Next-Generation Matching for a Modern Customer 360

Key Benefits

As an enterprise customer data platform solution, Informatica Customer 360 Insights turns raw data into intelligent insights.

- **AI-powered contextual matching** learns from business user decisions to create highly-connected views of the customer

- **Real-time reasoning** identifies patterns and trends in customer journeys and preferences

- **Infinite perspectives** provide each user with the view most meaningful to their use case

- **Extreme performance and scale** addresses every use case with all available data from any source

The State of Customer Data

The centuries-old mantra "The Customer is Always Right" has never been more relevant than in today's experience-driven environment. Customers want what they want, when they want it, and they want it personalized to their tastes and needs. Products, marketing, and procurement of goods and services need to be accompanied by the right experience, or customers will go elsewhere.

As customers interact with organizations through many different channels, they provide endless streams of information and insight that indicate their preferences, sentiment, and intent. However, those insights are often hidden in fragments of data or in unstructured formats that are often difficult to tap into.

For decades, organizations have strived to capture, manage, and make sense of customer data. The concept of a Customer 360 view is nothing new: it has always meant "everything about the customer." In today's digital age, what "everything" means has changed significantly, as there is more data and more types of data being captured and stored. Customer information is no longer limited to core attributes or transactions—digital transformation has produced additional methods of interaction including surveys like Net Promoter Scores (NPS), webchats, browsing history, social media posts, and much more. And along with these new types and formats of data there are new technologies you can use to create and manage modern Customer 360s as well.

Traditional Matching Techniques

Matching is a key capability for any Customer 360 solution. As data can be duplicated or fragmented and scattered across many siloed systems within an organization, matching is used to bring data together and identify unique entities. It is well known that there are multiple techniques for matching customer records, whether it is matching people to people or matching organizations to organizations. The two most common techniques are rules-based and probabilistic.

The rules-based approach compares attribute values using a variety of techniques (e.g., fuzzy matching, data frequency considerations, etc.) and then applies a set of conditions to determine what is a match. For example:

Rule 1: If NameScore is >= 90% match and BirthDateScore is an exact match and AddressScore is an exact match then result is a MATCH

Rule 2: If NameScore is an exact match and SSN_Score is an exact match then result is a MATCH

The probabilistic approach compares attribute values and then aggregates the results using a formula that yields a confidence score (or likeliness score). Any score above a threshold is considered a match. The most common approach in practice is a weighted-sum formula:

Score =   NameWeight * NameScore
          + BirthDateWeight * BirthDateScore
          + AddressWeight * AddressScore
          + SSN_Weight * SSN_Score

If score > match_threshold then result is a MATCH

While these techniques have been in practice for many years and are effective for matching, they do have limitations when considering today's diverse and sparse data that makes up a Customer 360.

Deduplication Only: The focus is on matching customers to customers in order to deduplicate them. However, there is also the need to match "objects" (such as interactions and unstructured documents) to customers, and this is a different type of problem to solve.

Limited Attribution: "Demographic attributes" typically use attributes such as name, address, phone and birthdate for customer-to-customer matching. These are obviously core attributes to use; however, relationship data and account/transaction data can improve match results (including identifying what is not a match) and, more importantly, play a critical role in matching "objects" to customers. For example: when an individual has included only name demographic information in a contact form on a website, how can you match that interaction to an internal customer record? It would require additional attributes that are in the form such as references to products or order numbers.

Structured Data Only: Traditional matching techniques, including the processes that use them, work with only structured data. Over 80% of data in an enterprise is unstructured, and a lot of it contains valuable insights about customers, including indicators of purchase or churn, sentiment, life events, and more. Not being able to match this data into a Customer 360 has become a competitive disadvantage.

Expert Tuning Required: A matching engine expert is normally required to tune the algorithm for an implementation, whether it be rules-based or probabilistic. Sometimes this is described as "art and not science," which is a different way of saying that this expert has their own interpretation and biases about what is needed. The tuning outcome can very easily be a sub-optimal algorithm that yields results with under matches and over matches.

Synthesis: Next-Generation Matching

Synthesis is a term used in next-generation customer data matching that ties unstructured elements and interactions to a customer record for a deeper understanding of what is happening in, and around, each customer touchpoint. Are you using machine learning to do this, or do you require an army of data stewards?

Synthesis is a next-generation matching technique that is focused on building out today and tomorrow's Customer 360. For example, it addresses matching prospects to customers, matching unstructured data and interactions to prospects and customers for the purpose of discovering non-obvious relationships. It uses "contextual attributes", machine learning, natural language processing, and a combination of probabilistic matching with declarative rules to accomplish this. And finally, it provides dashboards and reports so consumers of the data gain trust in what is done.

Matching Objects to Customers: Matching customers to customers is not sufficient to build out a Customer 360 required for today's analytics and operational use. Stitching in data such as accounts and transactions is normally straightforward, using cross-walk keys available in the data. However, there is a wealth of other data including unauthenticated interactions (such as webchats, emails and anonymous web visitors), and unstructured documents that can only be associated to customer records using a matching technology. Furthermore, this type of data contains valuable insights for analytics including indicators of churn, indicators of purchase, sentiment, and life events. These insights are only useful if they can be associated to that specific customer.

Contextual Attributes: Traditionally, customer matching focuses on demographic attributes such as name, address, and phone number. There is a considerable amount of other data in a Customer 360 that can be leveraged once you consider the customer's relationship with the business. This is especially important when matching objects and interactions (such as emails and webchats) into a Customer 360. It is possible to compare "contextual attributes" such as when did the interaction occur, what was it about, where did it take place, who was referenced in it, and what products and services were mentioned, to transaction and other interaction data associated to customers in order to match with a given confidence level.

Entity to Party Contextual Matching

From Webchat

John
IP Address: 184.122.10.1 (London)
Date: 2018-03-14
Subject: Delivery

John: Hi, I ordered a X3 server and a Z4 laptop last week. The server was delivered yesterday but I didn't receive the laptop yet, which was for my daughter who is leaving for university soon.

Agent: Yes, it is common for them to be delivered separately.

To Party

John Edwards
20 7234 3456
123 Main St, Toronto, ON, Canada

Transaction: Purchase
Date: 2018-03-09
Item 1: 94883-003 (X3 Server)
Item 2: 47242-004 (Z4 Laptop)

Transaction: Delivery
Date: 2018-03-13
Item 1: 94883-003 (X3 Server)

It is not possible to match the webchat to the party record with just the basic demographic data provided.

Confidence that this webchat is associated to this party is much higher once considering date, transaction and product info.

The same pattern applies for social data, emails, product reviews, unstructured documents and more.

Natural Language Processing: Unstructured text in its raw form is not well suited to being matched to customer records. There are, of course, powerful text search technologies available, but that is different than matching. Instead, we must apply Information Extraction (IE) to pull out the "contextual attributes" from the unstructured text and use it in matching. These contextual attributes then provide many more data points to use in the matching process.

Natural Language Processing

Tweet

BankCo mortgage rates are too high - time to switch to CreditUnion Co

Webchat

Agent: Hi John, how can I help you?

Dave: Hi. My mortgage expires on Nov 23 2017. What notification period is required for me to cancel it?

Agent: Sorry to hear you want to cancel your mortgage. Standard notification period is …

Company / Partners
Competitors
Topics
Products / Services
Dates / Time

Entity Information
 - Names
 - Addresses
 - Phone Numbers
 - …

Account Information
Relationship Information
Life Events

Machine Learning: There is a critical role for machine learning in next-generation matching. A matching algorithm can be learned using a supervised training approach where data stewards and subject matter experts label a properly selected set of match pairs as either matches or non-matches. These labeled matched pairs form a training set that is used to produce a matching algorithm. There are many benefits to this approach:

- **Unbiased:** The algorithm is learned from the data experts, as opposed to a consulting resource with biases manually tuning it.
- **Accurate:** In tuning an algorithm, there are a large number of variables to consider. Adjusting one variable has a rippling impact for all other variables, and this is why tuning by experts has been described as "art and not science," since it is impossible to consider countless permutations. Machine learning is designed to overcome this challenge and produce a more accurate matching algorithm. This is especially important when contextual matching is required and there are many more variables to consider.
- **Descriptive:** The machine learning process can indicate where it is doing well and—perhaps more importantly—it can describe the scenarios where it is not doing well, which are over-matches and under-matches. These scenarios can be addressed with declarative rules.

Probabilistic and Declarative Rules: There is often debate whether probabilistic or rules-based matching is better. The short answer is "both." The machine learning process produces a matching algorithm that is probabilistic but also a description of scenarios it is not able to satisfy. These scenarios are addressed through declarative rules.

Perspectives—Multiple Views of Synthesized Data

Informatica® MDM Synthesis will stitch together a full Customer 360 consisting of demographic, account, transaction, interaction, and unstructured data. Results of traditional matching algorithms are merged together to form entities. For example, if three distinct records are determined to be the same customer, then those records are merged together into a single entity. In contrast to traditional matching techniques, synthesis manages all the data in a graph. Data is related together with confidence levels. It is then possible to provide multiple views, or "perspectives," of a customer.

Customer 360 Examples

Synthesis connects more data points using many different match strategies to deliver context. In the example below, by connecting all the data from accounts, transactions, and interactions, we gain an understanding of John Edwards and his relationship to Jane Brown Edwards.

Synthesis connects data points using many different match strategies to deliver context

Reasoning infers intelligence that is then stored as part of the Customer 360. Through the process of reasoning, we gain insight into John's life events, household, journey, influencer network, and overall sentiment. Based on this insight, marketers can develop targeted campaigns, find others like him, and build personalized marketing offers.

Reasoning infers intelligence that is then stored as part of the Customer 360

In summary, Customer 360 requirements have changed; therefore, the solutions and matching techniques have also evolved to address those requirements. Same definition, different 360.

Next Steps

Learn more about Informatica Customer 360 Solutions.