Recommendation Systems
An Overview of System Types and Benefits – and Why You Should Implement With SAS®

Insights from a webcast

Featuring
Wayne Thompson, PhD, Manager of Data Sciences Technologies at SAS.
Contents

What Types of Recommendation Systems Exist Today? ................................................................. 1
How Are Recommendation Systems Currently Being Used? ..................................................... 1
How Are Businesses and Customers Benefiting? .... 2
How Do Recommendation Systems Actually Work? .................................................................. 2
  A Closer Look: SAS® In-Memory Statistics ................................................. 2
  Sophisticated, Back-End Processing ............................................................ 3
  A Real-World Scenario ............................................................................... 4
What Are the Advantages of SAS® Technology? ...................................................................... 4
  Support for Analytics on the Go ............................................................... 4
  Lightning-Fast, Interactive Collaboration ................................................. 4
  Support for a Wide Variety of Analytic Techniques .......... 4
Learn More ................................................................................................................. 5
If you’ve ever used Amazon, Pandora or Netflix, you’ve experienced the value of recommendation systems firsthand. These sophisticated systems analyze historical buying behavior and make recommendations to buyers in real time while they are shopping for a specific product. By supporting an automated cross-selling approach, they empower retailers to offer additional products or services that enhance the products already selected by customers. The emphasis here is on adding value – because if the customer sees value in the recommended items, it’s estimated that this can lead to a 5 to 20 percent increase in sales.

In many ways, it’s like magic – for both customers and marketers. Customers get personalized recommendations on additional products that are relevant, valued and helpful, and marketers can enhance offers in ways that proactively build better customer relationships, retention and sales.

But how do they actually work? And how are these systems being used in different contexts today?

In a recent webcast, Wayne Thompson, PhD, Manager of Data Sciences Technologies at SAS, explored recommendation systems in more depth. This paper summarizes his answers to common questions about the types of recommendation systems used today, who uses them and why, and the advantages of using them. He also delves into how the underlying technology works, with a special focus on SAS capabilities and benefits.

What Types of Recommendation Systems Exist Today?

According to Thompson, there are several types of recommendation systems:

- **Content-based**, which recommend items based on what customers looked at or purchased in the past.
- **Community-based**, which recommend items that similar customers (based on common preferences and tastes) have liked and purchased in the past.
- **Explicit**, which allow companies to use explicit ratings to build recommendation systems, like Yelp. These systems ask customers to assign a number of stars, for example, to rate something. These ratings are explicit scores that are used to build models.
- **Implicit**, which are built from data about items that were purchased together in the past without knowing explicitly how a customer rated the items.

Using these systems requires more than just building models. You also have to:

- Prepare data.
- Collect information about users, items and profiles.
- Aggregate and filter historical data.
- Evaluate and develop champion-challenger types of models.
- Deploy models in production systems so that consumers receive personalized recommendations automatically and in real time.
- Look at model decay and monitor and update models accordingly.

How Are Recommendation Systems Currently Being Used?

Recommendation systems have been deployed across a wide range of industries and contexts, especially as part of online shopping sites, explained Thompson. You’ll typically see them used in:

- **Supermarkets**: Registers that generate custom coupons for next purchases look at prior purchases and tailor coupons to include items you’d likely be interested in.
- **Book and music stores**: Companies can send customers emails following a purchase and recommend new books or albums – or in the case of Amazon, Pandora and similar sites, provide product recommendations in real time based on what a customer is currently looking at or listening to (or has in the recent past).
- **Investment firms**: Recommendation systems can analyze which stocks you would likely be interested in based on what similar customers have chosen (see sidebar, “Viseca Uses SAS to Optimize Recommendations”).
- **TV and movie services**: Companies like Netflix analyze each customer’s prior content choices and make recommendations based on them, as well as proactively recommend items in real time based on their browsing history.
- **Social network sites**: Sites like LinkedIn and Facebook use recommendation systems to suggest additional connections or friends based on a person’s existing network.
How Are Businesses and Customers Benefiting?

Recommendation systems are profitable investments for companies, Thompson said. For example, organizations typically realize:

- Stronger customer relationships by providing additional and often unique personalized service.
- Increased trust and customer loyalty.
- Higher sales and profitability.
- Higher click-through and conversion rates.
- New opportunities for promotion and persuasion.
- Deeper knowledge about customers.

But they also provide significant value for customers by helping them:

- Find things that are interesting or useful.
- Narrow down a set of choices.
- Explore options.
- Discover new things.
- And more.

How Do Recommendation Systems Actually Work?

To determine optimal recommendations for individuals or groups, you need analytics that can solve a tough computational and processing challenge, as well as crunch tons of data in real time.

A Closer Look: SAS® In-Memory Statistics

To better understand the challenges of analyzing big data in real time – and how recommendation systems overcome them – Thompson discussed an example involving SAS In-Memory Statistics. This application includes a recommendation system that can generate personalized, meaningful recommendations in real time using data stored in Hadoop – and with a high level of customization. “We’re focused on being the No. 1 analytical computation platform within Hadoop,” noted Thompson. “Many of our customers are moving to Hadoop because it lets them store their big data and process it both in parallel and in a distributed way.”

In addition, SAS In-Memory Statistics supports:

- **Interactive programming** so multiple users can concurrently and interactively analyze large amounts of data stored.
- **In-memory analytical processing** for fast analytic computations that are optimized for multiple passes across a distributed cluster.
- **Persistent data in-memory**, which provides speed and reduces latency.
- **Analytical data preparation**, which enables data access and manipulation, allowing you to transform and create variables and perform exploratory analysis.
- **Model development** so you can quickly create, evaluate and compare multiple statistical models.
- **Statistical algorithms and machine-learning techniques** for uncovering patterns and trends faster than ever before with a huge breadth and depth of analytical techniques.
- **Text analytics**, allowing you to analyze your unstructured (and structured) data using a wide range of text analysis techniques.

“All of these capabilities run on a market-viable computing platform that integrates seamlessly with third-party solutions from HDFS, Pivotal Greenplum and Teradata, among others,” added Thompson.
Sophisticated, Back-End Processing

“When all of your big data is stored in one place, it’s not feasible to bring it to your analytics software,” explained Thompson. “You have to bring your analytics to your data.”

As an example, imagine that all of your data is stored in Hadoop, and you have one head node (also referred to as a “general”) and four data nodes in your Hadoop file system. The computations for determining recommendations are divided up among the data nodes, with the best-user item combinations broadcasted back to the head node for decision making.

“Once you load data into memory within the Hadoop cluster, SAS software takes over,” Thompson said. “In this example, assume there’s an application sitting outside the cluster called Edge Node – it’s what communicates with the cluster, sending instructions to do something like ‘build a decision tree’ or ‘build a recommendation model.’ Because your data is already harnessed into memory, you can divide and conquer and compute in parallel with your recommendation models, and then return the results set to the SAS® LASR™ Analytic Server node sitting alongside or totally collocated inside the cluster.” This edge client just reads results sets – it’s not where computations are done. Computations are done in the cluster with the data nodes.

“It’s worth noting that the way SAS distributes processing within a Hadoop cluster is much more sophisticated than apps like MapReduce,” commented Thompson. “I can use a web client for SAS In-Memory Statistics – SAS Studio – to write recommendation code from anywhere. It’s not hard to do using PROC RECOMMEND, which is a procedure in SAS In-Memory Statistics. You simply add the data table that has your users, ratings and items. To build a model, you choose from different methods provided by the software – for instance, KNN or K nearest neighbor. In this case, I want to do a Pearson’s correlation as the similarity matrix.”

Once your code is written, explained Thompson, you submit it through the middle tier over to the cluster, which waits to receive processing requests. At this point, each of

Machine Learning at Scale: HDFS + In-Memory

- No MapReduce
- One data copy
- Concurrency
- Temporary columns
- MPP or SMP
- Thin clients
- Multiuser
- Interactive
- Real-time
- Point and click or programming

Figure 1: Principles of the back-end design.
the underlying data nodes computes a partial correlation coefficient and does it in a partitioned manner across each of the four data nodes. “It’s like having four people help with a certain task. If they work together to complete it, they can get done nearly four times as fast as if just one person did it,” explained Thompson. The individual coefficients then get broadcasted back to the head node for assembly. The result is a correlation coefficient that tells you how closely correlated the items you selected are relative to other items in the data set. This information then gets sent back to the web client. Figure 1 illustrates some of the basic principles of the back-end design of the SAS recommendation system, as just described.

A Real-World Scenario
To illustrate the power of SAS In-Memory Statistics, Thompson discussed a typical, real-world scenario: a company that needs to create a recommendation using book data. In this example, a data scientist has data that includes 278,000 users, 271,000 books and 1 million book ratings. “Let’s assume that she wants to develop a model and do a prediction using SAS In-Memory Statistics,” explained Thompson. “Using the software, she can quickly interact with the data, get distinct counts and calculate the average book rating – which was only 2.8 on a scale of 0 to 10. Next, she can develop a recommendation model using SVD, or the single value decomposition of a user-item ratings matrix, and run it on a distributed cluster with 16 data nodes.”

Given the size of the data set, this is a computationally intensive algorithm that could take hours or days to process using traditional analytical software. “But it only takes a few minutes to run in the SAS architecture previously described,” noted Thompson.

The results from the SVD include model diagnostics that show how the model preformed, as well as details on specific product recommendations for users. “Once the champion model is determined, then you can use the model code within your production system to generate specific recommendations for users,” Thompson said.

What Are the Advantages of SAS® Technology?
Thompson used this example to summarize the primary advantages of using SAS software to build a recommendation system.

Support for Analytics on the Go
“The principle of our design is enabling analytics on the go,” explained Thompson. “So you can access what you need using our thin client, called SAS Studio. Our software is easy to use, wherever you are – you just need a web browser.”

Lightning-Fast, Interactive Collaboration
Thompson noted that collaboration is essential to successful development of recommendation models. “That's why SAS allows multiple users to access the same data tables and work on them concurrently,” he added. “The new word for distributed is interactive. With SAS, people can build models very quickly, change settings, and see results sets as quickly as people can think. As a result, data scientists can work more efficiently and tackle more complex problems.”

Support for a Wide Variety of Analytic Techniques
SAS In-Memory Statistics provides both collaborative filtering and content-based approaches for use in recommender systems. In addition, the algorithms provided in SAS In-Memory Statistics support diverse approaches to developing recommender systems.

Examples of algorithms include:

- **K nearest neighbor**: A collaborative filter based on measures of association between items or users.
- **Cold starting**: A method for recommending typical products popular across your customer base to net-new users of your website.
- **Matrix factorization**: A way to create latent factors representing groups of items or families of items.
- **Association rules**: Rules for automatically recommending associated items to shoppers as they browse or place an item in their carts.
- **Clustering**: A way to reduce or decompress the number of users and items being analyzed when the user-by-item matrix is massive in size.
• **Slope One**: Algorithms used for collaborative filtering that include one of the simplest and yet highly accurate forms of nontrivial, item-based collaborative filtering based on ratings.

• **Ensemble methods**: A collection of various methods.

With SAS In-Memory Statistics, you can have different algorithms competing against each other in order to identify which works best for a specific business problem. You can also combine different techniques to determine the champion model.

**Learn More**

There are lots of recommender-building applications available today. What sets SAS technology apart is:

• **Precision** - because it provides an extensive set of in-memory machine learning techniques for developing recommendations in real time.

• **Flexibility** - because the recommender works with both rated and nonrated items.

• **Speed** - so you can quickly compose a list of best items for a customer or best customers for item.

• **Integration** - as recommendations can be integrated into all of your customer touchpoints.

Want to know more? Please visit:

• [sas.com/in-mem](http://sas.com/in-mem) to learn more about SAS In-Memory Statistics.

• [sas.com/consider-hadoop](http://sas.com/consider-hadoop) to read the TDWI Checklist Report Eight Considerations for Utilizing Big Data Analytics with Hadoop.

• [sas.com/hadoop](http://sas.com/hadoop) to learn about SAS and Hadoop.

• [sas.com/bigdatamatters](http://sas.com/bigdatamatters) to view the complete Big Data Matters webinar series.

You can also view the [webcast](http://sas.com/webcast) of Thompson’s presentation, which concludes with a demo of SAS In-Memory Statistics software being applied to a recommendation use case.