Panchiko

Dimitrije Radusinovic

Project Overview

The goal of this system is to support consumers in finding positively reviewed products that closely match their needs and preferences across a range of criteria like budget, comfort, performance, etc. The system uses human crowdsourcing to get as much information, data and reviews as possible form many different sources which will remove bias from branded marketing materials.

The workflow starts when a consumer send a request that outlines their specifications for the product that they are looking for. Crowd workers then search online stores and review sites to identify potential matching products, compiling links and basic specs into a shared database. Additional crowd tasks extract key product features, aggregate reviews, and annotate the dataset along important dimensions like cost, quality, comfort. Finally for the end we will have group of three people that will go through the data collected and vote what would be best product for consumer.

In contrast to opaque algorithms customized to business interests, the system can surface superior recommendations personalized to each individual by harnessing the crowd to obtain unbiased data from genuine users. Iterative product information and review collection also allows the system's performance to improve over time.

Research Question/Contribution

Key research questions this system can address:

- Can crowd workers collect, aggregate, and objectively present information to meet a requester's product needs? How accurately do crowd-generated recommendations satisfy consumer criteria versus those from existing tools?

- How many crowd iterations are needed to produce high-fidelity recommendations in categories with rich online data versus those with sparse data? Does starting with a small seed dataset enable the crowd to hone in on relevant products quickly?

- Do requester preferences and crowd annotations capture all necessary product aspects to make the "right" recommendation for an individual? What key data dimensions for different product categories yield the best predictions?

Related Work

Crowdsourcing complex tasks using micro labor markets is an emerging area of research. Early work by Kittur et al. [1] demonstrated using Mechanical Turk for remote user studies but focused on individual tasks. Little et al. [2] introduced TurKit for managing iterative tasks flows. Sorokin & Forsyth [3] used MTurk for tagging images in computer vision.

Our work builds on these by providing a general conceptual framework and toolkit for breaking down and coordinating complex, interdependent crowdsourcing tasks. We demonstrate applicability across a diverse range of domains including article writing, purchase decisions, and science journalism. The CrowdForge framework is novel in allowing multi-level, dynamic partitioning of work guided by the crowd itself.

Other related work has examined quality control methods in crowdsourcing. Sheng et al. [4] provided an early analysis of repeated labeling for quality assurance. More recently, Kamar & Horvitz [5] proposed modeling worker reliability for optimization. Our approach treats voting and verification as first-class tasks and shows benefits of merge-based quality control versus simple voting.

Stakeholder & User Information

The primary stakeholders for this crowdsourced product recommendation system are everyday consumers looking to find the best products for their needs. The system is designed to directly address common pain points consumers face today:

- Dealing with biased reviews and brand hype that make it hard to objectively evaluate products
- Fighting through inaccurate, manipulative recommendations from opaque algorithms optimized for seller profits over consumer interests
- Lacking the time and expertise to deeply research products across the exploding online landscape

By crowdsourcing honest reviews, criteria, and purchasing data from real users, the system gives consumers the power to cut through the noise. Personalized recommendations are tailored to each individual's stated budget, performance needs, aesthetic preferences, ethical values, and other dimensions supplied explicitly or inferred from past purchase data. The iterative collecting of reviews and refinement of criteria also enables recommendations to continuously improve over time as more collective intelligence gets incorporated.

On the supply side, workers that work for our system are critical stakeholders that make this system possible because without them this system would be impossible. By taking on simple microtasks like finding products, extracting key specs, annotating features, and consolidating reviews, workers collectively construct the rich dataset that feeds the recommendations. The participation model is designed to be flexible and accessible, allowing anyone to contribute a little or a lot towards this public good. Workers also receive fair monetary compensation for their efforts commensurate with task complexity.

At the end we can also say that the companies and sellers themselves that sells products that are on the platform also benefit from increased consumer trust and reduced churn thanks to transparency. While their interests are not center stage, ethical companies focused on delivering genuine value have an aligned incentive to provide the unfiltered data that enables great recommendations fitting consumer needs.

Researchers in fields like human computation, recommender systems, and collaborative filtering also have a valuable stake in the system's development. The unique crowdsourcing workflow provides a rich testbed to study questions around optimizing collaborative human systems, eliciting accurate crowd data, evaluating real-world recommendation performance, and analyzing effects of transparency on user trust.Insights from this human-centered recommendation approach could inform best practices for future platforms across many domains pursuing transparent, ethical recommendations through distributed collaboration.

Dataset

Procedures

Construction a good and big data is very important for powering personalized and relevant product recommendations. The dataset is assembled through a coordinated workflow across three recurring crowd task phases:

Product Collection

The collection phase focuses on rapidly logging breadth - workers continuously search major online stores and marketplaces to identify potential product options that match categories and features requested by consumers. Links, prices, specs, and descriptions are submitted for each product identified.The main goal of creating this dataset is to provide a good report across a wide range of product domains online.Strict policies prohibit any fabricated or copied data.

Product Annotation

As the dataset grows, additional annotation tasks enrich the products with supplemental attributes. Workers analyze pricing tiers, assigning entries to low, medium or high cost segments. Products are also given 1-5 star quality ratings based on reviews of durability, defect rates, and performance factors. Detailed feature tagging covers dimensions like comfort, ergonomics, materials, and technical specs while meeting regulatory and compliance standards. Ethics annotations trace sourcing and production policies regarding labor conditions and environmental impact. Workers can append any new descriptive categories appropriate for products lacking specifications.

Review Aggregation

The final phase aggregates opinions and feedback data around products. By searching many online customer review platforms, crowsoucers can summarize perspectives and recommendation that will help the system to make a right decision. Where review systems with quantitative ratings or tallyable metrics exist, statistics around score distributions are compiled. The goal is to enrich the dataset with community experiences - both positive and negative.

This three stage iterative process alternates between the collection, annotation, and aggregation tasks in parallel based on need. As new requests enter the system, uncovered product areas trigger more collection. Gaps in attribute detail prompt annotation surges. And availability of new review data sparks additional aggregation. Together it leads to systematic growth in the scope and depth of descriptive data coverage. Advanced algorithms then match products to consumers given their preferences.

Instructions for Workers

As a data labeler, you serve the critical role of enriching the recommendation dataset by analyzing existing entries and appending multi-dimensional attribute tags. The clarity and consistency of these crowd-sourced annotations enables personalized filtering and matching.

Price Segment

Classify the current market listing price as low, medium or high tier based on comparisons with competitive product offerings. Factor in discounts, sales, coupons that may offer consumers better effective pricing. Rationalize borderline cases with notes.

Quality Rating

Assign a 1-5 star rating evaluating cumulative factors like durability, reliability, defect rates, and overall performance excellence gleaned from customer reviews. Granular aspects like fitness for purpose may override coarse aggregations.

Comfort & Ergonomics

Tag specific features related to comfort and ergonomics like breathability, adjustability, strain reduction, accessibility accommodations, intuitive interfaces, physical ease of operation and servicing. Segment further by use contexts where appropriate.

Technical Specifications

Identify technical specifications like resolution, battery life, accuracy, connectivity protocols, power efficiency and other niche metrics useful for matching user requirements. Expand abbreviations, clarify comparisons.

Ethical Sourcing

Document available policies, certifications or initiatives related to ethical sourcing, labor conditions, environmental impact, social responsibility. Note tensions between aspirations and accountability.

Additional Descriptors

Append any other descriptive categories useful for capturing nuanced product differentiators lost in structured data. Preserve native segmentation schemas where consensus exists.

Aim to provide 9 or more distinct attributes across these dimensions to maximize multidimensional coverage unique to each product. Leverage all available details, customer perspectives and reviews to unlock wisdom in data.

Implementation Workflow & Diagram

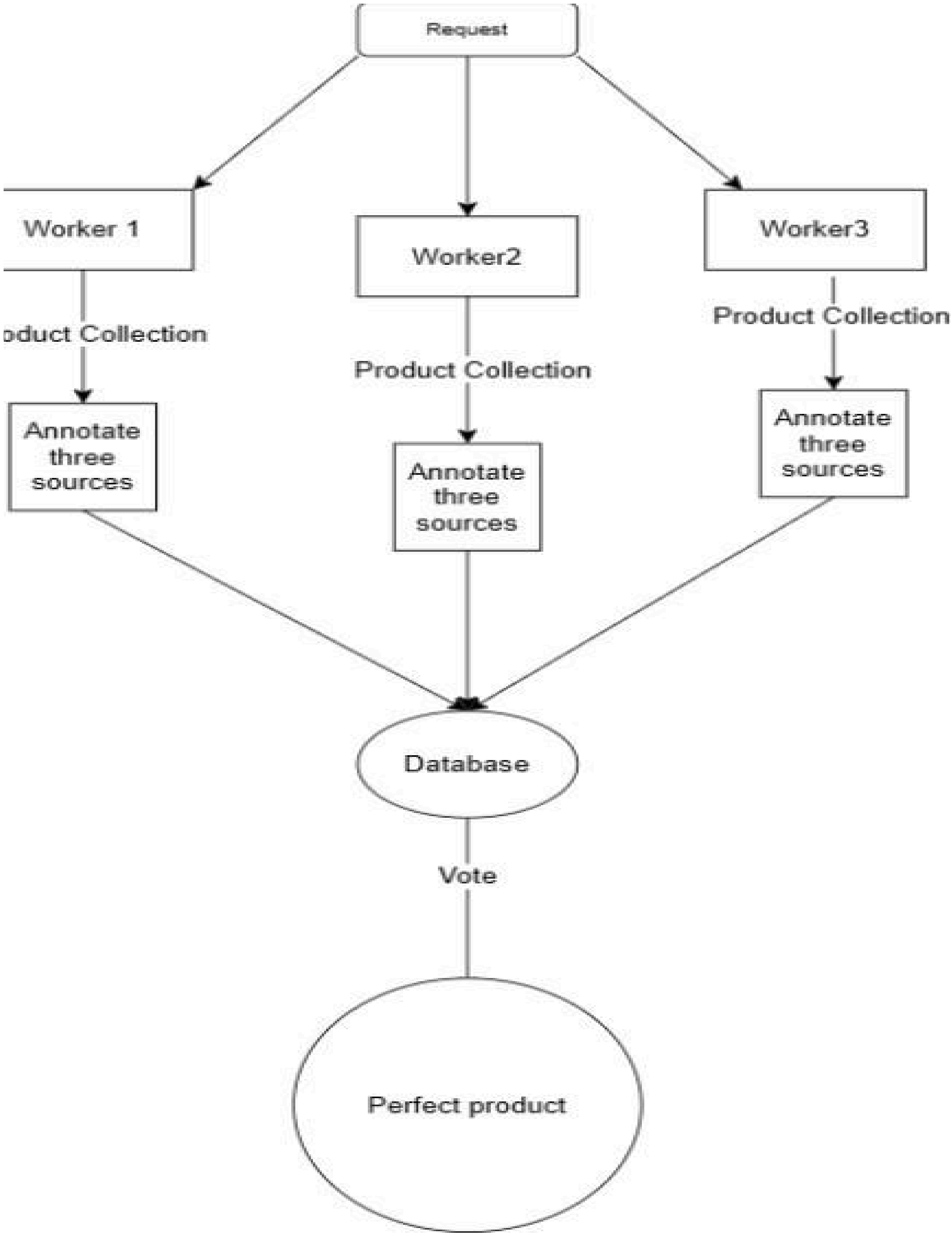
The overall crowdsourced recommendation workflow starts with consumers submitting a request detailing their ideal product requirements and preferences. This request enters into a database which triggers a workflow manager module.

Product Collection: Crowd workers search stores/reviews to submit links, categorization, annotations for requested and related products

Dataset Enhancement: Collected data is organized and augmented with worker insights to mitigate bias and emphasize real user perspectives

Recommendation Generation: The final output returns the most relevant products matching the user's needs based on the enhanced dataset

Ongoing human contributions improve recommendations over time as new products/reviews accumulate in the database. Analytics track crowd patterns and recommendation performance to optimize the workflow.


Request
Worker 1
Worker2
Worker3
oduct Collection
Product Collection
Product Collection
Annotate three sources
Annotate three sources
Annotate three sources
Database
Vote
Perfect product

Human-Based Elements

There are three key procedures in the crowdsourced recommendation workflow that leverage human intelligence that makes this system unique:

Product Collection

Workers use their judgment and domain knowledge to identify relevant products matching consumer requests. By searching across retailers and markets online, they can discover fitting options that automated scraping approaches may miss or struggle to parse due to complex specifications. Humans also discern nuances in product slotting across pricing tiers. Determining matches relies on perceptual skills and personal opinion.

Product Annotation

Humans are great at analyzing products to tag salient attributes like pricing segments, technical features, use case tradeoffs, and ethical sourcing considerations. While simple characteristics are programmatically extractable, workers quantify complex quality dimensions like comfort, ergonomics, and manufacturing precision using their reasoning skills. Furthermore, leveraging collective intelligence around appropriateness of categorization schemes sidesteps the need for top-down centralized knowledge engineering.

Review Aggregation

Summarizing opinions and experiences from customer reviews into consensus perspectives leverages humans' innate language capabilities. Identifying common use cases, crystallizing pros and cons, and destilling recommendations requires interpreting semantics beyond key words. The open-ended nature of synthesizing subjective text to find salient patterns remains challenging for algorithms without strong ground truth guidance from humannotated training corpora.

Instructions for Workers

Workers are essential partners in building an unbiased, high-quality dataset for personalized product recommendations. Core responsibilities across three recurring task areas require applying human judgment:

Product Collection

Leverage your shopping experience and domain familiarity to identify relevant products matching categories that consumers request. Search major retailers and specialty markets for potential options, noting products that algorithms may struggle to parse or overlook. Submit useful links, price data, specifications, and descriptions to help seed the dataset's coverage. While comprehensive inclusion is the goal, use your discretion to contribute products that balance both meeting needs and expanding breadth.

Product Annotation

Drawing on your reasoning skills, annotate the nuanced attributes of products already in the dataset along key dimensions like pricing tiers, technical features, use case tradeoffs, and ethical sourcing considerations. Quantify complex quality perceptions around comfort, ergonomics, precision engineering based on the collective intelligence of the crowd. Categorization schemes are expected to evolve based on your insights into salient facets. Flag any concerning biased assumptions or business conflicts of interest around product portrayals.

Review Aggregation

Synthesize customer opinions and experiences into coherent consensus perspectives. Summarize common use cases, crystallize notable pros and cons, and destill cogent recommendations that incorporate both praise and criticism. Interpret the semantics and contexts behind subjective text to identify meaningful patterns. The open-ended nature of this aggregation benefits tremendously from human language capabilities as a precursor to any automated content quantification.

The goal across all tasks is to accurately reflect available public data from diverse viewpoints without injecting personal opinions or biases. By coordinating the crowd's contributions, we can collectively build an inclusive dataset for powering personalized product recommendations.

Care for Crowd Workers

We aim to provide a supportive, ethical environment for all crowd workers powering the multi-stage recommendation workflow - spanning initial data collectors, labelers, reviewers, consolidators, and evaluators.

Equitable Compensation

Every task priced exceeds minimum wage matching median projected effort, scaling up for complexity. Bonuses further highlight uniquely insightful work independent of speed. We acknowledge skilled human cognition effort enables this system unlike fully automated approaches.

Redress Against Unfair Rejection

Workers can appeal disputed rejections with transparent explanations backed by evidence before independent internal review. Our team promptly examines contested cases during business hours sharing rationales while swiftly correcting rare mistakes.

Flexible Participation Options

Workers can contribute in concentrated blocks or split shifts fitting their personal schedules without punitive throughput expectations. Reasonable maximum duration guidance includes frequent screen break reminders to sustain cognitive focus and prevent fatigue.

Accessible Task Interfaces

Ongoing surveys continually refine interfaces emphasizing usability, accessibility to minimize unnecessary exclusion stemming from tools rather than skills. All perspectives expand collective wisdom so please suggest any areas for universal improvement.

Psychological Safety Culture

No single worker output constitutes absolute truth. We encourage respectfully challenging assumptions, enhancing context with multiple stakeholders. Progress flows from good faith attempts without fear of retaliation for voicing unconventional positions while identifying blindspots.

Software-Based Elements

Seamlessly integrating human intelligence into scalable workflows requires robust software capabilities including:

Workflow Manager

An orchestration module is the backbone, analyzing new requests, evaluating current dataset gaps, and optimally determining next crowd tasks to produce recommendations. Algorithmically routing jobs to balance supply and demand enables smoothly scaling human effort to match system needs.

Database Management

Effectively storing and processing the accumulated product options, attributes, and reviews collected by the crowd relies on a versatile database. Flexible organization and advanced search facilitate workers finding and enhancing relevant existing data in a growing knowledge repository.

User Portal

A sleek consumer-facing web portal handles intake of product requirements and display of personalized recommendations. The collection and formatting of details by humans leads to a truly natural interface for conveying needs and interacting with audience-driven results.

Crowd Integration

Embeddable crowd software lets the workflow tap into on-demand, scalable contributors from managed platforms like Amazon Mechanical Turk or Prolific. Targeted microtasks collect, evaluate, enrich dataset entries. API integration supports programmatic quality control and compensation disbursal as core incentives.

Deployment Plan

Executing this crowdsourced recommendation workflow requires a measured rollout approach:

Start Small

The system is initially tested in a limited beta phase with example recommendation requests across a few representative product categories. Before a larger rollout, this focuses on perfecting the fundamental crowd tasks, price structures, quality assurance systems, and cloud infrastructure provisions.

Scale Controlled

After hardening the foundations, the service formally launches by incrementally opening access and throttling volume rather than flooding immediately. Additional product domains unlock over time as operations processes mature. New task variants also phase based on need.

Optimize Targeted

With core flows stable, optimization focuses on maximizing value specifically for key user segments through enhanced personalization features and subsidized ethically sourced products. Analytics help balance automation with human effort at each layer.

Expand Judiciously

Budgets responsibly scale to support expansion but avoid runaway spending that jeopardizes ethical standards. Impact dashboards track holistic welfare alongside business metrics to guide judicious growth. Partnerships around aligned values help sustainably spread benefits.

Expected Output & Evaluation

Defining metrics to assess system performance is imperative for improvement. Key quantitative and qualitative measures help evaluate outputs.

Recommendation Relevance

Accuracy of crowdsourced product suggestions versus stated consumer criteria determines core utility. Rating scales quantify requester satisfaction.After a certain time, a good and positive rating creates reliability among customers, which leads to a larger number of customers.

Dataset Quality

Completeness of cataloged options, detailed meta annotations, salient review summaries and ethical sourcing tags proxy the information depth enabling strong matches. The collected data creates a big picture and gives us a better view of the overall statistics. These statistics are meant to help optimize the entire human effort of information gathering.

Crowd Contributions

Volume of product submissions, attribute annotations, and peer reviews demonstrate raw platform engagement. But labeling speed, inter-rater consensus, and explanation coherence better reflect work quality.

Responsible Returns

As recommendations convert to purchases, both consumer delight and company revenue expand.Of course, the concentration on the business itself of this system could lead to a decrease in social impact, which is not the goal of this system.Holistic scoring incorporates environmental sustainability and just employment indices to prevent dangerously narrow definitions of success.

Evaluating across metrics that capture the full spectrum of relevant outcomes allows continuously aligning system improvements with ethical values. This is of great importance in our system because it has the job of helping the improvement but also the innovation of this system.

Assumptions & Trade Offs

The hybrid human-computation workflow makes a few leaps of faith while recognizing necessary tradeoffs.

Enabling Assumptions

We assume the crowd possesses collective intelligence to gather, categorize, and annotate product data that meets individual consumer needs better than fully automated approaches. Additional iteration cycles based on user feedback are expected to incrementally improve recommendations. Decomposing tasks modularly is projected to enable scalable coordination without fully compromising end-to-end continuity.

Tension Tradeoffs

Pursing personalized accuracy for every recommendation needs balancing with practical turnaround times expected by consumers. There exists natural tension between customizing each suggestion and maintaining wide catalog breadth to satisfy diverse interests. Furthermore, modular task decomposition promotes distributed processing but adequate dependencies must propagate across decentralized work.

Alternatives Considered

Manual In-House Curation

Delegating all product evaluations and recommendations to specialized in-house staff allows consistency but proves challenging to scale across exponential niche varieties given the long-tail of consumer interests. Even experts have intrinsic biases.

Fully Automated Pipeline

Inferring subjective qualities like comfort, aesthetics and ethical sourcing solely algorithmically continues to falter in generalizing beyond detection of statistical patterns within limited training data. Lacking human judgment,which is the point of this system, also undermines trust for users.

Tailored Crowd Platform

Procuring a fully managed end-to-end crowd solution enables quality control through workflow guardrails but often locks into proprietary vendor pricing models that squeeze margins. Maintaining a perfectly curated roster introduces non-trivial overhead.

Must Have, Should Have, Would Be Nice to Have

Must Have Pillars

The multi-phase crowdsourcing workflow provides core coordination enabling any functionality. Shared data storage interconnects decentralized judgments. A consumer request portal allows intake specification. Without these human-driven components, no recommendations emerge.

Should Have Curators

Crowd workers filter and nominate options, replacing automated matching. Peer reviews on selected products provide credibility checks. Output explanations support transparency on provenance. Lacking these oversight roles forfeits customization, reliability and trust.

Would Be Nice Extensions

Enrichment flows around impact tracking and periodic reporting offer peripheral value via oversight. Mobile tasking improves participant access convenience but offline requests can still operate. These embellish an already functional human-centered system.

Data Speculation

	A	B	C	D
1	Product Links	Changelog	Notes	Price
2	JBL Tune 510BT Headphones, JBL Pure Bass Sound | Verizon	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
3	JBL Tune 710BT Wireless Headphones, Bluetooth Streaming | Shop Today (verizon.com)	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
4	Amazon.com: Bluetooth Headphones Over Ear,BERIBES 65H Playtime and 6 EQ Music Moc	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
5	Amazon.com: KVIDIO [Updated] Bluetooth Headphones Over Ear, 65 Hours Playtime Wireles	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
6	JBL Tune 510BT Wireless Bluetooth On-ear Headphones | eBay	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
7	Bluetooth Earbuds TWS 5.3 all phones laptop tablet Wireless Earphone Waterproof | eBay	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
8	Harman AKG Y500 Wireless Bluetooth Foldable On Ear Headphones Black Headset New 28	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
9	Beats by Dr. Dre Solo3 Wireless Club Collection Headphones, Brand New and Sealed | eBay	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
10	Beats by Dr. Dre Studio Buds+ Noise-Canceling True Wireless In-Ear Headphones - 1942535	medium-low price, comfort, Bluetooth	Request ID:01	Medium-Low
11				

Data Description

We received a product request from a user with ID 01. The requester was looking for medium-low priced headphones between $0-100, that are comfortable to wear and have Bluetooth connectivity.

Interpreting the Data

Reviewing the dataset, we see various headphone options that match the requested criteria to different extents. Most of the products fall within the desired price range and support Bluetooth. Multiple entries also mention comfort as a highlighted feature.

However, there appear to be two duplicate links referring to the same JBL Tune 510BT headphones. To maximize unique options shown to the requester, one of these duplicate links should be replaced with a different headphone product not already captured in the current list.

Impact & Next Steps of Different Data Scenarios

If fewer total products are found matching the necessary criteria like Bluetooth and wired capability, we may need to relax aspects like price range to increase options. Priority could be given to criteria explicitly mentioned as mandatory by the requester when filtering.

In cases with duplicate links, replacing duplicates with alternatives increases unique choice count. However, occasional benign duplication may organically reinforce particularly strong matches. Duplication thresholds could trigger automated consolidation.

For inconsistent or sparse annotation along dimensions like comfort and sound quality, crowd workers could be allocated to fill in missing ratings. This helps enable multi-criteria sorting and filtering for product matching. Lacking annotations altogether diffs against a baseline of requester guidance only.

In all cases, additional requester interaction to clarify needs, revise criteria weights, or provide feedback on results would produce further refinement. Dialog tightens the human-centered alignment loop.

Conclusion

Our crowdsourced system demonstrates the potential for collaborative human intelligence and opinions to deliver personalized, unbiased, and ethical suggestions matching individual consumer needs. The synergistic coupling of target crowd effort with scalable software infrastructure provides a pathway to overcome current limitations around overly automated recommendations propagated by opaque algorithms.

When we combine human judgment, perception, inference and linguistic interpretation through iterative microtasks, our system can then create a real picture of the product.Distributing discrete contributions across a diverse crowd also promotes increased access and agency for participatory work while directing skills where most impactful. However, thoughtfully nurturing both innovation and inclusion simultaneously remains an enduring commitment requiring continuous re-centering around human welfare alongside business returns. The long-term aim is advancing recommender systems from opaque manipulation of consumers towards more empowering tools focused on responsibly elevating people - both end users and crowd workers alike.

References

[1] Kittur, A., Chi, E., Suh, B. Crowdsourcing User Studies With Mechanical Turk. In Proceedings of CHI (2008).

[2] Little, G., Chilton, L.B., Goldman, M., and Miller, R.C. Turkit: human computation algorithms on mechanical turk. In Proceedings of UIST (2010).

[3] Sorokin, A., Forsyth, D. Utility data annotation with Amazon Mechanical Turk, First IEEE Workshop on Internet Vision at CVPR (2008).

[4] Sheng, V.S., Provost, F., Ipeirotis, P.G. Get another label? improving data quality and data mining using multiple, noisy labelers. In KDD (2008).

[5] Kamar, E., Horvitz, E. Incentives for truthful reporting in crowdsourcing. In AAMAS (2012).

[6] Request Workflow Diagram. Draw.io diagram created by Dimitrije Radusinovic on 12/15/2023.