Representation Learning for Optimal Cutting Time Detection in Milk Coagulation

Abstract

This project explores the use of self-supervised and contrastive learning to model milk coagulation, with the goal of identifying the optimal cutting time, a critical decision in cheese production. The dataset consists of 12 image sequences, each with only one frame labeled by an expert as the optimal point. The aim is to learn a representation space where distances reflect the progression toward coagulation, enabling the model to reason about how close a given frame is to the optimal cutting point.

Context and motivation

Cheese is the most important processed dairy product for direct consumption or as an ingredient. Its production process depends on several critical stages such as coagulation, syneresis, and ripening. Coagulation sets the course for the rest of the process: the exact moment the curd is cut determines whey syneresis, solids retention, and ultimately the final texture and yield.

The gel or curd must be cut when it has the appropriate rheological and microstructural properties. Therefore, determining the final coagulation time is one of the key variables in the process and is subject to high variability. However, in most factories, this decision is still made based on a fixed time or by visual, subjective inspection by an operator.

This project aims to identify the optimal cutting time, a key decision that directly impacts both yield and final product quality.

Student profile

I am currently in the first year of my PhD at the IIIA (Artificial Intelligence Research Institute, CSIC). I would like to take advantage of this TFM to align it with my doctoral research and ensure continuity between both works. My PhD advisor will also be involved as a co-supervisor for this project.

My thesis focuses on the milk coagulation stage, and I am currently finalizing a review article based on the following paper:

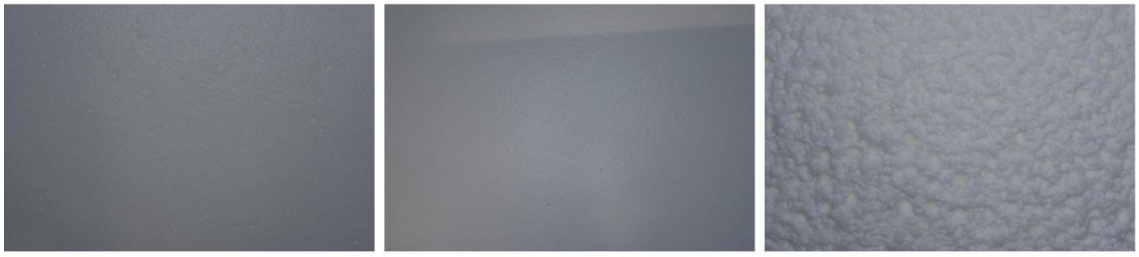
https://www.sciencedirect.com/science/article/pii/S0166361524001015

Proposal

The idea is to build upon the analysis and dataset used in the above study. The dataset consists of periodic image sequences from 12 coagulation processes, with frames labeled as either before cutting time (Pre CT) or at cutting time (CT):

Set	1	2	3	4	5	6	7	8	9	10	11	12
Pre CT	77	90	108	89	54	45	63	72	68	59	60	77
CT	17	12	20	23	23	25	21	24	37	35	29	34
Total	94	102	128	112	77	70	84	96	105	94	89	111

Set 1 Set 2 Set 3



Set 4 Set 5 Set 6

Set 7 Set 8 Set 9

Set 10 Set 11 Set 12

Each sequence differs in length and in the moment at which the optimal cutting time occurs. It is also important to note that coagulation is an event that always occurs in milk for cheese making, whether induced by acid or enzymatic addition.

The goal of this project is not to approach the problem as a binary classification task (as done in the reference paper), but rather to estimate how far each frame is from the

optimal cutting time. In other words, the aim is to model the progression of the process and understand the temporal dynamics leading up to the cutting point. The objective is to learn a representation space where distances reflect this progression, allowing the model to distinguish images based on their proximity to the optimal cutting time.

The data consists of temporally ordered image sequences, where a single frame is labeled as the optimal cutting time by an expert operator. However, intermediate timestamps are not available, and the labeled cutting point may not be perfectly precise, as it is based on subjective assessment. For this reason, self-supervised or contrastive learning approaches are well-suited to capture the underlying progression dynamics. Additionally, the limited expert supervision makes this a compelling few-shot learning scenario worth exploring.