



◀ **MAIN** ▶

Session 2.0

29.08.2022

OTTO-Campus

Hamburg & Digital



**KEY-
LEARNINGS
der Sessions**

OTTO

Machine Learning hat NLP gelöst...oder? Ein Plädoyer für hybride Lösungsansätze in angewandter Sprachtechnologie



- Machine Learning hat NLP in den letzten zehn Jahren weit nach vorne gebracht und die Latte in vielen Standard-Tasks deutlich höher gehängt.
- Wir müssen endlich aufhören, Machine Learning als die einzige innovative Technologie im Umfeld von KI und NLP zu betrachten und verstehen, dass wirklich nur sehr wenige Probleme erschöpfend mit Machine Learning zu lösen sind.
- Wirklich leistungsstarke NLP-Lösungen ergeben sich aus der intelligenten Verknüpfung unterschiedlicher Lösungsansätze zu einer hybriden Gesamtlösung.
- Die Stärke hybrider Systeme liegt in der komplementären Kombination von Systemen mit unterschiedlichen Fähigkeiten und Fehlerprofilen.

Rethinking semantic product search: new query-centric approaches for low-call segments



- Approaches to semantic product search with NLP language models can be divided into query- and product-centric ones, with research moving to embrace query-centric solutions
- Problems can often be formulated through both lenses, and it pays to reflect very carefully about the problem definition as it predefines the subsequent solution.
- In low-recall segments, it pays to shift the problem analyses from product-centric to query-centric.
- “Low recall” can also be redefined not as a problem of “hidden relevant products that are similar to the found lexical products”, but as a problem of “specialized tail queries that are similar to other tail and head queries, which link to more relevant products”.
- In conclusion, query-centric AI-solutions with NLP-language models in low-recall segments provide an excellent corollary to product-centric lexical search engines in modern hybrid search architectures that we employ at Otto.

Neural Search mit Sentence Transformers



- Sentence Transformers können Synonyme und den Kontext einer Suchanfrage verstehen
- Vortrainierte Transformer Modelle haben out-of-the-box ein akzeptables Sprachverständnis
- Für spezifischen Jargon einer Krankenkasse ist ein Re-Training mit eigenen Daten nötig
- Training mit eigenen Daten verbessert Ergebnisse deutlich
- Schlanke Bereitstellung mit FastAPI möglich

NLP as a service: From PoC to Production with GPT-3



- Large Language Models like GPT-3 give amazing results
- Try to focus on an actual user problem to get started
- Do a POC with your team in OpenAI's Playground
- NLP-based products should have a user-in-the-loop!
- Implement a prototype using the API and get user feedback as soon as possible

Vom Paper zum produktiven Service: Implementierung eines Transformer basierten Recommender Systems



- Lieber langsamer vorankommen und dafür korrekt implementieren, als Fehler zugunsten von Zeitgewinn zulassen.
- Arbeitet immer zuerst an der Stelle des Projektes mit der größten Unsicherheit.
- Zu glauben etwas sei richtig implementiert, ob wohl es falsch ist, und darauf weitere Entscheidungen zu basieren, stellt eines der größten Risiken für das Projekt dar.
- Vorsicht beim in batch negative sampling.
- Die Loss Function sollte wirklich verstanden werden.

Deep Multimodal Recommendations: Solving the cold start problem



-  Read more papers! By comparing your production systems to the current state-of-the-art, you can discover substantial opportunities for improvement.
-  Infrastructure is important! The amount of data involved in processing images require unique technical solutions that can take a long time to implement without prior experience.
-  Do more hackathons! They may be fun, but they are a valuable tool for product development, too.
-  Complicated is not always better! Start with a model that you understand theoretically and can technically handle.
-  Go Live ASAP! Even on Friday afternoons.

Automatisierung von Machine Learning Prozessen für Recommencer Systeme



- Machine Learning Modelle und der Code, der diese Modelle erzeugt, ist Teil des Produktions-Codes und sollte auch entsprechend behandelt werden.
- Wenn Modelle sich ändern (z.B. neu trainiert werden), ist es oftmals sinnvoll vor dem Deployment zu messen, wie sich die Änderungen auf das Verhalten des Gesamtsystems auswirken, in dem die Modelle integriert sind.
- Ergeben sich starke Änderungen im Gesamtverhalten, muss man validieren, ob diese Änderungen positiv oder negativ sind (bei Online Plattformen z.B. via A/B Tests).

Die Post-Deployment Phase eines KI Produktes



- In the post-deployment phase frequent changes are necessary, thus strive for an easily maintainable and deployable AI product.
- Best practices from system engineering are a great start to achieving an easily maintainable and deployable product. Yet, they are insufficient for an AI product.
- MLOps today is still in a very messy state with regard to tooling, practices, and standards. So, there is no single right solution.
- Today's right is maybe tomorrow's wrong. Don't be afraid to challenge past decisions.

Von der Auswahl geeigneter KI-Use Cases bis zur Prävention einer „Lost Generation“ of Cognitive Automation



- Apply the use case assessment
 - ... to reduce hype and fear in your organization
 - ... to foster collaboration between business and IT departments.
 - ... to divide and conquer use cases thus making them feasible for successful implementation and deployment.
 - ... to also say “no” to unsuitable use cases, i.e., not following a technology-push paradigm that seeks to implement AI for the sake of AI.
- Automate in a human-centered and sustainable way.
- Mindset: “Automate to Innovate”.



- Die Grundidee des Manifold Learnings – die wichtigsten Information in hochdimensionalen Datensätzen liegen auf niedrigdimensionalen Mannigfaltigkeiten, die wir approximativ berechnen können
- Die wichtigsten Methoden – PCA, Multi-Dimensional Scaling, und insbesondere UMAP
- Praktische Anwendungen in den Bereichen Feature Engineering, Dimensionality Reduction und Data Visualization
- Kombinierbar mit anderen interessanten Methoden wie Random Projections
- Verfügbar in Standard-Python-Libraries wie sklearn und umap-learn

Dynamic Pricing in Logistics. How does AI-assisted Pricing work?



- Dynamic Pricing consists of multiple data products.
- Domain knowledge and business input are indispensable for Dynamic Pricing to work.
- First, we need to be able to quantify price elasticity in a market.
- Start with aggregate price predictions and continue towards personalized pricing.
- Market data are important, even if they are only measured by proxies.

Data Science for Social Good in Berlin: How Data Science can help NGOs



- If you organise a datathon/hackathon, it's very important to invite data ambassadors who would manage and prepare data for further usage by volunteers
- If you work with NGOs, it's extremely important to regularly nudge them and remind them about projects you work on together
- If you want to help NGOs with data-related issues, be ready to work with simple EDA/data management tasks, most NGOs do not have big data and are not interested in machine learning or AI
- Always check and explore properly the data that you get from NGOs, there could be many mistakes and missing values
- You need to learn how to explain all your findings and methods in layman's terms to be on the same side as NGOs

Datenminimalismus – Bessere AI mit weniger Daten?



- Es lohnt sich mit wenigen Trainingsdaten zu starten, um festzustellen, wie die Modelgüte mit der Datenmenge skaliert
- Damit lässt sich abschätzen welche Datenmenge benötigt wird, um eine erwartete Zielmodelgüte zu erreichen
- Die Reduktion der Trainingsdatenmenge hilft die Daten besser zu verstehen
- Eine intelligente Auswahl der Trainingsdaten reduziert Kosten, ohne die Modelgüte zu verschlechtern
- Kein Datensatz ist perfekt.

Data Science Infrastruktur – vom Experimentieren zur Industrialisierung



- Das erfolgreiche Realisieren von Data Science Anwendungsfällen erfordert eine dedizierte Infrastruktur. Hier lohnt es sich, frühzeitig eine modulare Architektur zu verfolgen, um unterschiedliche Projekte über über eine gemeinsame Plattform abzubilden.
- Mittels Infrastructure-as-code und CI/CD-Techniken lassen sich viele Schritte des Lifecycles von Data Science Projekten automatisieren.
- Beim Experimentieren und Industrialisieren von Data Science Anwendungsfällen sind wiederkehrende Teilaufgaben beim Programmieren zu lösen. Hier können selbstentwickelte Libraries helfen.
- Data Science erfordert ein hohes Maß an Kollaboration verschiedener Rollen innerhalb der Organisation. Über ein geeignetes RBAC-Konzept können diese Rollen auf einer gemeinsamen Projektinfrastruktur zusammenwirken.

Bringing machine learning applications into production



- Successful companies in their ML Ops journey have a strong focus on three aspects: People, Process and Technology.
- Companies adopt machine learning cloud solutions for better performance, agility of resources, availability of AI services for quick wins, managed ML services to remove the heavy lifting of building new features around experiment tracking, model monitoring and alike and because of ML Ops functionalities.
- A typical ML Ops pipeline consists of four parts: data, (re-)training, inference and build pipeline. All of those can be available as template to data science teams.
- Building robust ML Ops pipelines requires a stable foundation such as multi-account deployments and well-designed access management to build ML Ops pipelines on top.
- Company-wide ML Ops guardrails can be used as a mechanism to streamline adoption throughout the organization.

Data Juniors rekrutieren - ein Perspektivenwechsel



- Die starke Fokussierung auf Data Professionals mit vielen Jahren Berufserfahrung in genau den Tools die ein Data-Team selbst einsetzt ist eine kurzfristige Wette.
- Die richtigen Data Juniors zu rekrutieren und einzuarbeiten kann ein entscheidender Wettbewerbsvorteil sein, wenn Teams die richtigen Voraussetzungen zur Entwicklung schaffen.
- Die Erstellung eines Einarbeitungs-Konzept sollte dabei große Aufmerksamkeit, insbesondere von erfahrenen Mitarbeitern bekommen.
- Bei den Erwartungen an die technischen Fähigkeiten von Juniors (aber zuweilen auch Erfahrenen) sollte es weniger um Tools und mehr um Konzepte gehen.
- Effektives Lernen geht häufig besten mit anderen zusammen. Sei es durch ständig wechselndes Pair-Programming, kurze Team Projekte oder das Vorbereiten von eigenen Lerninhalten.