Name: Huaiqin Shi , Zid:z5467071

1. Scope

1.1 Domain：

Fabric Defect Post-Processing Recommender

1.1.1 Intended Users：

Quality Inspection Engineer

Process engineer

production line supervisor

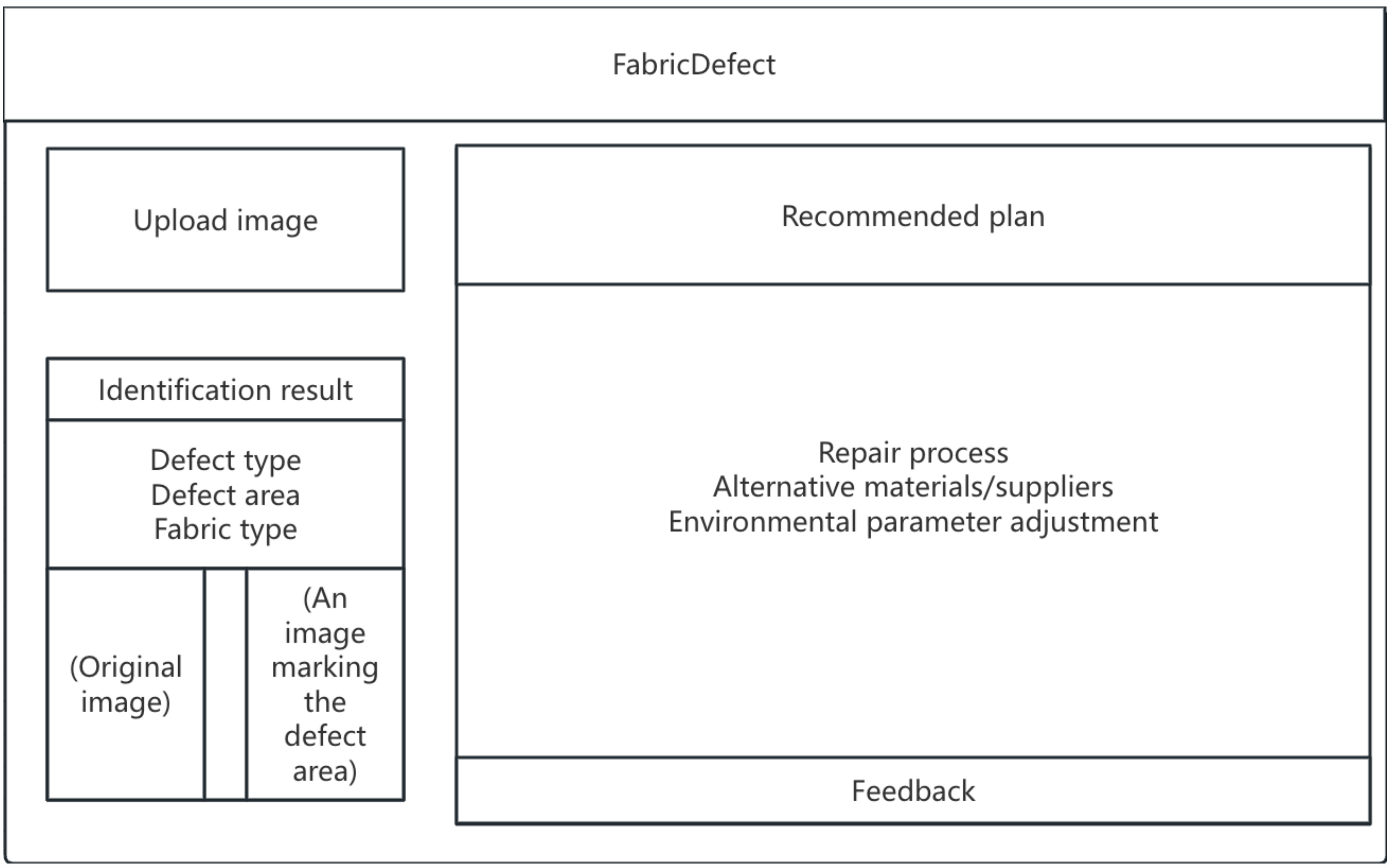
1.2.1 Recommended quantity

2

1.2.2 User Interface:

Web interface

1.3.1 Page sketch



1.3.2 Explanation

Collect user feedback regularly (weekly) and fine-tune the model in batches (incremental learning).

1.4.1 Cold start and dynamic update

cold start:

new defect type:New defect type: Use Cosine on defect-embedding to recommend nearest similarity processing.

new user:Preset rules according to "Fabric Type → Common Schemes".

1.4.2 dynamic update

Collect user feedback regularly (weekly) and fine-tune the model in batches (incremental learning).

Eliminate or retrain the "low adoption rate" schemes.

1.5 business model

1.5.1 SaaS Subscription

Basic version: several devices per factory

Enterprise Edition: Unlimited devices + API calls, multiple users, multiple permissions

1.5.2 Value-added

Customized development of expert knowledge graphs

Data reports and visual dashboards

2 Dataset

2.1 Source

(kaggle)https://www.kaggle.com/datasets/nexuswho/fabric-defects-dataset

2.2 Scale and Quality

2739 images (around 2G)

field	type	source	说明
fabric_type	Category	Annotation	Cotton, Polyester…
defect_type	Category	CNN Classified output	Crease, Knot…
defect_area_ratio	Numerical value(0–1)	Mask Calculation	(Defective pixels/Total pixels)
machine_id	Category	Factory database	M101, M202
ambient_temp, RH	Category	Sensor data	Ambient temperature (°C), relative humidity (%)
repair_action	Category	Expert annotation	reprocess, patch, reject…
repair_cost, time	Numerical value	Expert experience estimation	Expert experience estimation
user_rating	整数 (1–5)	User feedback	The satisfaction score of the recommended plan
adopt	Numerical value(0–1)	User feedback	Whether the user adopts it
Process	Category	Annotation	The methods of fabric processing

2.3 Usage method

Utilize train_test_split from sklearn.model_selection to hierarchical random split and perform a 5-fold CV on the training set and select the hyperparameters and model structure. Finally, the validation set is used for early stop, and then the test set is used to report the final performance

fabric_type, defect_area_ratio ,defect_type,machine_id, ambient_temp, RH,Process are used to build model

2.4 Limitations

The dataset only contains undamaged and defective images and does not include labels such as fabric. You need to label them ourself

3. Method

3.1 Content-Based Classification

3.1.1 Principle

Train the classifier based on the structured features of defects and fabrics to directly predict the optimal repair_action.

3.1.2 Scene

New users/New fabrics (Cold Start)

3.2 collaborative filtering

3.2.1 Principle

1. Based on the adoption or scoring of previous defect-fabric combinations by different quality inspectors, a "user × solution" interaction matrix is constructed;
2. Through matrix factorization potential factors between users and schemes are mined, enabling the preferences of similar users and the handling patterns of similar defects to draw on each other.
3. By using the inner product of the learned user factors and solution factors, predict the most likely repair solution that a quality inspector is likely to adopt in a new defect scenario, and provide personalized recommendations accordingly.

3.2.2 Scene

There are sufficient user feedback logs

3.3 Feasibility

Content-Based Classification has approximately 2,700 structured records, with a feature dimension of around 10. XGBoost/LightGBM trains at an extremely fast speed for this scale of data, with a total GPU time consumption of less than 30 minutes. It can also be completed on the CPU within 1 to 2 hours

The time cost of Collaborative Filtering matrix factorization and incremental update is also extremely low, which can complete project experiments and evaluations

4. Evaluation

4.1.1 Offline indicators

Category	Indicator	Explanation
Ranking accuracy	Precision@1 / Recall@3	Top-1、Top-3 Recommended accuracy rate
Classification performance	Accuracy, Macro-F1	Classification performance of each repair category
Regression prediction	MAE / RMSE	Prediction errors for repair_cost, time

4.1.2 Online indicators

Indicator	Explanation
Adoption rate	The proportion of users accepting the recommended plan
Average satisfaction	The average score of user_rating
Recommendation delay	A single recommendation takes a long time

4.2 Trade-offs and weights

Determine the key indicators and business priorities, assign corresponding weights to the importance, and make compromises based on the weights. The user adoption rate is the most important

4.3 Timeliness and Resources

Online recommendation: Lightweight real-time computing, with extremely short time consumption, can ensure immediate response

Regular full-weight training
Frequency: Once a week
Reason: Accumulate sufficient new interactions/feedback to ensure the stability of the model

4.4 User Experiment design

The interface process is very intuitive, and the "expert explanation" is a plus.
It is suggested to add a small pop-up window below the plan to display the before/after comparison chart of similar cases to assist in making the final decision.