ECONOMIC ASPECTS OF ANALYSIS RELATED TO THE DEVELOPMENT OF A MULTIMODAL EMOTION RECOGNITION

Aharon RUSHANYAN
Ph.D. applicant at Armenian State Politechnical University

Key words: economic behavior, market sentiment analysis, multimodal analysis, convolutional neural networks, emotion-driven market trends, emotion recognition

Introduction

Emotion recognition has gained importance in economics as it helps decipher consumer behavior, employee emotions and has become a valuable tool for analyzing economic trends. Multimodal systems that integrate visual and audio cues have become powerful tools for assessing these emotional dynamics [Ekman & Friesen, 1978, 527].

In the context of economics, understanding emotions can provide important information about the marketplace, consumer decision-making, and employee well-being. By combining audio and visual cues, multimodal emotion recognition systems can more accurately assess emotional states, which is particularly important in the context of such fields of economics as marketing, financial forecasting and human resource management [Baltrušaitis et al. 2019, 20]. This article discusses multimodal emotion recognition based on convolutional neural networks (CNNs) [Ayeni 2002, 12]. The potential of recognition systems in the field of economics, as well as the methodology, practical implementation and scientific novelty of using these systems for economic analysis are presented.

Methodology

The first step in building a multimodal emotion recognition system is to collect relevant data.

Data collection

- Consumer data: product testing records or advertising reviews.
- Market sentiment data: public reactions during major economic events (e.g., political announcements, stock market news).
- Employee surveys: video and audio collected during interviews or work environment assessments.

Feature Extraction

- Audio MFCC (Mel-Frequency Cepstral Coefficients) feature allows capturing emotional tone [Simonyan & Zisserman, 2015, 14].
- Visualization: extracting frames from video to detect facial expressions expressing emotions such as joy, stress or interest [Viola & Jones, 2001, 20].

Model-based architecture

- Audio CNN, trained to analyze MFCC features and detect patterns of audio tones.

- Video CNN trained to categorize emotions such as happiness, surprise or anger based on facial expressions.[[Viola & Jones, 2001, 20].
- The connection layer combines the output of audio and video CNNs for final emotion classification [Zhang et al. 2018, 15].

System operation

- Pre-processing of economic datasets to extract visual features.
- Training CNNs on labeled data with economic context (e.g., customer feedback annotations).
- Applying the trained model to analyze new data for real-time or retrospective analysis.

Literature review

Emotion recognition has attracted much attention in various fields including psychology, artificial intelligence, and economics. Traditional approaches such as text-based sentiment analysis have limited ability to capture non-verbal cues [Pang & Lee, 2008, 135]. The research on "Facial Action Coding Systems" developed by Ekman and Friesen emphasizes the importance of facial expressions in conveying emotions [Ekman & Friesen, 1978, 527]. Similarly, audio-based emotion analysis highlights how tone and pitch reflect sentiment [Schuller et al., 2013, 12].

In economic applications, sentiment recognition has been used to assess consumer reactions to advertisements, to analyze the workforce, and to assess investor sentiment during financial events. Deep learning, especially convolutional neural networks (CNNs), recent developments in this area have improved the accuracy of multimodal systems, making them ideal for integrating audio and video cues.

Scientific novelty

Integrating multimodal emotion recognition into economic analysis is an innovative application of deep learning. While traditional economic models rely on numerical and textual data, the addition of audio and visual cues enriches the understanding of human behavior. This approach creates a link between psychology and economics, emphasizing the role of emotional drivers in economic activity [Loewenstein & Lerner, 2003, 23]. Increases the accuracy of sentiment analysis with additional methods, as well as provides the ability to track sentiment in real time, allowing dynamic responses to market trends.

Research and Analysis

To demonstrate the real use of multimodal emotion recognition, program that analyzes consumer sentiment during product feedback.

Step 1: Data Preprocessing

Audio augmentation

Additional enhancements are needed to the audio data processing to improve the reliability and robustness of the model. Adding audio noise to generalize the data across different audio conditions to increase the sensitivity of the model to audio frequencies.

```
def augment_audio(y, sr):
    # Add noise
    noise = np.random.normal(0, 0.005, y.shape)
    y_noise = y + noise
    # Time stretching
    y_stretch = librosa.effects.time_stretch(y, rate=0.9)
    # Pitch shift
    y_pitch = librosa.effects.pitch_shift(y, sr, n_steps=2)
    return [y_noise, y_stretch, y_pitch]

# Example usage
y, sr = librosa.load('path/to/audio.wav', sr=None)
augmented_audio = augment_audio(y, sr)
print(f"Generated {len(augmented_audio)} augmented samples")
```

Figure 1. Audio augmentation

This step ensures that the voice CNN is exposed to a variety of conditions, which improves its generalization ability and enhances its adaptability in realistic scenarios.

Video data augmentation

The following transformations are applied to video data: image flipping and rotation to make the model more robust to changes in camera angle; lighting adjustments to adapt the model's capabilities to different lighting conditions.

These additions increase the diversity and robustness of video CNNs.

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator

def augment_frames(frames):
    datagen = ImageDataGenerator(
        horizontal_flip=True,
        rotation_range=20,
        brightness_range=[0.8, 1.2]
    )
    augmented_frames = [datagen.random_transform(frame) for frame in frames]
    return np.array(augmented_frames)

# Example usage
augmented_frames = augment_frames(frames)
print(f"Original frames: {frames.shape}, Augmented frames: {augmented_frames.shape}")
```

Figure 2 : Video augmentation

Step 2: Merge at attribute level

Instead of combining data from audio and video CNNs in the final decision step, we can combine them at the feature level.

Advantages: The model is simultaneously trained on audio and video modules to create deep insights. Improves the efficiency of detailed tone discrimination. and visual cues.

```
from tensorflow.keras.layers import Flatten, Dense, concatenate

def build_fusion_model(audio_model, visual_model):
    # Flatten CNN outputs
    flat_audio = Flatten()(audio_model.output)
    flat_visual = Flatten()(visual_model.output)
    # Combine feature vectors
    combined_features = concatenate([flat_audio, flat_visual])
    # Add dense layers for joint learning
    x = Dense(128, activation='relu')(combined_features)
    x = Dense(5, activation='softmax')(x)  # Final classification layer
    # Create the model
    fusion_model = Model(inputs=[audio_model.input, visual_model.input], outputs=x)
    return fusion_model

fusion_model = build_fusion_model(audio_cnn, visual_cnn)
fusion_model.summary()
```

Figure 3: Merge at attribute level

Step 3: Multimodal sentiment analysis in real time

During a customer feedback session, the system processes audio and video streams in real time. Simultaneous processing of audio and video data enables real-time sentiment analysis.

```
import threading

def process_audio_stream(audio_queue, model):
    while True:
        if not audio_queue.empty():
            audio_data = audio_queue.get()
            features = extract_audio_features(audio_data)
            emotion = model.predict(features.reshape(1, 40, 1, 1))
            print(f"Audio Emotion: {np.argmax(emotion)}")

def process_video_stream(video_queue, model):
    while True:
        if not video_queue.empty():
            frame = video_queue.get()
            frame = cv2.resize(frame, (64, 64))
            emotion = model.predict(frame.reshape(1, 64, 64, 3))
            print(f"Visual Emotion: {np.argmax(emotion)}")

# Using threads for parallel processing
audio_thread = threading.Thread(target=process_audio_stream, args=(audio_queue, audio_cnn))
video_thread = threading.Thread(target=process_video_stream, args=(video_queue, visual_cnn))
audio_thread.start()
video_thread.start()
```

Figure 4. Real-time multimodal sentiment analysis

Step 4: Integrate an economic application

For example, sentiment scores in product reviews.

```
def calculate_sentiment_scores(predictions):
    sentiment_scores = {
        'Positive': 0,
        'Neutral': 0,
        'Negative': 0
    }
    for emotion in predictions:
        if emotion == 0:  # Assume 0 represents positive
            sentiment_scores['Positive'] += 1
        elif emotion == 1:  # Neutral
            sentiment_scores['Neutral'] += 1
        else:  # Negative
            sentiment_scores['Negative'] += 1
    return sentiment_scores

# Example predictions
audio_predictions = [0, 1, 2, 0, 0]
visual_predictions = [0, 2, 2, 1, 0]

# Combine and score
all_predictions = audio_predictions + visual_predictions
scores = calculate_sentiment_scores(all_predictions)
print(f"Sentiment Scores: {scores}")
```

Figure 5. Assessing the sentiment of product reviews

Performance rating: a sentiment-based evaluation method that summarizes the results of consumer feedback. Data visualization to analyze overall sentiment trends.

Step 5: Creating a monitoring panel

We use principles from Tufte's foundational work [Tufte, E. R. 2001, 197 pages] alongside the Streamlit library to create a monitoring panel and visualize sentiment.We can create the following dashboard to analyze sentiment during economic events (e.g., product releases or political announcements).

```
import streamlit as st

def display_dashboard(sentiment_scores):
    st.title("Live Emotion Dashboard")
    st.bar_chart(sentiment_scores)

# Simulate sentiment scores
example_scores = {
    'Positive': 15,
    'Neutral': 5,
    'Negative': 8
}
display_dashboard(example_scores)
```

Figure 6. Creating a monitoring panel

This system clearly demonstrates how multimodal sentiment recognition can be applied to economic scenarios, facilitating consumer analytics and market sentiment tracking. A dataset of consumer reactions to a new product was analyzed using a multimodal emotion recognition system. The system accurately recognized emotional patterns such as:

- Positive responses: smiles and applause from satisfied customers.
- Neutral reactions: minimal facial expressions and a balanced tone indicating indifference.
- Negative reactions: frowning eyebrows and a sharp, disappointed tone indicating dissatisfaction.

Through pattern analysis, we found that the ability to accurately assess emotions can influence pricing, marketing, and product design strategies, providing data-driven decision-making that is relevant to consumer needs.

- Market readiness: products with a high positive sentiment index have a higher chance of success.
- Feedback. Visual cues provide additional context for voice sentiment, improving the overall perception of consumer feedback.

Conclusion

Multimodal sentiment recognition systems based on convolutional neural networks (CNNs) have the potential to revolutionize economic analysis. Using audio and visual data, these systems can provide detailed information about consumer sentiment, market trends, and employee sentiment. The integration of these signals has been shown to have the potential to improve decision making and provide competitive advantage across sectors. Future research could explore the integration of advanced techniques such as attention mechanisms or transformational models, further improving accuracy. In addition, the creation of larger datasets specific to the economic context will increase the applicability of these systems in real-world situations. Emotion recognition is not just a technological advancement, it is a paradigm shift that increases the understanding of human factors in economics. activities to a new level. This paper focuses on the economic implications and use cases of multimodal emotion recognition, while maintaining the technical depth necessary for practical implementation.

References

1. Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press.
2. Schuller, B., et al. (2013). Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. IEEE Signal Processing Magazine.
3. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval.
4. Baltrušaitis, T., et al. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence.
5. Tufte, E. R. (2001). The Visual Display of Quantitative Information.

6. Viola, P., & Jones, M. J. (2001). Robust Real-Time Face Detection. International Journal of Computer Vision.
7. Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations (ICLR).
8. Zhang, Z., et al. (2018). Emotion Recognition Based on Multimodal Features and Multilevel Fusion. IEEE Transactions on Cybernetics.
9. Joshua Ayeni (2002) Convolutional Neural Network (CNN): The architecture and applications
10. Loewenstein, G., & Lerner, J. S. (2003). The Role of Affect in Decision Making. Handbook of Affective Sciences. Oxford University Press.

Aharon RUSHANYAN

Economic aspects of analysis related to the development of a multimodal emotion recognition

Key words: economic behavior, market sentiment analysis, multimodal analysis, convolutional neural networks, emotion-driven market trends, emotion recognition

This article, we investigate how systems of emotion recognition based on multimodal support through CNN can be useful in economic analysis. Traditional economic theories rely mainly on quantitative and qualitative information and essentially exclude vital non-verbal emotional cues in the determination of human actions. Engaging audio and visual pathways, this innovative approach, links psychological knowledge with economic one and thereby enhances comprehension of purchase decisions, market forces and workplace affective states. The approach entails capturing the audio and video data, feature extraction for the acoustics of the audio through MFCC and the human face using facial landmarks, and the use of CNN. The fusion layer integrates all such modalities for improving the accuracy of prediction. The practical use entails use of sentiment analysis and integrating incidences of collecting sentiment scores concurrently and in standard activities such as product release and effect of a policy. The effectiveness of the system to decode emotions makes easier for firms and policy makers to make economic decisions with an incorporation of emotions, a concept that is a milestone in the sphere of economical analysis. Further modifications are possible in subsequent studies with bigger datasets and more complicated networks associated with this system. The article also shows how multimodal emotion recognition systems can revolutionalise conventional economic models using the might of artificial intelligence technology and behavioral information.