REAL TIME RECOMMENDERS IN E-COMMERCE.

“When building a recommendation system, there’s no silver bullet. You’ll always do better if you can factor in domain-specific knowledge.”
Index

Shouldn’t online information be timely and relevant in this age? 02

Real-time recommenders that respond to individual behavior 04
The right offer, but unfortunately, out of context 04
Getting started with recommenders 04
  Pointing out use cases 04
  Collecting data 05
  Testing use cases 05
Technology should not be the limiting factor 05
  Integrating recommenders in existing infrastructure 05
Best-of-breed platform 06

Success stories 07
Transavia: Personal recommenders to improve online user experience 07
  Better service by collecting customer information 08
  The increased importance of the data team 08
NPO – Tuning to different viewing behavior by offering relevancy and transparency 09
  The right process for innovation 09
Setting up the data platform and developing use cases 10
Understanding behavior through dashboards 10
Personal recommendations 10
Transparent use of information 10

Open source clickstream collector: Divolte 11
Advantages of Divolte 11
Shouldn’t online information be timely and relevant in this age?

Relevancy. Marketers have been raving about the right message at the right time for years. Relevancy is not concerned with channels. Omnichannel is the only way.

Web shops that introduced recommendations started a revolution, but custom-built recommenders remained out of reach for most organizations. The availability of new (open-source) technology has made the realization of custom recommenders much easier. You’d expect most online retailers to personalize their front-end to the preferences of each website visitor in real-time these days. But in reality, practice proves differently.

According to Big Data Survey, the national data research initiated by Big Data Expo and GoDataDriven, 50% of organizations don’t personalize their website at all. Only 10% of those who do personalize, automate in real-time, despite the many technological solutions available to adjust a site to the preferences of a visitor (group). One-third of organizations in travel claim to have some form of automated personalization on their website. Increasing relevancy by providing a personalized user experience doesn’t have to be difficult. In this whitepaper, we share tips and best-practices to personalize a site using real-time recommenders.
Real-time recommenders that respond to individual behavior

Exploiting the opportunities in customer data is possible and affordable. Technology that stores customer data on one, integrated channel and analyzes large volumes of data in real-time has developed rapidly and is available to every organization.

Recommenders are predictive models that aim to determine the needs of individual relations. The more data an organization has at its disposal, the better the recommendations. Recommenders can, for example, suggest next best actions, interesting products or services. It’s the art of recognizing which step, which product or which service is the best for every individual relation RIGHT NOW, within every digital touchpoint.

The right offer, but unfortunately, out of context

Online companies have focused on customer profiles and personas for some time now. Often, these are determined based on purchases made in the past. But historic purchases are only partial pointers to a person’s current interest. Product owners can play it safe and personalize their website based on preferences, with the downside of not being very specific. An alternative is to use remarketing; displaying offers of viewed products on external websites.

You’ve probably had this experience while browsing; you keep seeing banners with products you’ve viewed somewhere on a website weeks before. Worse, in some cases, you’ve already purchased the item. To some extent, the offer is relevant, but it’s simply too late. Ideally, personalization is used within the context of the current session so that a visitor’s buying intention and your offer are in sync.

It is now technologically possible to analyze customer behavior in real-time. The next step is not only to analyze but to use this customer data in real-time. The better the underlying algorithms are refined, the better you can cater to the needs of an individual. For this reason, off-the-shelf solutions will always be outperformed by custom-built recommenders that factor in domain knowledge.

The quality of personalization improves significantly by combining purchase history with real-time clickstream data. By ingesting behavioral data as well as other data sources, a model can be trained to understand customer needs. For website visitors to find what they need quicker and better, different customers or customer groups can be offered different search results. It may even lead to exciting product or services suggestions that the customer had not thought of before.

“Ideally, personalization is used within the context of the current session so that a visitor’s buying intention and your offer are in sync.”

Getting started with recommenders

Recommenders can play a role in any stage of the customer journey. Developing recommenders begin with pointing out use cases, collecting data and testing the business value of a use case by running experiments.

Pointing out use cases

Product recommendations based on what other visitors have viewed or bought may be the first use cases that spring to mind. Amazon.com once made global headlines with its recommenders and many web shops have followed suit. Today, it is possible to react to individuals by closely examining their behavior. This information can be used not only to recommend products but also to offer the best next action. By combining behavioral data of the current session with other available data sources, you can provide a fully tailored proposition for every individual.
Smart recommenders enable organizations to offer unparalleled service for a fraction of the cost and effort traditionally needed. To name a few applications:

- **Visitor type recognition** Within seconds, travel organizations like Transavia (see success story later in this whitepaper) recognize whether a visitor is looking for a holiday or a business trip. Based on this, they personalize the checkout process.

- **Buying intention** Web shops determine the buying intention of a website visitor based on their on-site searches. It makes a huge difference whether a visitor is looking for “TV” or for “Samsung Ultra HD TV Curved.” In the former case, you would inform the visitor about all the options. In the latter, you send the visitor straight to the checkout process.

- **Accelerate product introductions** By including new products as recommended products for specific user groups, organizations can quickly discover if these products appeal to certain visitors. When visitors respond positively, the recommender can be trained to show these products to comparable visitors, which works for existing products as well as for products with a cold start.

- **Better adaptability to change** By closely monitoring changes in behavioral patterns, you can spot market developments and upcoming trends at an early stage and adapt your product offering accordingly.

- **Internet of Things** Recommenders can also be applied to IoT, like analyzing thermostat data to provide personal recommendations for improving energy efficiency or home insulation.

- **Positive effects on the long tail** Often, recommenders lead to interesting suggestions in the long tail. Dutch national broadcaster NPO (see success story later in this whitepaper) experienced an uplift in older content when they started using personal recommenders online.

**Collecting data**

To develop data-driven applications, it is essential to have as much good quality data available as possible. Recommenders can be developed by combining numerous data sources, like web server log files, reviews, product information and clickstream data. The collection of streaming data allows organizations to adapt to the current needs of individual relations.

Many organizations use analytics data that has been collected on a third-party platform. The downside is that you provide an external company with your analytics data, which only comes available for your purposes after you invest substantial amounts of money in buying this data. Using a clickstream collector on your organization’s infrastructure ensures continuous ownership and accessibility of all the clickstream data.

Because the purpose of data is not always apparent initially, it’s advisable to collect and store as much of it as possible and to refrain from disposing or aggregating it. A scalable data platform, like open source solution Hadoop, is a good option for storing data for future use.

**Testing use cases**

Because the pace of change is so fast and ever accelerating, long-term project planning is futile by definition. The ability to adapt to current needs, demands and developments quickly and in an agile way are far more valuable than a rigid strategy for the next five years.

Experimentation enables organizations to adjust product development and develop predictive models. It is essential for testing the waters and keeping a finger on the pulse of the client or customer. Organizations like Uber, Airbnb, Booking.com and Zalando invest in the development of their data science teams and have demonstrated substantial growth with this approach.

A data scientist can be described a curious developer with a statistical background who understands both the numbers and the business impact. A data scientist experiments with data and trains models with the purpose of creating value for the organization.

**Technology should not be the limiting factor**

Technologically there is no reason not to personalize the entire journey for your clients. Still, many organizations find adopting the right technology and a new way of working challenging. Replacing a complete infrastructure within an organization is something that is impossible to do overnight. Many organizations also find it difficult to develop new propositions and innovate, simply because the existing team members are not yet skilled for this, and there is little time available to experiment.

Organizations that choose to implement open-source technology and make time available to experiment prove capable of implementing recommenders within an existing infrastructure most quickly.

**Integrating recommenders in existing infrastructure**

To avoid silo formation and technological lock-in, it is preferable to store website data (and data from other sources) in a central data platform. This separates data, algorithms and front-end (see “E-commerce Stack”). It is not realistic to replace existing infrastructure that took years to implement overnight. Fortunately, it is perfectly possible to integrate recommenders with existing content management systems or e-commerce platforms, like Sitecore, SAP Hybris, IBM, or Oracle.

All that is required is a clickstream collector (see the final section of this whitepaper, “Divolte”), a data platform and a smart data scientist. Recommendations from an algorithm can simply be made available to the front-end via an API.
Best-of-breed platform
Organizations used to build their entire platform from scratch or purchase an existing e-commerce platform (as-a-service). The former requires a huge investment in development and operations, with the risk of always running behind in regards to the adoption of innovation. The latter raises the question if an all-in-one solution can provide the best option on every single requirement. The one thing that is certain, is that you create vendor lock-in. Good enough as long as all is well, but when you decide you want to move to a new platform, the high threshold of switching platforms can be a real pain.

Nowadays, it is very well possible to combine applications that excel in specific features to form a best-of-breed solution built from a patchwork of solutions. Data should not be stored within the separate applications but should be ingested in a separate data platform that is based on open source technology Hadoop.

By implementing an intelligence layer that sits between the data platform and the front-end, data, intelligence and user experience in the front-end are separated so that each of them can serve its purpose best.

Organizations that have built their platform, develop the capability to continuously upgrade platform features as soon as better technology becomes available. By developing their own algorithms, these organizations are able to personalize all of their customer touchpoint, in near real-time.
Success stories

Transavia: Personal recommenders to improve online user experience

In an industry where price pressure increases exponentially, Transavia aims to make a difference by pro-actively offering the right service at the right moment during the full customer journey. It’s a transparent approach where all conditions and costs are made clear directly from the start.

Personal recommenders play a major role in this strategy, creating a fully-personalized customer journey. This experience starts with the recommendation of destinations based on individual website behavior.

“Viewed destinations are compared in real-time with historical website data. Based on this information we recommend destinations that have the strongest match with the search behavior of every single website visitor,” says Charles Verstegen, Senior Revenue Development Manager at Transavia.

Personal recommenders are not limited to the destination finder. Transavia sees opportunities to increase the relevancy of other product and service offerings as well.

“Take a business traveler. Generally speaking, this type of customer benefits more from an offer for more legroom than for bigger luggage allowance, while this is exactly opposite for leisure travelers. By taking these characteristics into account, we can recommend the right bundle to different types of visitors. For our customers, this leads to an immediate improvement of the user experience,” says Verstegen.

Transavia expects to add real-time price information to the flight pages soon, allowing customers to easily compare prices between different destinations that are likely relevant to the current session. Based on price or availability, a visitor that actively searches for flights to Valencia might be persuaded to book a flight to Malaga instead.
The ultimate goal is to provide the right information, service and offers to every individual customer in every step of the customer journey, from finding their destination to returning home.

**Better service by collecting customer information**
To get to know customers better, it is essential to gather as much relevant information as possible. Transavia uses a broad range of customer-related information to find correlations and train models. The machine learning models are fed with data like group size, duration, group composition, season, a moment of the day and, of course, destination.

Finally, to personalize the website, the generated recommenders were made available for the Sitecore website via an API. GoDataDriven implemented Divolte and worked with Transavia’s data team to develop the recommenders.

Divolte was not the first clickstream collector to be implemented at Transavia. Various solutions were installed for various purposes. Quickly after the initial implementation of Divolte, Transavia decided to consolidate as many clickstreams into Divolte as possible.

“The big advantage of Divolte is the flexibility. Because of the open source character of Divolte, the clickstream collector lacks the limitations that we experienced in other solutions,” says Verstegen. “Our data team is very excited about the possibilities they now have with Divolte. For example, now our web team is able to tag data exactly as the data team prescribes.”

**The increased importance of the data team**
Transavia always had a strong foundation for data-driven modeling of flight optimization, but the rapid expansion of data and data sources created opportunities for more advanced features. The airline couldn’t seize these opportunities within existing structures and processes, so in the spring of 2016, Transavia introduced a dedicated data team composed of professionals from both business and IT departments.

The data team has full autonomy for innovation, development, and operations of data applications. The data specialists communicate with all teams involved, like architecture or e-commerce, to reconcile and exchange information on a daily basis. The team adopted an agile process for projects, which divides complex projects into clear sprints.

Smart data applications and projects to optimize revenue and operational excellence are crucial for Transavia’s long term strategy. Models that optimize crew planning or the prediction of potential disturbances, like maintenance, will be the data team’s key focus.
For over 60 years, the Nederlandse Publieke Omroep (NPO), Public Broadcasting Organization, has been producing and broadcasting radio and television. Every week, NPO reaches 85% of the Dutch population, with a larger presence online. To provide relevant content for online viewers, NPO recently implemented several smart data applications.

Personal broadcasts
Because of changes in viewing behavior and an increased number of channels, it has become increasingly difficult for media companies to retain viewer attention. To remain relevant in the coming decade as a public broadcaster and to make sure that viewers don’t get lost in the overwhelming offer of video content, NPO set an objective to become a personal broadcasting organization.

NPO began pointing out potential use cases, like a/b tests, automatically generated playlists and personal newsletters to offer consumers personal viewing experiences. To successfully develop these use cases, it pointed out three main themes; dashboards, recommenders and the introduction of NPO ID.

The right process for innovation
Product innovation required implementing an agile methodology and working in multidisciplinary teams.

“For NPO, agile working in multidisciplinary teams was not common practice, to put it mildly. We found out that it was not only necessary to adopt our way of working, but that the whole office needed a makeover. We have made necessary adjustments to establish a modern and flexible workspace through the full line of business,” said Marcel Collette, then manager of information systems at NPO.

Because NPO did not employ data engineers and data scientists at the time, the broadcaster decided to in-source these skills. The consultants from GoDataDriven helped kickstart the project and also actively shared knowledge with all team members.

“A strong startup vibe could be felt within the new data team, especially in the early days. Speed was everything,” according to Erik van Heeswijk, then ad interim project manager responsible for big data strategy development at NPO. “In a later phase, the focus shifted more to internal support and process management within the organization.”
Real time recommenders in e-commerce

Setting up the data platform and developing use cases
Software is often developed for commercial purposes. For NPO, features don’t always fit with their public character and guide role. Webshops put a lot of effort into designing a tight sales funnel or focus on conversion optimization. For NPO, these topics are not relevant. What has always been very relevant to NPO is transparency. It is important that users know what personal details are stored and what they are used for, which means that NPO only uses data from individual uses when permission has been granted to provide relevant recommendations.

NPO has built their infrastructure using proven, open-source technology like Python, Java, Divolte, Hadoop, and Spark.

“By building the platform and data-driven applications from the ground up, we have been able to experience the entire learning curve and learn to understand how everything works,” explained Collette.

Much effort has been put in the realization of a central data platform based on HDFS and Spark that combined data from the various brands within the NPO organization. The implementation of the platform started with hosting. To process streaming internet data in an optimal and scalable way, NPO chose to implement Divolte as clickstream collector and to host the platform in the cloud.

“We had very little experience with cloud hosting. It was important that our ICT department built up this expertise, so they were closely involved directly from the start. When the cloud environment proved to be solid and secure, internal support instantly increased,” said Erik van Heeswijk. “The visualization of data in dashboards and the demos after every sprint also contributed to the internal support.”

Understanding behavior through dashboards
For the initial dashboard development and data collection understanding, NPO initiated two pilot projects. One with a daily platform (NOS) and one with a general platform (NPO.nl). Editors were actively involved and determined how to collect data and what metrics and visualizations should have focus. Their specific domain knowledge provided context to what works and what doesn’t.

Setting the right KPI’s was essential. If content is valued based on the number of people that have viewed the whole program, then NPO should cap video duration to five minutes and offer content of cats and Katja Schuurman.

Based on the collected data, various visualizations were developed, including the most popular articles on the NOS website, the life expectancy of an article (how long an article will remain relevant), and the current number of viewers compared to the expected number.

Personal recommendations
NPO uses open-source solution Divolte to collect all website interactions in real-time. The collected data can be used to create dashboards as well as recommendations based on viewed content.

Recommendations based on the video content on the site allows NPO to offer visitors a more relevant program. Views on long tail content have increased, leading to content from the entire video catalog of NPO being served. “Not only do we see that content is watched more often and longer, but the value of older content has also increased due to the recommenders,” explains Erik van Heeswijk.

Correlations between viewing behavior and programs from different NPO brands are analyzed. Every color is a different brand; every circle is a program. The bigger the circle, the longer the playtime.

Transparent use of information
Transparency is very valuable to NPO. There is a clear separation between anonymous browsing and a personal environment available behind a log-in.

“As a public broadcasting company, we have to take privacy very seriously. We carefully protect our user data. An organization like Channel 4 acts as a reference for us. Anonymous website visitors receive recommendations based on general trends and editor picks. But logged-in users will receive more personal recommendations soon,” according to Marcel Collette.

“Visitors that create an NPO ID will receive better service and content they will appreciate. Building up a relationship between the viewer and NPO is crucial for this relationship and we can only achieve it with a clear proposition. Members can always access their personal data and have it deleted from our databases if they choose.”

By understanding the viewing behavior on the NPO websites better, the public broadcasting organization has been able to develop relevant recommendations that will lead to even more engaging content in the future.
Open source clickstream collector: Divolte

Divolte Collector is a scalable, high-performance, open-source solution to collect large amounts of clickstream data in Hadoop. Divolte has been developed by GoDataDriven in collaboration with several organizations, including Wehkamp, NPO, Transavia, De Persgroep and Hallmark, with the ambition to offer a flexible solution to collect clickstream data.

Divolte collects both website and app data, allowing developers to create recommender engines and banner optimization systems. Organizations that use Divolte benefit from data collection with a first-party cookie and the flexibility to store data in any location.

Advantages of Divolte

Open Source: Divolte is available free-of-charge for everyone and is developed by the community.

Full control: Data is stored in your own database. You have full ownership of all data.

No sampling, no aggregation, all raw data is and remains available.

Flexible: Data teams rave about Divolte’s flexibility, enabling web developers to tag exactly as the data team prescribes.

Easy to integrate: You integrate Divolte by adding a tag to your website, that’s it.

Scalable: Divolte has been developed with high-traffic websites in mind and is capable of handling hundreds of thousands of web events per second.

More information? Go to www.godatadriven.com/divolte or send an email to signal@godatadriven.com.

About GoDataDriven

Since 2009, GoDataDriven has been at the forefront of Big Data and Data Science innovation in the Netherlands. The GoDataDriven team consists of a unique combination of experienced data scientists and excellent data engineers, all with state-of-art technological knowledge.

GoDataDriven develops custom data-driven solutions for leading enterprises. Whether it is about setting up data infrastructure, improving data quality, search optimization for an e-retailer, fraud detection for a bank or sales forecasting for a production company; bottom-line, GoDataDriven’s solutions always provide real value for the client.

Clients include eBay, Wehkamp, Eneco, KLM, Schiphol Airport, ING, Rabobank, Bakkersland, Transavia, and NPO.

For more information, please visit www.godatadriven.com or contact Rob Dielemans at robdielemans@godatadriven.com or +31(0)35 67 29 069.