Detection of fashion trends and seasonal cycles through the analysis of implicit and explicit client feedback

Roberto Sanchis-Ojeda
Stitch Fix
One Montgomery Tower,
Suite 1200
San Francisco 94104, CA
rojeda@stitchfix.com

Daragh Sibley
Stitch Fix
One Montgomery Tower,
Suite 1200
San Francisco 94104, CA
dsibley@stitchfix.com

Paolo Massimi
Stitch Fix
One Montgomery Tower,
Suite 1200
San Francisco 94104, CA
pmassimi@stitchfix.com

ABSTRACT
In this contribution we describe a new approach to detecting seasonal and fashion trends, by statistically modeling how clients' reaction to style units change with time. In our framework, client reactions are required to take the form of binary outcome variables (e.g., buy vs. do not buy, click vs. do not click). Client behavior can then be studied with generalized linear models and mixed-effect models that include temporal features. We discuss how the coefficients of such models inform which styles are going in or out of season or fashion and demonstrate these methods using simulated data.

Categories and Subject Descriptors
H.4 [**Information Systems Applications**]: Miscellaneous

Keywords
Time-series analysis, Fashion, Generalized Linear Models

1. INTRODUCTION
Understanding how style preferences change with time is of critical importance to a fashion retail organization. Fashion trends can appear cyclically, with styles re-emerging after decades of absence. Fashion preferences can change on a yearly timescale as a result of changing weather. Other styles become very popular, but then quickly disappear (sometimes for the greater good of humanity). Identifying and anticipating which, if any, of these trends are affecting client preferences would enable more effective control of an organization's inventory.

Big data and statistical modeling provide a new and powerful method to quantify, verify, and falsify our understanding of fashion trends. However, there is a limited literature about detecting fashion trends using large datasets. One potentially powerful source of relevant data is social media. Studies have examined the content and tags of Twitter and Instagram posts [12], or have attempted to identify fashion topics on Twitter [4]. Google trends has also been used to identify changes in how types of clothing have been searched [17]. The whole web is full of information about fashion, with thousands of fashion blogs and online magazines, and aggregating such information could change the way to identify fashion trends [5].

In this contribution we will focus on detecting temporal trends in client interactions with an existing assortment of styles, rather than mining secondary data sources to infer what will become fashionable. In that sense, our work is related to the extensive literature on recommender systems, particularly methods which model customers preferences changing over time. The case of time-aware collaborative filtering seems to be well studied ([10], [2]), along with other types of time aware recommender systems ([15], [16]). An application of time-aware collaborative filtering to a large fashion sales dataset from Amazon clearly showed temporal patterns on the user preferences over a period of 10 years [8].

However, our current goal was not to recommend merchandise, but rather to identify and describe temporal trends. As a result, our approach is more statistically oriented: we propose to model changes in the underlying distribution of client interactions with different styles.

This contribution is organized in sections, with section 2 describing how client-style interactions are fundamentally Bernoulli trials. In section 3 we describe problems with aggregating these Bernoulli trials into Gaussian distributed ratios prior to analysis for temporal trends. In section 4 we describe a more statistically and computer intensive approach to modeling temporal trends in the Bernoulli trials for a single style. In section 5 we describe how to properly aggregate data for the case of a large number of styles, and how to model the aggregated data and still obtain statistically robust results. Through these sections, we demonstrate via simulations how to build a system to identify and classify fashion trends. Finally, in section 6 we describe how one could go beyond this mostly linear models into more complex non-linear models.

2. TURNING CLIENT RESPONSE INTO A BINARY VARIABLE

Interaction	Resp.	ClientID	Style ID	Date
Website click	0	23128	1	2016-01-01
Website click	1	88912	2	2016-01-01
Website click	0	34913	3	2016-01-02
Cart decision	1	34913	3	2016-01-02

Table 1: Extract from the typical database of client interactions with the different style units. Here style ID 1, 2, 3 could refer to a blouse, a scarf and a pair of pants for example.

Our method begins by conceptualizing every client interaction with a piece of clothing as a Bernoulli trial. Many interesting interactions of client C_i with a unit of style S_j will yield one of two outcomes, one that can be considered positive (purchase) and another one negative (no purchase). Within this framework, a unit of style can be flexibly defined. It may correspond to a fundamental piece in the hierarchy of styles, with a defined size, shape, fabric, color and other important attributes. Our method will simply assume that style units are defined so that when a client interacts with a style, they are always having a similar experience. As a result, changes in the underlying distribution of positive outcomes can be interpreted as a consequence of changing fashion trends. If pieces of clothing with different colors or other attributes happen to be part of the same style unit, adding a variable that could capture that difference in our statistical models is recommended.

To illustrate, consider a hypothetical "Fashion Online Store" (FOS). The client C_i logs into FOS website where she is greeted with an initial screen that shows her several pieces of clothing, that correspond to different styles S_j. The client clicks several of the options to check prices, higher resolution pictures and feedback from other clients. Among them, the client decides to include two items, each from one different style into the virtual shopping cart. At the end of the session, the client C_i decides to purchase one of the items, and remove the other from the cart.

Within this example many behaviors can be represented as binary variables. For example, we could record all the interactions with the n style units that were shown in the original screen, assigning 0 to all the negative interactions (not clicked) and 1 to the positive interactions (clicked). Similarly, we could record an extra interaction for each of the styles that were clicked, assigning a 0 to those that did not make it to the cart, and 1 to those that made it. Finally, the last interaction will give 0 to the item that was not bought and 1 to the one that was bought. A more traditional retail store could record instances where a given item was picked to try in the fitting room, and record also how many of those items were purchased. It could also record if a given sold item was returned within the returning period. In any case, after a certain period of time the business will have gathered a table with individualized interactions for each style and each client, and with an assigned positive or negative outcome for each of them (see Table 1).

3. USING THE NORMAL APPROXIMATION TO DETECT TEMPORAL TRENDS

One great benefit of using Bernoulli variables is aggregation, since by definition the total number of positive outcomes obtained after n Bernoulli trials with mean p follows a Binomial distribution with exact same value of p, and with a total of n draws. If we focused on one single interaction described in the previous paragraph, the best estimate of p given a data set with n_{pos} positive outcomes from n_{total} number of interactions is simply $p_{est} = n_{pos}/n_{total}$. In the case where n_{total} is large enough, the underlying binomial distribution of positive outcomes can be approximated by a Gaussian distribution with a width of $\sqrt{n_{total} * p_{est}(1 - p_{est})}$, giving a margin of error to define confidence intervals for the mean p_{est} of the distribution equal to the expected Gaussian width divided by n_{total}.

Via the "Website click" interaction, we have simply described the well studied problem of calculating click through rates with proper confidence intervals. The analysis of click-through data is extensive ([1], [6]). However, to our knowledge, the study of temporal trends is mostly restricted to short-term changes, that unfold over a few hours. For instance, studying if there are differences in click-through rate between day and night. Our temporal trends are expected to be much longer and smooth, over a timescale of months or even years.

For example, let us assume we have been gathering information on a given interaction for thousands of clients during a two year period. These interactions may be aggregated by month to obtain monthly average estimates of p, with associated 68.3% confidence levels (1-sigma). To simulate this scenario we could first decide how many client interactions happened in a day using a Poisson distribution with a mean of 3 interactions. For each interaction, we obtain a random Bernoulli draw and repeat the process for two years. Two different style units are analyzed, the first unit (style ID = 1) has a ratio of positive interaction responses p that starts at 0.7 and drops 0.2 per year, and second style unit (style ID = 2) is kept constant at 0.5. The left panels of Figure 1 show a plot of the estimates $p(t_i)$ and their uncertainties $\sigma(t_i)$, where i represents an index that runs through all month-long windows. As we can see, the negative temporal trend is much easier to visually detect for style 1.

To quantify this temporal trend, one could fit a linear regression model to the estimates of the positive outcome ratio, and obtain the best fit parameters and their associated uncertainties. Styles whose interactions have coefficients that are statistically significantly different from 0, could be considered as styles that are going in or out of fashion. If the number of draws per month is not uniform, we can use a generalized least square regression model in which we can set the weights of each estimate to be proportional to the number of interactions during that month or the inverse square of the estimated uncertainties.

This method for detecting linear temporal trends has a variety of limitations. There are several free parameters embedded within this analysis, whose values may affect the resulting conclusions. In particular, the window for the aggregation process is crucial. A short window will result in a low number of Bernoulli trials, breaking the assumptions of the normal approximation. A long window will result in a small number of aggregations, making it harder to fit

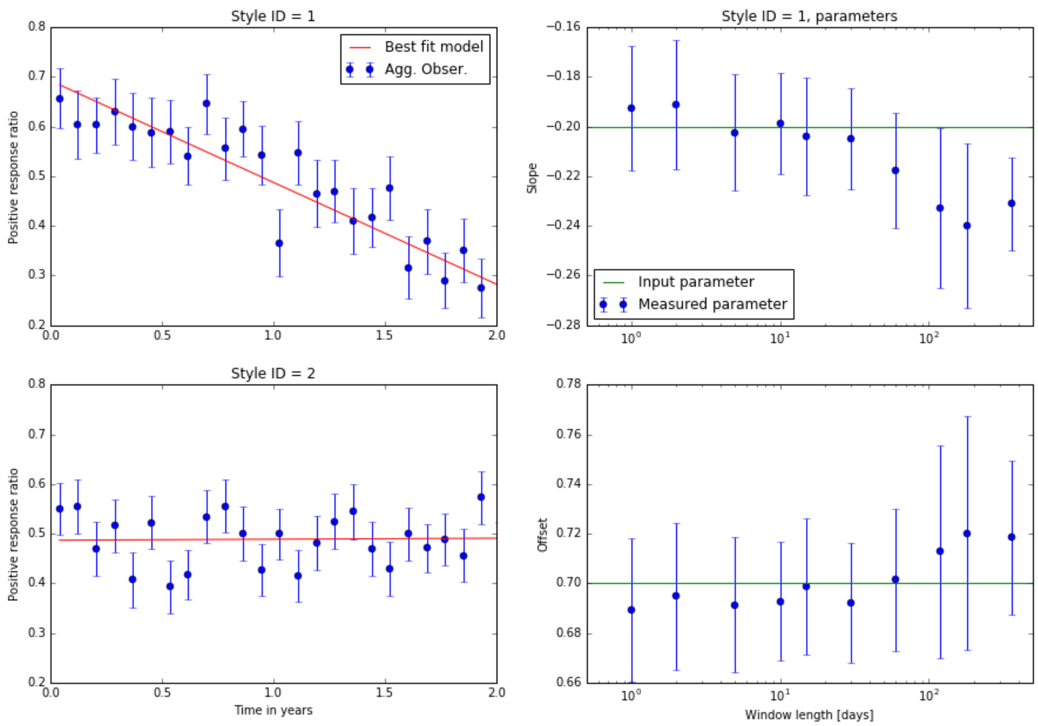

Figure 1: Left panels show the measured positive response ratios for two different styles, one with a simulated downward temporal trend and the other kept constant, with uncertainties estimated using the Gaussian approximation. A generalized least square linear fit can recover the input parameters within reasonable confidence intervals. Right panels show how these estimated parameters change with the temporal window used to aggregate the data, showing the problems associated with this approximated method.

a proper linear regression. Moreover, a long window can break our assumption that all grouped Bernoulli trials come from the same distribution, since p is likely to be changing with time if we have monitored the right interaction. Additionally, the estimated parameters depend on the selected window of aggregation, as shown in the right panels of Figure 1. This problem can be exacerbated in the presence of gaps in our data, which are likely to happen with the most seasonal style units (beach shorts, heavy scarves), or when we attempt to detect more complex temporal trends like cyclic trends.

4. FITTING ONE STYLE UNIT WITH GLMS

A robust method for identifying fashion trends should yield similar conclusions regardless of how we organize our data. This issue becomes even more acute when multiple factors are simultaneously affecting client behavior. In these situations, we should prefer a method that lets us directly model these additional sources of variance. Once we have modeled these effects we can marginalize over them to provide a more robust and sensitive detector of temporal effect.

In the case of a single style, our proposed method is to fit the individual Bernoulli trials with a Generalized Linear Model (GLM, [9]) with a set of temporal features and additional nuisance features. For example, "Website click" interactions for a given style, can be modeled as Bernoulli trials with a probability p that is affected by a linear decaying trend and a yearly sinusoidal cycle. Essentially, we can describe our client interaction variable as a Bernoulli variable with a mean probability of success $p(t, X)$

$$p(t, X) = p_0 + m * t + c_t * \cos\frac{2\pi(t - t_0)}{T} + \sum_i c_i * X_i(t) \quad (1)$$

where m is the temporal slope, c_t and t_0 are the amplitude and phase of the yearly cyclic trend, and T is the seasonal scale selected (one year in this case). The additional variables $X_i(t)$ can either be categorical or continuous, and can describe client features or other things that change with time. With a fix seasonal scale T, this model is still nonlinear because of t_0 but with a bit of algebra we can transform the cosine term into a linear combination of one sine and one cosine, which can be evaluated independently of the model parameters, allowing us to rewrite cyclical term as

$$c_t * \cos\frac{2\pi(t - t_0)}{T} = C_{sin} * \sin\frac{2\pi t}{T} + C_{cos} * \cos\frac{2\pi t}{T} \quad (2)$$

where the coefficients C_{sin} and C_{cos} are time-independent and can be written as

$$C_{sin} = c_t * \sin\frac{2\pi t_0}{T} \qquad C_{cos} = c_t * \cos\frac{2\pi t_0}{T} \quad (3)$$

The information contained in Table 1 can be used in combination with standard routines that fit GLMs using formulas, such as the statsmodels package in Python or the base glm() function in R, to obtain model coefficients with their associated uncertainties and z-scores that will allow us to understand the strength of different model features. One can also use standard feature selection techniques to asses whether the data justifies the presence of temporal trends, and then interrogate the model to classify styles as changing or not with time. In Figure 2 we can see an example of an interaction with the described level of complexity, and how a GLM fit was capable of obtain reliable parameters and uncertainties for all the different effects.

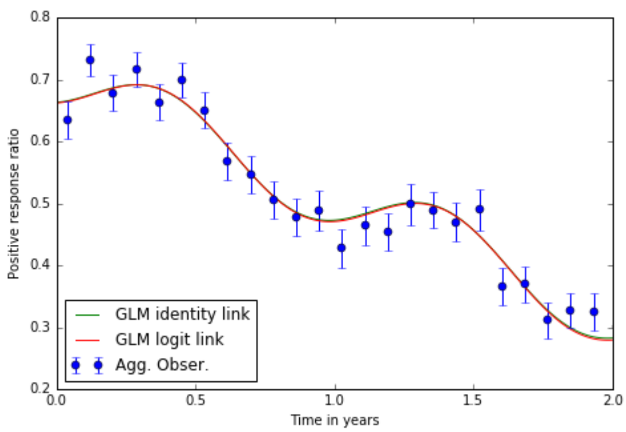

Figure 2: This plot shows the result of a simulation very similar to the example of Style ID = 1 in Figure 1. In this case, the number of samples per day was obtained from a Poisson with a mean of 10. The mean probability of positive outcome from which we obtain the Bernoulli draws had an additional sinusoid with an amplitude $c_t = 0.05$. GLM models with logit and identity links were fitted to the Bernoulli data, with results agreeing very well with the input parameters. The monthly aggregated probabilities of positive outcome are plotted in blue only to show how well the models fit the underlying Bernoulli data.

The example Figure 2 also shows that we have a choice over which link function to use when working with GLMs. The use of an identity link function sets the mean of the Bernoulli distribution to be a linear combination of the selected features. This is a natural choice, since this is precisely the kind of behavior that we have assumed for the styles going in and out of fashion. However, an identity link is unbounded, and could in principle reach values of the mean outside of the expected range for a probability of (between 0 and 1). In such cases, a logit function, that essentially turns the GLM into a logistic regression model, can help us fit the data, at the expense of a more complex interpretation of the coefficients.

5. STUDYING A LARGE NUMBER OF STYLE UNITS

In the example of one single style unit described in Figure 2, a single machine was enough to obtain coefficients for all features and temporal trends in a short amount of time. If we want to consider dozens of styles units, we can take the approach of fitting a model to each style unit and summarize the results a posteriori. We can also take the approach of fitting all unit styles together at once, which will give us the choice of selecting which coefficients are shared by all style units and which ones are allowed to be individual to the style.

Generally speaking, each style unit will have its own base rate of positive outcome, so each one will add an independent offset to the model. In some occasions we might be interested in measuring the global temporal properties of a set of style units, so we could use one shared coefficient for each temporal feature and obtain an average temporal strength over the different styles. It is important to note that within a GLM framework, style units with a larger number of client interactions recorded (larger n_{total}) will contribute more towards these averaged coefficients. The same will happen to any nuisance features added to remove other sources of temporal variations.

This approach becomes prohibitive when we pass certain thresholds for the number of style units, model features and Bernoulli trials per style unit. As a rule of thumb, if the number of parameters of the GLM is on the order of thousands and the number of Bernoulli trials on the millions, fitting a model will start to represent a heavy burden in a single machine. One can always resort to parallelized versions of these tools, such as the glm() function of Spark R, but this complex approach is not recommended during the exploratory phase of the project and might not be available to everybody. Here we propose two complementary solutions to alleviate this problem, categorical aggregation and Generalized Linear Mixed-effect models (GLMMs, [14]).

5.1 Categorical aggregation

The first proposed solution, categorical aggregation, refers to the process of aggregating the Bernoulli trials into appropriate binomial draws that are expected to share the same underlying probability of positive outcome. This is very similar to what was discussed in section 3, but in this case we are not going to use the normal approximation to estimate uncertainties, we will use the number of positive and negative outcomes in a given group as the input for our GLMs. It is crucial that each group shares the same probability of positive outcome according to our model, so feature selection will have a big influence over how we proceed. The most obvious aggregation group will likely be the style ID, since we probably want to fit each style unit independently. We call this procedure categorical aggregation because it will require to turn any continuous feature into categories, so we can properly group our observations by each value of each category. If for example we wanted to use the height of the clients as a feature, we will need to group them in categories (for example short, regular and tall) and then make sure that each categorical group has only information from one of the categories. In this particular example, we can be as extreme as to define one category for each possible height within a 1 inch margin, and we could essentially obtain the same result as with a continuous feature. We would only need to use the mean height of each group as a continuous feature in our model rather than a categorical variable.

This method has its limitations, because as soon as we have

Interaction	# Pos.	# Neg.	Style ID	Date
Website click	20	35	1	2016-01-01
Website click	27	30	2	2016-01-01
Cart decision	2	6	3	2016-01-01
Website click	38	28	3	2016-01-02

Table 2: Extract from the aggregated table that allows for a more computer efficient analysis of temporal trends.

more than a few features, the total number of groups, which might be as large as the product of the number of levels for each feature, can become prohibitive. During the exploratory phase, using a low number of levels per feature is recommended, and more granular levels can be chosen for the final model if the right computer resources are available.

Categorical aggregation through the time component will also be required if we want to have a solid improvement in computation speed. Again, we face a trade-off when selecting the length of each temporal category, where a longer window will reduce the number of data points in our model, increasing computational speed, but it will also reduce our temporal resolution. However, in this case we have no statistical argument to select a large window, as we do not need to gather a large number of trials per category for the model to fit properly. Our recommendation is to select a window length that allows us to have at least 20-40 temporal categories for an initial fit, and then rerun the analysis with a larger number of categories to make sure that the coefficients do not change significantly. Pushing the length of the window to a reasonable minimum, like for example one day, is probably best to avoid biases.

The outcome of our categorical aggregation process will be similar to what we can find in Table 2. Here we have the positive and negative responses for each interaction type and style unit, grouped by date. We could in principle leave time as a categorical variable, but it would be more time efficient to turn it into a discrete variable, for categorical aggregation, calculating first the amount of time between each date and an arbitrary zero point. If we call that t, we can now model the probability of positive interaction $p(t, X)$ using the formula from the previous section (see 1). By aggregating over the right categories, we have generated a set of Binomial draws that are as simple to model as a Bernoulli trial, but with a massive reduction in data volume. To implement GLMs with Binomial draws in R, the formula will change from "$Resp \sim t + \ldots$" to "$Cbind(\#Pos, \#Neg) \sim t + \ldots$", whereas the statsmodels package will require the slightly different version "$\#Pos + \#Neg \sim t + \ldots$".

Within this framework one can generalize to the case of several style units, as we would only need to add Style ID into the formula. If it is included as another term in the sum, then it will yield an offset for each style unit, and all the temporal coefficients will represent a weighted average over all styles. If we add the term as an interaction by multiplying one or all of the present terms, we will get the corresponding coefficients for each style. However, when we are interested in individual coefficients for each style, running the code for each style might be more efficient, in order to avoid using too much computer memory at once.

5.2 Using GLMMs for great speed gains

The type of approach described in the previous subsection can run into computation problems if the number of style units is very large. Even after categorical aggregation, the prohibitive number of coefficients might make GLMs a poor choice to solve the problem. Here is when the second proposed solution becomes important, since using GLMMs allows us to increase computation speed without compromising the integrity of our statistical analysis. GLMMs are also very convenient because they allow us to set some features as fixed effects that are shared across all style units and others treated as random effects, so that we can identify particular style units that have a higher or lower coefficient for a given effect. A simple version of a GLMM could assume that the fixed part of the model depends linearly on time, and add the offset and slope of that linear trend as a random effect specific to each style. The output would be a fixed effect coefficient for the mean offset and temporal slope, and random coefficients for the same variables for each style unit that express how much each style differs from the mean behavior of all styles combined. The power of GLMMs resides in the fact that all the offsets are forced to be random draws from a Gaussian distribution, instead of being completely independent coefficients. We recommend this approach when the number of style units is large enough to constrain the random effect distribution meaningfully.

5.3 A temporal trend classification algorithm

When using GLMMs to fit a dataset with a large number of style units, including certain temporal features as random effects can potentially be used to build a classification algorithm of the styles based on the values and uncertainties of the random effect coefficients. In order to showcase this potential, and to check that our models behave well with large data sets, we designed two simulations based on the example described in Figure 2. We draw Bernoulli samples every day for 2 years, with the number of samples drawn from a Poisson distribution, but in this case we come back to the original Poisson mean of 3.

In the first simulation, we simulate 1000 different styles independently, where each one has a simulated probability of positive outcome of $p_i(t) = p_{0,i} + m_i * t$. The values of $p_{0,i}$ are drawn from a Gaussian distributions centered at 0.6 and width 0.1, whereas the values of m_i are drawn from a uniform distribution between -0.1 and 0.1 (per year). We made sure that these parameters were mathematically acceptable, checking that the simulated $p_i(t)$ were bound between 0 and 1. We then performed temporal categorical aggregation with a length of 7 days, which improved our speed without compromising the integrity of the analysis. With the aggregated data in hand, we ran a GLMM model with time as the only feature, using a logit link, since using an identity link caused some convergence problems, and we are not interested in the coefficients per se, just in their ranking capabilities. The fixed coefficients were in great agreement with the global input parameter distributions.

The lme4 [3] and sjPlot [11] R packages offer 95% confidence intervals for the random coefficients, so we can classify the styles according to simple statistical arguments. First we add the temporal slope fixed coefficient to the confidence intervals for slope random coefficients. The fix coefficient uncertainties are assumed to be small enough to be negligible when compared to the random coefficient uncertainties. Those styles where the entire confidence interval for the slope random coefficient is negative are classified as going out of fashion, whereas those with a completely positive confidence interval are the trending styles. Since we have kept the coefficients used to simulate the interactions for each style, we can now see which fraction of the items is classified as trending or out of fashion giving the initial input. The upper panel Figure 3 shows the results, where it is clear that the classification does very well the larger the temporal slope (as expected, because of a higher statistical power) and has a really small rate of false positives (simulated styles with a negative slope classified as trending and vice versa). It is important to note that the shape of this curve depends critically on the fact that all styles have a similar number of observations, so most of them have a very similar margin of error. The values at which the percentage classified starts dropping could have also been reduced by simply simulating the temporal trends with more observations. Finally we can reduce the false positive rate even further by using 99% confidence intervals as an example.

A similar classification algorithm can be build to detect the most seasonal unit styles according to our client interactions. In this second simulation, we simulate 1000 different styles independently, where each one has a simulated probability of positive outcome of $p_i(t) = p_{0,i} + c_i * cos(2\pi(t - t_0))$. The same input distribution was used for $p_{0,i}$, whereas c_i was drawn from a uniform distribution from 0 to 0.1. The parameter t_0 was drawn from a uniform distribution form 0 to 1 year, in order to avoid having a very similar set of simulated sine and cosine coefficients. The same bound check and temporal categorical aggregation was applied in this case. The GLMMs model that we ran was very similar to the one used in the first simulation, but substituting the t term by a sum of a cosine and a sine term, as described in section 4.

The fixed coefficients agree well with expectations, and we add them to the random coefficients in order to classify the styles. The treatment is a bit more complex, since a large amplitude of seasonality can be transformed into a large sine coefficient and a small cosine coefficient (or vice versa) depending on the original value of t_0. We decided to classify a style as seasonal if at least one of the sine or cosine confidence intervals was incompatible with zero, and the results validate this approach (see lower panel of Figure 3). More complex analysis where the correlation between the different sine and cosine coefficients might help improve the precision of this algorithm.

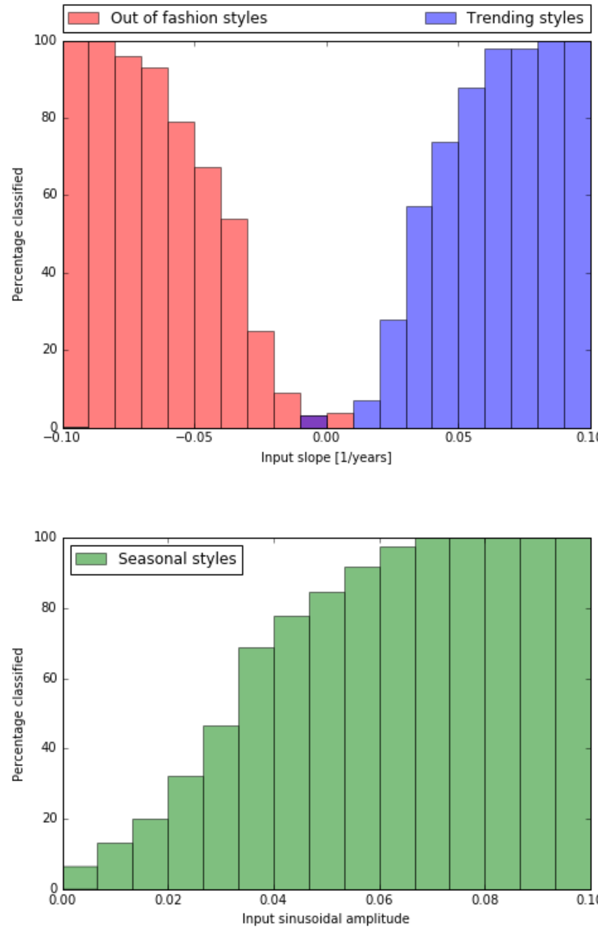

Figure 3: Outcome of our fashion trend classification algorithm for two different simulations. The upper panel simulates styles with a positive or negative linear trend in the probability of positive outcome, whereas the lower panel represents styles with a yearly sinusoidal trend. In both cases we bin the simulated styles by their simulated input parameters, and show the detection rate of our classifier. As expected, the algorithm is good at classifying styles with large temporal effects, and the rate of false positives (small amplitudes or even wrong sign) is small. This rate could be reduced by increasing the level of confidence required for the classification.

6. DISCUSSION AND FUTURE DEVELOPMENTS

In this paper we explored several statistical methods for identifying and quantifying linear and cyclical fashion trends. First, client interactions represented as Bernoulli trials were temporally aggregated and the resulting ratios were modeled with simple linear regression methods. We discussed the limitations and drawbacks to this approach, and demonstrated the advantage of using methods such as Generalized Linear Models and Mixed-Effect Models that take Bernoulli or Binomial draws as inputs and can model additional sources of variance.

We anticipate that choosing, exploring and defining the right additional features will be the hardest problem in most cases, as many important factors change with time in real business applications. By controlling these nuisance factors, we may increase the probability that we detect the fashion trends of interest.

Finally, we presented simulations to demonstrate how this process might identify styles that are going out of fashion (trending) by finding the style units that have a significantly negative (positive) temporal slope; and identify seasonal styles when coefficients for sine and cosine features significantly differ from zero.

Throughout this paper we have only discussed two different types of temporal effects, a local linear trend and a cyclic trend represented as a sinusoid. Not only these representations are simple and useful, but they are also convenient, since they allow us to use linear models. Extensions of this framework could include several coefficients to describe the changes with time as a high order polynomial. Initial attempts to carry such fits show that regularizing the fits becomes critical to avoid over-fitting. They could also describe time as a completely categorical variable, and obtain one coefficient per month or quarter, to study the temporal structure of those coefficients via other methods.

In a general sense, the GLM models employed here are simply obtaining the coefficients that maximize the log-likelihood of a modeled underlying distribution with a time dependent mean given the data at hand. Within a Bayesian framework, one could design a functional p that could depend on a series of features and parameters in a non-linear way, and create a series of tools to obtain the model parameters that maximize the log-likelihood of the model. A great application of this would be to leave the seasonal timescale T of the cyclical component as a free parameter, to obtain the true periodicity of the trends. One could then use Markov Chain Monte Carlo ([13], [7]) tools to explore parameter space and obtain the appropriate parameter uncertainties. Of course, such methods will be more time consuming and should only be justified in the presence of very complex temporal patterns.

7. ACKNOWLEDGMENTS

We thank the other Data Scientists of the Merchandise Product Development team at Stitchfix, Andrew Almand-Hunter, Erin Boyle, Chris Rinaldi and Zhou Yu for their contributions to the concepts developed in this paper, and their feedback on the manuscript. We also want to thank Eric Colson and the rest of the Algorithms & Analytics team at Stitchfix for cultivating an atmosphere that encourages innovative solutions to our current business challenges.

8. REFERENCES

[1] D. Agarwal, B.-C. Chen, and P. Elango. Spatio-temporal models for estimating click-through rate. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 21–30, New York, NY, USA, 2009. ACM.

[2] L. Baltrunas and X. Amatriain. Towards time-dependant recommendation based on implicit feedback. In *Workshop on context-aware recommender systems (CARS'09)*, 2009.

[3] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[4] S. Beheshti-Kashi. Twitter and fashion forecasting: An exploration of tweets regarding trend identification for fashion forecasting.

[5] S. Beheshti-Kashi, M. Lütjen, L. Stoever, and K. Thoben. Trendfashion - A framework for the identification of fashion trends. In *45. Jahrestagung der Gesellschaft für Informatik, Informatik 2015, Informatik, Energie und Umwelt, 28. September - 2. Oktober 2015 in Cottbus, Deutschland*, pages 1195–1205, 2015.

[6] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *Proceedings of the 27th International Conference on Machine Learning ICML 2010, Invited Applications Track (unreviewed, to appear)*, June 2010. Invited Applications Track.

[7] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[8] R. He and J. McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web*, pages 507–517. International World Wide Web Conferences Steering Committee, 2016.

[9] R. W. M. W. J. A. Nelder. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384, 1972.

[10] Y. Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 447–456, New York, NY, USA, 2009. ACM.

[11] D. Lüdecke. *sjPlot: Data Visualization for Statistics in Social Science*, 2016. R package version 2.0.0.

[12] L. Manikonda, R. Venkatesan, S. Kambhampati, and B. Li. Evolution of fashion brands on twitter and instagram. *CoRR*, abs/1512.01174, 2015.

[13] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *JCP*, 21:1087–1092, June 1953.

[14] D. G. C. N. E. Breslow. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

[15] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 363–372. ACM, 2013.

[16] Y. Zhang, M. Zhang, Y. Zhang, G. Lai, Y. Liu, H. Zhang, and S. Ma. Daily-aware personalized recommendation based on feature-level time series analysis. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1373–1383. International World Wide Web Conferences Steering Committee, 2015.

[17] O. Zimmer and Y. Horwitz. Google fashion trends report - spring 2015. `https://think.storage.googleapis.com/docs/google-fashion-trends-report-spring2015.pdf`, 2015.