Sales Forecasting, Polls vs Swarms

Gregg Willcox, Louis Rosenberg, Hans Schumann
Unanimous AI, San Luis Obispo CA, USA
gregg@unanimous.ai

Abstract. Sales forecasts are critical to businesses of all sizes, enabling teams to project revenue, prioritize marketing, plan distribution, and scale inventory levels. To date, however, sales forecasts of new products have been shown to be highly inaccurate, due in large part to the lack of data about each new product and the subjective judgements required to compensate for this lack of data. The present study explores product sales forecasting performed by human groups and compares the accuracy of group forecasts generated by traditional polls to those made using Artificial Swarm Intelligence (ASI), a technique which has been shown to amplify the forecasting accuracy of groups in a wide range of fields. In collaboration with a major fashion retailer and a major fashion publisher, groups of fashion-conscious millennial women predicted the relative sales volumes of eight sweaters promoted during the 2018 holiday season, first by ranking each sweater's sales in an online poll, and then using Swarm software to form an ASI system. The Swarm-based forecast was significantly more accurate than the poll. In fact, the top four sweaters ranked by swarm sold 23.7% more units, or $600,000 worth of sweaters during the target period, as compared to the top four sweaters as ranked by survey, (p = 0.0497), indicating that swarms of small consumer groups can be used to forecast sales with significantly higher accuracy than a traditional poll.

Keywords: Swarm Intelligence, Artificial Intelligence, Collective Intelligence, Sales Forecasting, Product Forecasting, Customer Research, Market Research, Customer Intelligence, Marketing, Business Insights

1 Background

Accurate sales forecasting is critical to businesses of all sizes, enabling teams to project revenue, prioritize marketing, plan distribution, and scale inventory levels. In recent years, AI has been used to assist in sales forecasting, but traditional AI tools are heavily reliant on historical sales data. Unfortunately, there are many situations where little to no historical sales data exists to support forecasting. This is especially true when launching new products or features or entering new markets [1]. When historical data is sparse, human judgement methods such as focus groups, customer interviews, expert opinions and customer intention surveys are often used in place of the missing data to forecast the demand of new products [2]. In one study of the prevalence of new product forecasting techniques, customer and market research was found to be the most widely used technique for forecasting the sales of new products, and that it was used in the majority (57%) of all new product forecasts [3].

While human judgement methods such as customer surveys are widely used in new product forecasting, the overall forecasting accuracy of these techniques is low: a new product's forecasted sales have been found to be on average 42% off from the true unit sales, and the more novel the product, the higher this error rate [3]. More accurate forecasting techniques are therefore necessary to improve the efficiency of bringing new products to market.

The present study proposes a novel methodology for sales forecasting using an emerging form of AI called Artificial Swarm Intelligence (ASI). If effective, ASI offers a major benefit, as it does not require historical data but instead uses real-time customer input. The study then compares the accuracy of traditional polling, using consumer surveys, to real-time "customer swarms" using ASI when making forecasts.

Previous research has shown ASI to be a powerful method for amplifying the predictive accuracy of networked human groups [4, 5]. A variety of prior studies, across a wide range of tasks, have indicated that real-time "human swarms" can produce significantly more accurate forecasts than traditional "Wisdom of Crowds" methods such as votes, polls, and surveys [6]. One study tasked human groups with predicting a set of 50 soccer matches in the English Premier League. The results showed a 31% increase in accuracy when participants collaborated as ASI swarms [7]. The ASI swarms also outperformed the BBC's machine-model known as "SAM" over those same 50 games [8]. Similar increases have been found in other studies, including a study at Stanford Medical School that showed small groups of radiologists were able to diagnose pneumonia with significantly higher accuracy when using Swarm software than taking votes or diagnosing as individuals [9].

Although previous research has shown that ASI technology can empower groups to outperform individual forecasters as well as traditional crowd-based methods, no formal study has been conducted in the domain of corporate sales forecasting.

1.1 From Crowds to Swarms

When collecting input from human groups, the phase "Wisdom of Crowds" is often used when aggregated input is used to generate output of higher accuracy. [10-12] Also referred to as Collective Intelligence, these methods date to the early 1900's and typically involve collecting polling data from individuals and computing a statistical result. When comparing "swarms" and "crowds," the primary difference is that in crowd-based systems, the participants provide isolated input that is aggregated in external statistical models, whereas in swarm-based systems the participants interact in real-time, "thinking together" as a unified system. In other words, crowds are statistical constructs while swarms are closed-loop systems in which the participants act, react, and interact in real-time, converging together on optimized solutions.

ASI systems are generally modeled on biological systems such as fish schools, bird flocks, and bee swarms. The present study uses the Swarm software platform from the company Unanimous AI. The core technology, called Swarm AI, is modeled on the collective decision-making processes employed by honeybee swarms [7]. This framework was chosen because honeybee populations have been shown to reach highly optimized decisions by forming real-time systems [13]. In fact, at a structural level, the decision-making processes observed in honeybee swarms are very similar to the decision-making processes observed in neurological brains [14,15].

When reaching decisions, swarm and brains both employ large populations of simple excitable units (i.e., bees and neurons) that operate in parallel to (a) integrate noisy data about the world, (b) weigh competing alternatives when a decision needs to be made, and (c) converge on preferred decisions as a unified system. In both brains and swarms, outcomes are arrived upon through competition among sub-populations of simple excitable units. When one sub-population exceeds a threshold level of support, the corresponding alternative is chosen by the system. In honeybees, this enables the group to converge on optimal decisions across a wide range of tasks, for example when selecting the best possible hive location from a large set of options. Researchers have shown that honey bees converge on the best possible solution to this life-or-death decision approximately 80% of the time [16,17].

1.2 Creating Human Swarms

Unlike birds, fish and bee, humans have not evolved natural mechanisms to establish feedback-loops among members. Fish for example, when moving in schools, detect faint vibrations in the water around them. Birds, when flocking, detect subtle motions propagating through the formation. Honey bees, when reaching decisions as a swarm, use body vibrations called a "waggle dance" to encode their views. To enable human groups to form similar systems, specialized software is required to close the loop across all members. A software platform (**Swarm**) was created to allow online groups to form real-time systems from anywhere in the world [4,18]. Modeled on the decision-making process of honeybees, the Swarm platform enables groups of online users to work in parallel to (a) integrate noisy evidence, (b) weigh competing alternatives, and (c) converge on decisions together as a real-time closed-loop system.

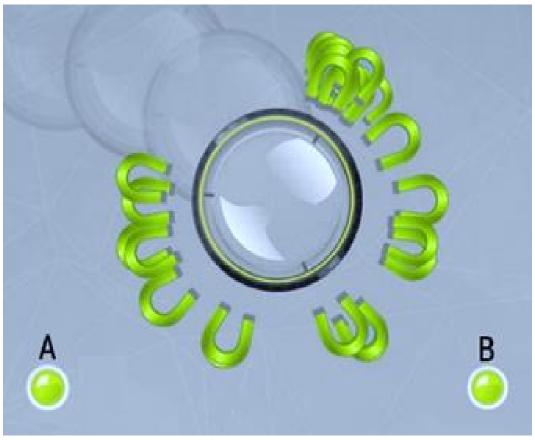

Fig. 1. Real-time ASI choosing between options

As shown in Figure 1 above, human swarms answer questions by moving a graphical puck to select among a set of options. Each participant provides input by moving a graphical magnet with a mouse, touchpad, or touchscreen. By adjusting their magnet in relation to the moving puck, real-time participants express their intent on the system as a whole. The input from each individual is not a vote, but a continuous stream that varies freely over time. Because all members can vary their intent continuously in real-time, moderated by AI algorithms, the group collectively explores the decision-space, not based on the input of any single user, but based on the emergent dynamics of the system as a whole. This enables complex deliberations to emerge in real-time, empowering the group to converge on the solution that best represents their combined knowledge, wisdom, and insights.

It is important to note that participants do not only vary the orientation of their individual intent, but also modulate the magnitude by manipulating the distance between their magnet and the puck. Because the puck is in fluid motion throughout the decision-space, users need to continuously update the position and orientation of their magnet so that it stays close to the puck's outer rim. This is important, for it requires participants to remain engaged throughout the decision-making process, continuously evaluating and re-evaluating their individual thoughts and feelings with respect to the question at hand. If they stop moving their magnet in relation to the changing position of the puck, the distance grows and their applied sentiment wanes.

2 Method

To compare the predictive accuracy of ASI to traditional polling in sales forecasting, the relative sales performance of eight sweaters was forecast using both surveys and swarms. The sweaters were all part of a new line designed for the 2018 holiday season by a major fashion retailer, so there was little historical data that could be used to accurately forecast the relative sales performance of each sweater. In addition, each sweater was a different color or pattern, or featured a printed graphic, so forecasting the relative sales of each sweater was highly subjective.

Fig. 2: Example Sweaters used for Sales Forecasting

The intended market for the sweaters were millennial women, so a group of thirteen millennial women were convened as a representative group of 'experts' to forecast the sales of each sweater. All of the participants in this study self-identified as interested in fashion with no previous sales forecasting experience. Participants were also not co-located for this experiment.

To compare the forecasting accuracy of traditional surveys to ASI swarms, the group was asked to predict relative unit sales (i.e., the ordered rank by sales) of the eight sweaters, first as individuals using an online survey, then by "thinking together" as an AI-optimized system using Swarm. The online survey consisted of one task: ranking the sweaters in order of total unit sales over the holiday period, from most (1st) to least (8th). The mean rank of each sweater over all surveys was used to generate an ordered list of sweaters, from most (1st) to least (8th) forecasted unit sales.

Next, all participants logged in remotely to the *swarm.ai* website, and used the Swarm software to collectively rate the sweaters on two critical metrics, Trendiness (from 1 to 5) and Breadth of Appeal (from 1 to 5). After rating the sweaters on those metrics, the group used a "process of elimination" method to sequentially rank the eight sweater styles on projected sales units, from least to most. Approximately three months later, the actual unit sales of each item over the holiday period was reported to the researchers by the fashion retailer, and the performance of each Sales Forecasting method was compared.

3 Results

The actual unit sales of each sweater, as percent of total sales, are reported in Table 1 below along with the rankings generated by the survey and the swarm respectively. Sweater names and sales numbers have been anonymized for confidentiality.

Table 1. Actual Sales Performance along with Forecasted Sales Rankings

Sweater	Percent of Total Sales	Survey Average Forecasted Sales Rank	Swarm Forecasted Sales Rank
A	23.4%	3	2
B	21.7%	7	3
C	16.8%	1	1
D	15.4%	2	5
E	12.0%	8	6
F	8.4%	5	4
G	1.2%	4	8
H	1.1%	6	7

4 Analysis

The results shown in Table 1 above reveal that the swarm generated a far superior ranking than the survey, correctly identifying the top three sweaters by sales volume, as well as correctly identifying the bottom two. That said, a rigorous analysis is required to compare the rankings methods with statistical significance. To perform this comparison, the forecasted rank and real-world performance (in unit sales) of each item was used as a measure of the quality of the ranking: the more the unit sales of each item reflected the forecasted rank, the better the ranking.

Specifically, the quality of the rankings was compared by calculating the percentage of total unit sales accounted for by the top n items in each ranking (Table 2). The cumulative unit sales generated using the swarm to select sweaters is greater than or equal to the cumulative unit sales of the survey for all cutoffs **n**, indicating that participants were regularly able to predict the sales performance of each sweater more accurately as a swarm rather than by survey. In other words, if this group was asked to predict the top **n** sweaters by unit sales performance, they would have selected a far better-performing set of sweaters by swarming than by providing input through a traditional poll.

Table 2. Cumulative proportion of total unit sales and p-value of difference

Number "*n*" of Sweaters Selected	Top N Survey Unit Sales	Top N Swarm Unit Sales	p-value
1	16.8%	16.8%	0.9954
2	32.2%	40.2%	0.4140
3	55.6%	61.9%	0.0001
4	56.8%	70.3%	0.0497
5	65.2%	85.7%	0.0394
6	66.3%	86.8%	0.1573
7	88.0%	98.8%	0.0046
8	100.0%	100.0%	1.0000

Statistical Significance: to calculate the probability that the swarm outperformed the survey by chance, the average survey ranking of each sweater was bootstrapped 10,000 times by randomly resampling participant surveys with replacement from the original pool of surveys. An ordered list of forecasted sweaters was generated for each bootstrapped survey using the average ranking of each sweater in that bootstrap.

The cumulative sales for each bootstrap was recorded, and the proportion of bootstraps was calculated where the survey matched or outperformed the swarm's cumulative sales performance for each of the top N-ranked sweaters (a p-value). The p-values for each rank are listed in Table 2. We find that the swarm significantly outperforms the survey ($p<0.05$) for the selection of the top 3, 4, 5, or 7 sweaters.

To visually compare swarm and survey performance, the cumulative unit sales of each method is plotted in Figure 2, including a black dotted line that marks the cumulative sales that would have been achieved by perfectly ranking the sweaters.

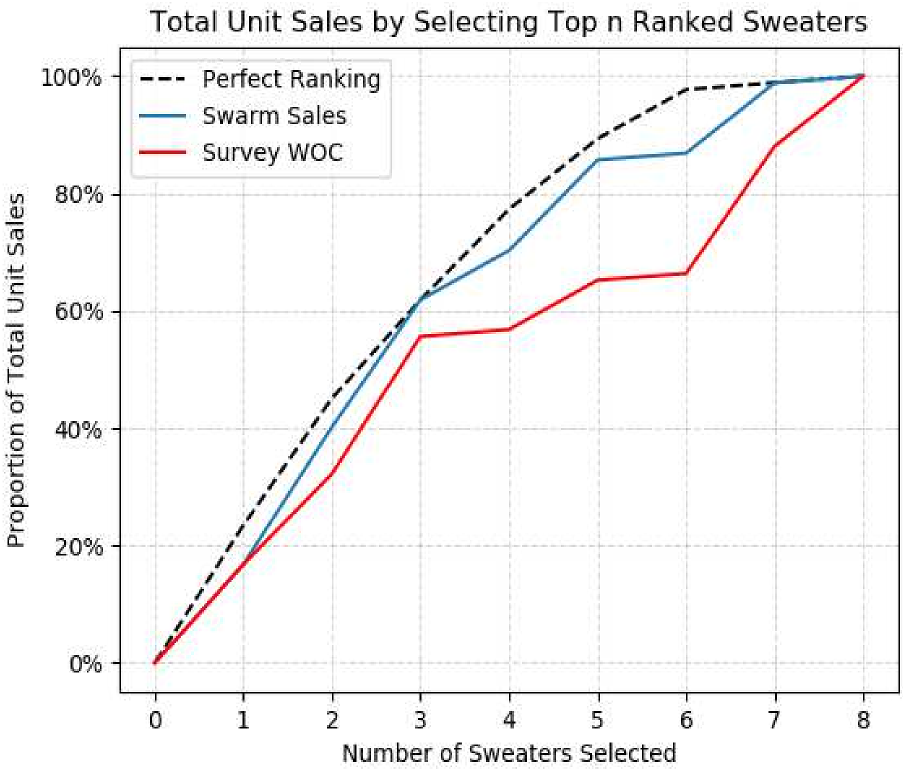

Fig. 2. Unit Sales for top N ranks by the swarm and survey methods

Looking at Figure 2 above, we can see that if a business wanted to predict the top selling products from a set of offerings, for example to plan inventory levels, they would get more accurate insights by using the swarm over the survey. For example, we can compare the actual sales of the top four sweaters selected by the swarm to the top four sweaters selected by the survey. We find that the total sales generated by the swarm-based selections was $3.1 million (90.1% of maximum possible) compared to only $2.5 million (73% of maximum possible) for the survey-based picks, as illustrated in figure 3.

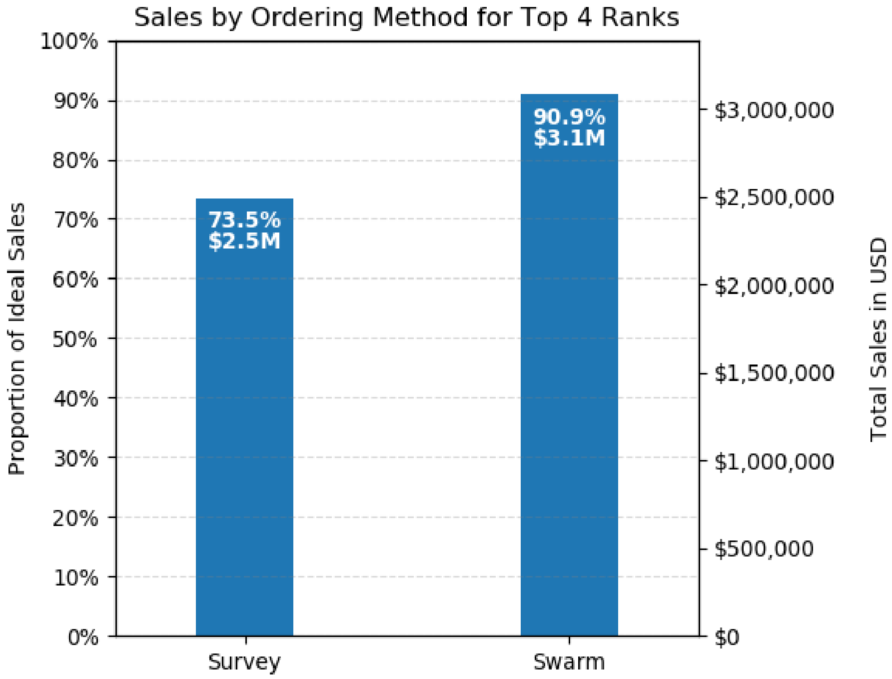

Fig. 3. Total sales (USD) and proportion of ideal sales for the top four sweaters selected from each forecasting method

To more directly compare the quality of overall rankings between the swarm and the survey, the sum of the unit sales at each cutoff was calculated as the area under the curve (AUC) in Figure 2. In fact, the swarm achieves an AUC of 94.4% of the best possible cumulative unit sales, while the survey achieves an AUC of only 81.0%. A histogram of the bootstrapped survey vs. swarm proportion of ideal sales AUC is shown in figure 4: we find that the swarm forecasts significantly outperform the survey forecasts using this aggregate metric (p<0.001).

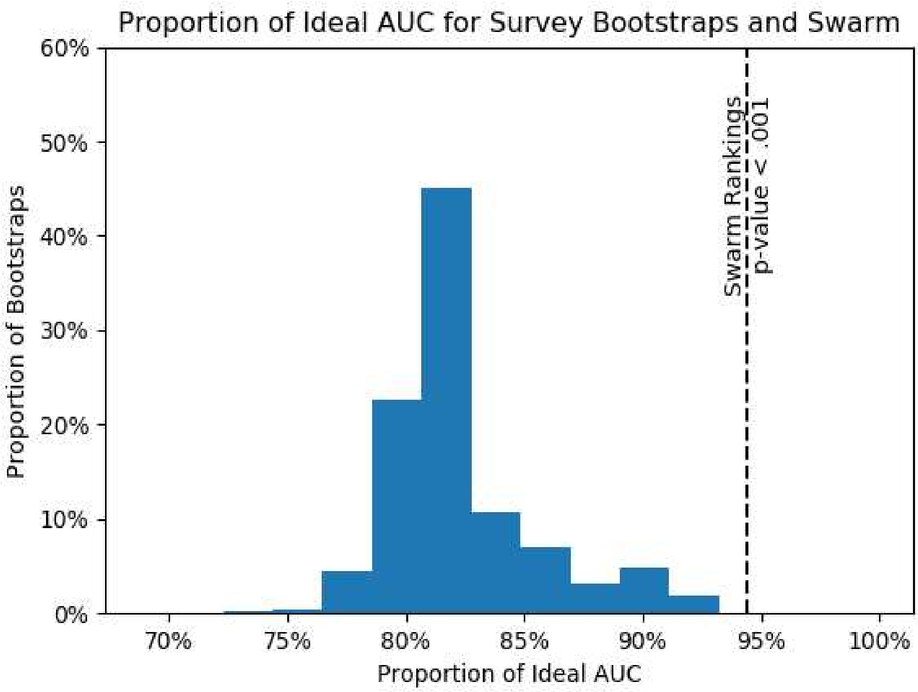

Fig. 4. Cumulative Unit Sales of All Ranks from Survey Bootstraps and the Swarm

5 Conclusions

Sales forecasting of new products is inherently challenging: by definition there's little historical data on the market perception of the product. The inaccuracy in contemporary sales forecasts is often in excess of 42% of true unit sales, which significantly impairs the efficiency of business processes, and suggests that new methods are needed to improve the accuracy of sales forecasts.

In this study, we introduced Artificial Swarm Intelligence systems as a tool for sales forecasting using consumers, and compared the forecasting accuracy of consumer surveys, a widely used sales forecasting tool, to the forecasting accuracy of consumer swarms. Each method forecasted the sales of eight new sweaters designed by a large fashion retailer for the 2018 holiday season. The ASI systems forecast the sales rankings of the sweaters significantly more accurately than the survey ($p<0.001$), so we can be confident that the group performed better when swarming than when polled due to more than random chance alone.

While the results from this study are very promising, the exploration was limited to forecasting the ranked sales performance of a relatively small set of products with a relatively small set of human participants. In future studies, it would be valuable to expand the number of participants and items considered. It would also be valuable to test the use of ASI technology for sales forecasting across a variety of different regions, time-frames, and channels.

Acknowledgements

Thanks to Bustle Digital Group for supporting this project by sourcing participants and coordinating with the retail partner. Also, thanks to Unanimous AI for the use of the swarm.ai platform for this ongoing work.

References

1. Kahn, Kenneth B. "Solving the problems of new product forecasting." Business Horizons 57 (5) 607-615.
2. Lynn, Gary S., Schnaars, Steven P., Skov, Richard B. "A survey of New Product Forecasting Practices in Industrial High Technology and ow Technology Businesses." Industrial Marketing Management 28 (6) 565-571.
3. Kahn, Kenneth B. "An exploratory Investigation of new product forecasting practices" The Journal of Product Innovation Management 19 (2002) 133—143
4. Rosenberg, Louis, "Human Swarms, a real-time method for collective intelligence." Proceedings of the European Conference on Artificial Life 2015, pp. 658-659
5. Rosenberg, Louis. "Artificial Swarm Intelligence vs Human Experts," Neural Networks (IJCNN), 2016 International Joint Conference on. IEEE.
6. Rosenberg, Louis. Baltaxe, David and Pescetelli, Nicollo. "Crowds vs Swarms, a Comparison of Intelligence," IEEE 2016 Swarm/Human Blended Intelligence (SHBI), Cleveland, OH, 2016.
7. Baltaxe, David, Rosenberg, Louis and N. Pescetelli, "Amplifying Prediction Accuracy using Human Swarms", Collective Intelligence 2017. New York, NY ; 2017.
8. McHale, Ian. Sports Analytics Machine (SAM) as reported by BBC: http://blogs.salford.ac.uk/business-school/sports-analytics-machine/
9. Rosenberg, Louis, et al. "Artificial Swarm Intelligence employed to Amplify Diagnostic Accuracy in Radiology" IEMCON 2018, Vancouver, CA; 2018
10. Bonabeau, E. Decisions 2.0: The power of collective intelligence. MIT Sloan management review, 50, 2 (2009), 45.
11. Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N. and Malone, T. W. Evidence for a collective intelligence factor in the performance of human groups. science, 330, 6004 (2010), 686-688.
12. Surowiecki, J. The wisdom of crowds. Anchor, 2005.
13. Seeley T.D, Buhrman S.C 2001 "Nest-site selection in honey bees: how well do swarms implement the 'best-of-N' decision rule?" Behav. Ecol. Sociobiol. 49, 416–427
14. Marshall, James. Bogacz, Rafal. Dornhaus, Anna. Planqué, Robert. Kovacs, Tim. Franks, Nigel. "On optimal decision-making in brains and social insect colonies." Soc. Interface 2009.
15. Seeley, Thomas D., et al. "Stop signals provide cross inhibition in collective decision-making by honeybee swarms." Science 335.6064 (2012): 108-111.
16. Seeley, Thomas D. Honeybee Democracy. Princeton Univ. Press, 2010.
17. Seeley, Thomas D., Visscher, P. Kirk. "Choosing a home: How the scouts in a honey bee swarm perceive the completion of their group decision making." Behavioural Ecology and Sociobiology 54 (5) 511-520.
18. Rosenberg, Louis, and Willcox, Gregg. "Artificial Swarms find Social Optima". 2018 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA 2018) – Boston, MA; 2018.