

POP: Mining POtential Performance of new fashion products via webly cross-modal query expansion

–Supplementary Material–

Christian Joppi^{*1}[0000–0003–4495–9515], Geri Skenderi^{*2}[0000–0001–9968–7727],
and Marco Cristani^{2,1}[0000–0002–0523–6042]

¹ Humatics s.r.l

{name.surname}@sys-datgroup.com

² University of Verona

{name.surname}@univr.it

The supplementary material discusses the following topics, in order:

1. Additional ablation study on the signal forming step of POP (Sec. 1);
2. Details on the training data and forecasting horizon on the VISUELLE dataset (Sec. 2);
3. Complete results on all Fashion Foward (FF) dataset partitions for the new style popularity forecasting task (Sec. 3);
4. Qualitative demonstrations of the downloaded images from our temporal cross-modal query expansion, along with interesting insights discovered on the fashionability of different categories and colors (Sec. 4);
5. Inspired by latest, established guidelines in CV & ML research, a discussion on ethical, social and economical implications of our work (Sec. 5).

1 Additional POP ablations

This section provides an additional ablation related to the ones shown in Sec. 4.1 and Tab. 3 in the main paper, concerning the **signal forming step** of the POP signal. This is done by varying the embeddings of the cleaned images by CL that are used as input to Equation 3 (Sec. 3.5 in the main paper), based on the following strategies:

- *Negative*: it indicates the average distance of \mathbf{z} with the *pruned unfashionable images* $\{ \mathbf{x}'_i \}_{i=1, \dots, M''^{(t-k)} }^{(t-k)}$, substituting the positive ones from Equation 3;
- *Positive and Negative*: here we fed into the forecasting approach two signals, the original POP and the *Negative* one, making POP a multivariate (2D) time series.

We discover that the *Negative* approach gives some boost, probably accounting for how much the probe has to be dissimilar to unfashionable items. On the other hand, *Positive and Negative* shows a decrease in comparison to using only the

* indicates equal contribution.

Strategy	First Order Setup		Release Setup	
	WAPE	MAE	WAPE	MAE
<i>Negative</i>	52.68	28.78	53.90	29.44
<i>Positive and Negative</i>	52.97	28.94	54.35	29.69
POP	52.39	28.62	53.41	29.18

Table 1. Alternative versions of the signal forming step, comparing to the one proposed in Sec. 3.5, represented here as POP, on both the *first order setup* and *release setup* of VISUELLE. Lower is better for both metrics.

pruned fashionable images, probably because the two signals are complementary (Tab. 1). Nevertheless, all the ablated approaches still provide an exogenous time series that helps models perform better when they consider it as additional input.

2 Additional details on VISUELLE

The VISUELLE dataset from [3] contains the following, multi-modal information for each product: i) Images, ii) Text tags, iii) Google Trends, iv) Sales curves.

- **Images:** RGB images with an average size of 577×227 ;
- **Text:** Three types of text tags (category, dominant color, fabric) that come partially from the technical sheet of the product (category, fabric), while the dominant color has been automatically extracted by the authors of the dataset.
- **Google Trends:** Three 52-step long time series for each product, one for each of the tags listed above, extracted by the Google Trend platform in robust way [2] for the 52 weeks prior the sale start date;
- **Sales curve:** 12-step long weekly time series, reporting the sales of a particular item.

The testing product set is composed of the 497 most novel products of the two most recent seasons (Spring-Summer and Autumn-Winter 2019). The rest of the dataset (5080 products) is used for training. We *utilise all of this data* for the experiments in Sec 4.1 of the main paper. **Additional forecasting results** are displayed in Fig. 1 (in terms of WAPE), exhibiting the behaviour of the best performing model for all possible forecasting horizons.

3 Complete results on Fashion Forward

We show the full new style popularity forecasting results on all dataset partitions of Fashion Forward in Tab. 2, providing an extended version of Tab. 4 from the main text. The results show how POP can be used to estimate the intra dataset popularity trends of products from Fashion Forward with relatively high accuracy for all partitions. We would like to emphasize again that we use POP as input to forecasting models and compare the predictions with the ground-truth testing set that comes from the FF time series.

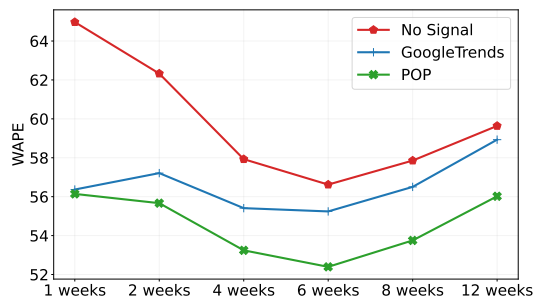


Fig. 1. WAPE for different forecasting horizons and exogenous signals, using GTM-Transformer [3] on the VISUELLE dataset. After six weeks there is a long enough history to model tendencies in the sales without considering product discounts or replenishments, unlike longer horizons. This is also reflected in the WAPE values, which keep increasing for forecasting horizons longer than six weeks. POP improves the forecasts for any horizon.

4 Qualitative Results

In this section we report a qualitative analysis of the POP signal, which in the main paper was limited to Fig.3 due to space limitations. These results give additional insight on the significance of our time-dependent, data-centric approach.

4.1 Under the hood of the POP signal

In Fig. 3 and Fig. 4 we report two examples of the (automatically) downloaded images used for the formation of the POP signal. In both figures, the probe images from which we extract the textual attributes to index the search are depicted. The analysis for each figure is reported in the corresponding caption. We also report in the figures some pruned images by the confident learning step, marked by a red cross.

4.2 The importance of a time-dependent query

In the main paper, we report in Tab. 3 various ablation studies on POP. Here we focus on two aspects in particular: the “Time Dependent Query Expansion”, and the “misaligned Past”. The obtained results suggest that exploiting (un)fashionable images not related to the date of delivery on the market gives worse results in terms of forecasting. Fig. 5 qualitatively demonstrates why this is the case. As it is visible, what made a garment of a particular type and color fashionable in 2017 (Fig. 5, top) does not correspond to the same visual elements that can be found in 2019 (Fig. 5, bottom). More specifically, throughout the

Global Average												
Signals	Mean		Last		Drift		AR		ARIMA		SES	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
Oracle	0.136	0.170	0.093	0.114	0.174	0.222	0.271	0.403	0.136	0.167	0.094	0.116
GoogleTrends	0.846	1.000	0.846	1.000	0.846	1.000	0.846	1.000	0.846	1.000	0.846	1.000
POP	0.152	0.192	0.116	0.144	0.182	0.229	0.281	0.418	0.235	0.293	0.125	0.156
Dresses												
Signals	Mean		Last		Drift		AR		ARIMA		SES	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
Oracle	0.155	0.197	0.130	0.158	0.203	0.263	0.307	0.409	0.173	0.209	0.129	0.157
GoogleTrends	0.849	1.000	0.849	1.000	0.849	1.000	0.849	1.000	0.849	1.000	0.849	1.000
POP	0.119	0.157	0.108	0.127	0.173	0.216	0.229	0.334	0.162	0.193	0.109	0.130
Shirts												
Signals	Mean		Last		Drift		AR		ARIMA		SES	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
Oracle	0.122	0.149	0.075	0.097	0.148	0.190	0.301	0.371	0.126	0.159	0.080	0.103
GoogleTrends	0.840	1.000	0.840	1.000	0.840	1.000	0.840	1.000	0.840	1.000	0.840	1.000
POP	0.144	0.175	0.109	0.152	0.166	0.215	0.274	0.336	0.139	0.189	0.111	0.151
Tops&Tees												
Signals	Mean		Last		Drift		AR		ARIMA		SES	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
Oracle	0.132	0.165	0.074	0.087	0.172	0.212	0.206	0.429	0.108	0.133	0.073	0.087
GoogleTrends	0.848	1.000	0.848	1.000	0.848	1.000	0.848	1.000	0.848	1.000	0.848	1.000
POP	0.193	0.245	0.131	0.153	0.206	0.257	0.341	0.585	0.405	0.497	0.156	0.186

Table 2. Results across all the Fashion Forward [1] datasets.

spring/summer season of 2017, the green kimonos tend to be heavily associated with white patterns and the color white in general. In 2019, the kimonos are almost all in different shades of green or even dark green.

5 Ethical Concerns and Societal Impact

Ethical implications could, in principle, arise from the web image search: observed images can, for example, contain copyrighted images. Nevertheless, just as a normal user would use Google Images to gather an opinion of what could be trending in fashion, so do we, albeit automatically. In particular, we do not need to personally look at the web images (apart for the ones shown in the paper and here for explanatory purposes), since POP is a just a numerical time series. **As for the societal impact**, our approach can be highly beneficial for fast fashion, which is the third most polluting industry in the world. Having a precise estimation of sales or popularity can improve the situation by solving supply chain issues and our pipeline can play a leading part. Ameliorated forecasts with

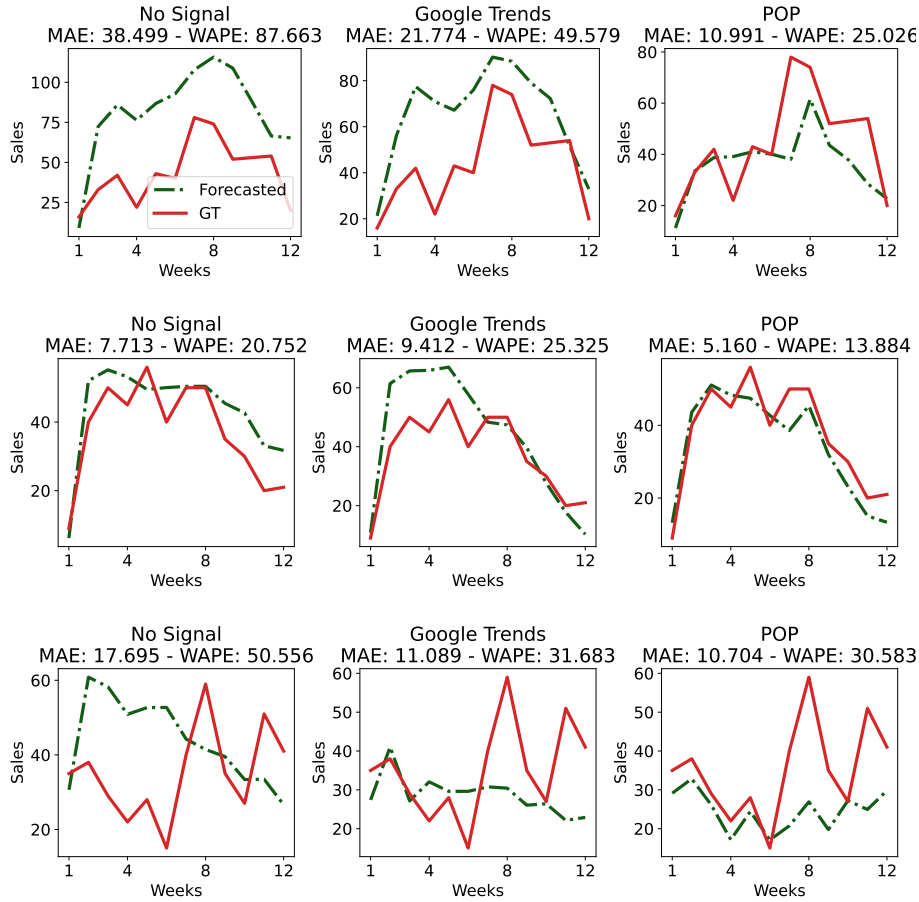


Fig. 2. Qualitative results on VISUELLE, considering all the 12 time-steps. In all the cases, using POP gives better results than not using it or using other, similar exogenous series.

POP also has a big impact at the economic level, in terms of profit. The best forecasting model on the VISUELLE dataset (which uses our generated time series, as seen in Tables 1 and 2 in the main paper), allows to spare 21% w.r.t. ordinary guidelines for new fast-fashion products, reducing a loss of \$4.390.400 US dollars to \$3.491.600 US dollars, assuming a general price of 28\$ per piece for all products (independently on the category).



Fig. 4. Examples of images downloaded for the query ‘Violet Long Sleeve’ (after pruning by confident learning). The “Fashionable Violet Long Sleeve” items seem to have a darker tone in most cases. Very long sleeves fade into dresses, indicating the length of the garment as an important aspect for making it fashionable. Curiously, “Unfashionable Violet Long Sleeve” contain brighter colors, short garments (like pyjamas) with writings or printed images. Pruned images are marked with a red cross.

«Fashionable» Green Kimono Dress (over the years)



Fig. 5. Examples of Fashionable downloaded images for particular time-dependent queries. In this particular case, for the query "green kimono dress", it can be seen how the notion of fashionability can have significant variations over time. Notably, green kimonos in 2017, as seen in the latter half of the first figure, tend to be heavily associated with white patterns and the color white in general. In 2019, this trend appears to be dying out, with the kimonos being of different shades of green or even dark green.

References

1. Al-Halah, Z., Stiefelhagen, R., Grauman, K.: Fashion forward: Forecasting visual style in fashion. In: ICCV (2017)
2. Medeiros, M.C., Pires, H.F.: The proper use of google trends in forecasting models (2021)
3. Skenderi, G., Joppi, C., Denitto, M., Cristani, M.: Well googled is half done: Multimodal forecasting of new fashion product sales with image-based google trends. arXiv preprint arXiv:2109.09824 (2021)