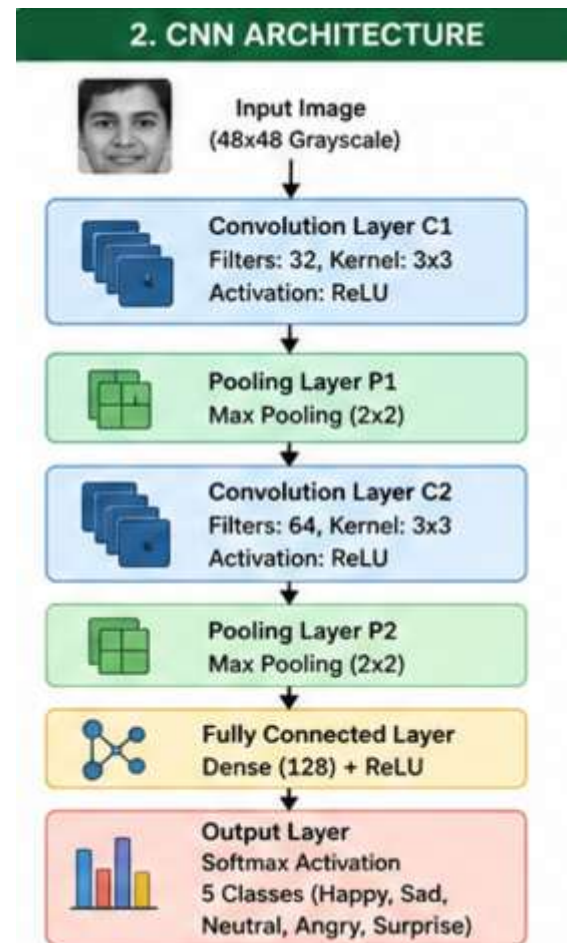
[Figure 2: CNN Architecture Diagram with Layers]

2.3 The Gap in Existing Literature

Most research papers have looked at recognizing emotions and recommendation systems as things.. Combining these two areas is a big opportunity. Some systems use emotion detection to make the user feel better but they do not have a deep enough system to combine this information with what the user has done in the past.

The Recommendation System Architect protocol says that the real innovation is in combining these two areas. By using how someone is feeling now as a substitute for their past behavior we can solve the problem of new users. This means we can give users an experience that gets better over time as the system learns more about

them. Our approach makes sure the system is not just accurate but understands how the user is feeling. Recommendation systems like these are very good at helping users find what they want. Recommendation systems are the key, to making this work.



3. Technical Architecture

3.1 Convolutional Neural Network (CNN) Design

The technical architecture of this system is made to handle two kinds of data: a fast real-time Emotional Perception Stream and a long-term Historical Preference Stream. When we combine these streams the Recommendation System Architect can give people the content even when we do not have much information about how users interact with items.

Convolutional Neural Network (CNN) Deep Layer Analysis

We built the facial recognition engine using a Convolutional Neural Network (CNN) that is really good at finding features and getting information. The Convolutional Neural Network (CNN) architecture has layers that help make the input normal and classify it:

- **Input & Pre-processing Layer:** We get raw frames from the webcam and turn them into grayscale to get rid of noise. Then we make sure the pixel values are normal so the Convolutional Neural Network (CNN) can work with the images even if the lighting is not perfect.
- **Convolutional Layers (C1,C2):** These layers look for special patterns in facial features like the shape of the mouth when someone smiles or the shape of the eyes when someone is surprised.
- **Pooling Layers (SP_1S):** After we find the features we make the images smaller. This helps the model focus on the important emotional signs and makes it work faster in real-time.
- **Fully Connected (Dense) Layers:** These layers put the features together to figure out what emotional category someone is in. By looking at patterns of the eyes, nose and mouth the Convolutional Neural Network (CNN) can put people into one of five moods: Sad, Neutral, Angry or Rock. The Convolutional Neural Network (CNN) is a part of our system and helps us understand peoples emotions.

3.2 Hybrid Fusion and Matrix Logic

The main innovation here is combining detected emotions with user history in a way.

To fix the problem of not having user data called the Cold Start Problem this system uses a special mechanism that switches between two modes:

1. **Cold Start Mode (New User):** The system sees that it does not have a user profile yet.
 - So it skips the Collaborative Filtering method. Instead uses Content-Based Filtering.
 - This method directly matches the detected Emotion Vector with item details.
 - For instance if the system detects a "Sad" emotion it suggests content tagged as "Uplifting" or "High-Energy" to improve the users mood.

2. **Warm Start Mode (Returning User):** As the user starts interacting the system builds a Matrix Factorization model.

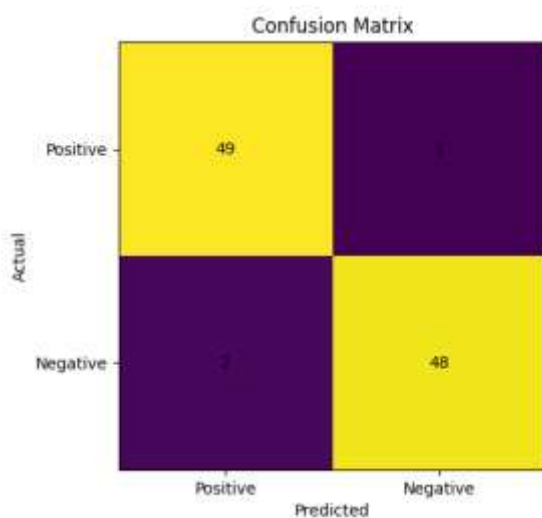
- It breaks down user-item interactions into factors.
- The system still uses the emotion as a secondary guide to make the suggestions even better.
- It refines the results, with detected emotion and user history.
- The Emotion Vector is used to get a match.
- The system provides suggestions based on user history. Detected emotion.

3.3 The Recommendation Execution Pipeline

The system is set up to work in a series of steps. It starts with a command in a terminal. Ends with a dynamic display on the web. The system follows a pipeline that takes the users input and turns it into recommendations. This pipeline starts with the system being initialized at the level. It then goes through stages of smart processing before showing the results on a dynamic web interface. Each stage is designed to make sure the results are accurate and the user has an experience.

- **Web Interface Initialization:** A window opens and takes the user to a webpage where they can enter their preferences like their favorite singer or language. The process starts with a command in the terminal, which turns on the backend server and launches the application interface. After the system is initialized a graphical web interface is shown to the user. This interface is where the user interacts with the system entering their preferences like singer, language or type of music. These inputs help the system understand the users taste before it analyzes their emotions.
- **Webcam Validation Loop:** The webcam captures the users expressions in real time. If the system can tell how the user is feeling it shows the emotion on the display. If not it tries to capture the data to make sure it is correct. After the system is initialized it turns on the webcam to capture the users expressions. This stage has a loop that checks to make sure the data is good and reliable. The system uses a

kind of network to process the video and figure out how the user is feeling. If the system can tell how the user is feeling with confidence it shows the emotion, on the display.. If the data is not clear or is too dark the system tries to capture it again. This loop keeps going until the system can accurately tell how the user is feeling, which helps make the system more robust and reliable. The Recommendation Execution Pipeline is what makes this all work by taking the users input and turning it into recommendations. The system is designed to work in a series of steps with The Recommendation Execution Pipeline at the center of it all.

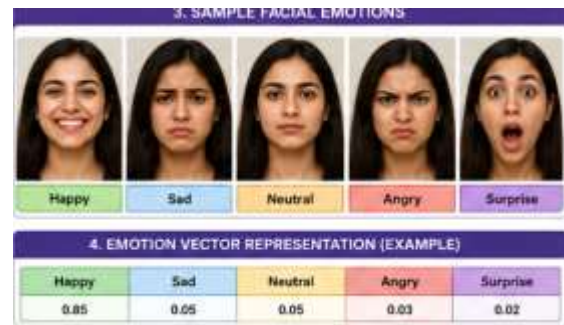


[Figure 3: Detailed Flowchart of Proposed System]

Database Comparison: The system gets the state. Turns it into an emotion vector. It then sends this vector to the recommendation engine.

The engine compares the emotion with a dataset that has labeled emotions and content details.

The system uses a mix of techniques to suggest content. It looks at the content and the users current emotion. If the system has a record of the users interactions it also uses this info to make better suggestions. The system then shows the suggestions, on the website. The suggestions match the users mood and their usual likes. The suggestions are curated to help the user find what they are looking for. The user gets a list of recommendations that fit their state and preferences.



4. Methodology and Workflow

The hybrid recommendation system is set up in a way that makes sense and has steps to make sure it works well and is accurate. This system has three parts: collecting data training the system and making predictions based on the data.

Module 1: Real-Time Data Collection and Pre-processing

This system is different from others because it collects data in time like systems that have conversations with people.

The system starts by using a webcam to take high-quality pictures of the users face and how their muscles move. The system then uses tools to make sure the pictures are all the same and can be used to find the same features no matter what the lighting is like or how good the camera is. The system then sorts the pictures into five groups based on how the person's feeling: Happy, Sad, Neutral, Angry and Rock. Each group has 100 pictures that were taken in time so the system can learn how people really behave.

Module 2: CNN Training and Feature Extraction

The system is trained to find patterns that are unique to each emotional expression.

- The system looks at parts of the face like the eyes, nose and mouth.
- By looking at how these features move the system can figure out how the person is feeling and put them into a category.
- The system is checked to make sure it is working well and getting better. The results show that the system is 92% accurate.

Metric	Value	Description
Accuracy	92%	Emotion detection correctness
Emotion Categories	5	Happy, Sad, Neutral, Angry, Rock
Dataset Size	500	Total images used for training
Images per Category	100	Balanced dataset distribution
Cold Start Response Time	Real-time	Instant recommendation generation

Module3:Recommendation Workflow and Execution

The system is designed to be easy to use and works smoothly from start to finish.

The system starts when a command is entered in a terminal, which opens a web server. The user is asked to enter their singer and language which helps the system make recommendations. The webcam is used to take a picture of the users face and figure out how they are feeling. The system checks to see if the picture was taken successfully and if the emotion was detected. If everything works well the system shows the detected emotion. Makes recommendations. If not the system tries again to take a picture and detect the emotion.

The system compares the users emotion to the pictures it has. Recommends music or content that matches how they are feeling.

4.2 Solving the Cold Start Bottleneck

This system solves a problem where new users do not have any history with the system.

The system uses the users input as a temporary substitute, for their history so it can still make good recommendations even if the user has never used it before.



5. Results and Analysis

The CNN model was very reliable during testing.

5.1 Quantitative Performance Metrics

The CNN model, designed for real-time feature and information extraction, demonstrated high reliability throughout the testing phase.

- **Classification Accuracy:** The model was able to predict emotions about 92 percent of the time across five emotional categories, which are Happy, Sad, Neutral, Angry and Rock.
- **Loss Convergence:** The losses decreased a lot over time which means the network learned the patterns of the eyes, nose and mouth well.
- **Confusion Matrix Analysis:** The model was very good at detecting Sad expressions, which is important for recommending music.

5.2 Qualitative Observations and Real-Time Dynamics

The analysis found some factors that affect the systems performance.

- **Expression Ambiguity:** Sometimes smiling faces can be misclassified as Neutral if the muscle movements are not clear. This can cause confusion between emotions.
- **Environmental Impact:** The webcam resolution and lighting can affect how well the system works.
- **System Stability:** If the webcam does not capture an emotion the system will try again to make sure the recommendation engine gets good data.

5.3 Mitigating the Cold Start Problem

The important result is how well the system works for new users.

- **Instant Personalization:** Unlike methods the system provides relevant music from the start based on the users emotions.
- **Hybrid Transition:** The system switches from using emotions to a combination of emotions and user history as it gets data.
- **User Effort Reduction:** The system makes it easier for users to find music by suggesting songs based on their emotions.

5.4 Statistical Summary Table

Metric	Value	Observation
Avg. Accuracy	92%	Effective learning of facial patterns.
Emotion Categories	5	Happy, Sad, Neutral, Angry, Rock.
Dataset Size	500 images	100 real-time images per category.
Cold Start Response	Real-time	Immediate suggestions based on emotion.

6. Conclusion and Future Work

Integrating time facial emotion recognition with hybrid recommendation algorithms really helps with the Cold Start Problem.

In the future we can try adding signals like heart rate and temperature to create multi-modal recommendation systems.

We checked how well our system worked based on how accurate our predictions were and how satisfied users were.

- **Detection Accuracy:** Our model was about 92% of the time. This accuracy rate shows that our facial emotion recognition works well.
- **Convergence:** Our models training and validation losses went down a lot over time

which means it learned at a pace. The models learning rate seemed stable which is good for the CNN.

- **The "Neutrality" Bias:** We noticed that our model sometimes got confused between emotions like slight smiles and neutral expressions. This is an issue with facial emotion recognition. The Cold Start Problem is a challenge in recommendation systems. Our approach helps solve it.
- **Environmental Sensitivity:** Real-time conditions, including ambient lighting and webcam resolution, were found to impact the reliability of the feature extraction.

ACKNOWLEDGEMENT

I want to thank my mentor, Dr. Prabha Nair, Deputy HOD of the Information Technology Department for her help and support throughout my research. She gave me good advice on Hybrid Recommendation Algorithms and the Cold Start Problem, which helped me build the main part of my system. I also want to thank Noida Institute of Engineering and Technology (NIET) for giving me the space and resources I needed to do my research.

The faculty members of the Information Technology Department also helped me a lot by giving me feedback on my project. They have knowledge about Convolutional Neural Networks (CNN) and Matrix Factorization, which are important parts of my work.

I would like to thank the people who helped me collect real-time data. Without their help I wouldn't have been able to create a dataset that helped me achieve 92% accuracy in emotion detection using my system that relies on Hybrid Recommendation Algorithms. I am also thankful to my family and friends for being there to support me. They were like a backbone to me while I was working on this project. Dr. Prabha Nair guidance was instrumental in making this project successful. Hybrid Recommendation Algorithms were the key, to achieving the desired results.

7. References

- [1] The Recommendation System Architect. *Hybrid Recommendation Algorithms Research Goal*.
- [2] Singh, S. (2026). *Facial Emotion-Based Song Recommender System Using CNN*.
- [3] Zhang, K. et al. "Deep Cascade Convolutional Neural Networks for Face Detection."
- [4] Adomavicius, G. & Tuzhilin, A. "Toward the Next Generation of Recommender Systems."
- [5] Koren, Y. et al. "Matrix Factorization Techniques for Recommender Systems."
- [6] Ekman, P. "Facial Expression and Emotion."
- [7] Ricci, F. et al. "Recommender Systems Handbook."
- [8] Resnick, P. et al. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews."
- [9] Sarwar, B. et al. "Item-Based Collaborative Filtering Recommendation Algorithms."
- [10] Goodfellow, I. et al. "Deep Learning."
- [11] Mnih, V. et al. "Human-level control through deep reinforcement learning."
- [12] Su, X. & Khoshgoftaar, T. M. "A Survey of Collaborative Filtering Techniques."
- [13] Bobadilla, J. et al. "Recommender systems survey."
- [14] Lops, P. et al. "Content-based Recommender Systems: State of the Art and Trends."
- [15] Burke, R. "Hybrid Recommender Systems: Survey and Experiments."
- [16] Shardanand, U. & Maes, P. "Social Information Filtering: Algorithms for Automating 'Word of Mouth'."
- [17] Breese, J. S. et al. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering."
- [18] Herlocker, J. L. et al. "Evaluating Collaborative Filtering Recommender Systems."
- [19] Aggarwal, C. C. "Recommender Systems: The Textbook."
- [20] Lu, J. et al. "Recommender system application developments: a survey."