Big-Data in der Bundesverwaltung

Fachgebiet: Arbeitsmarktstatistik
Betreuer: Dr. Arno Schmidhauser
Experte: Siddhartha Arora (IBM Switzerland Ltd.)
Industriepartner: Staatssekretariat für Wirtschaft, Bern

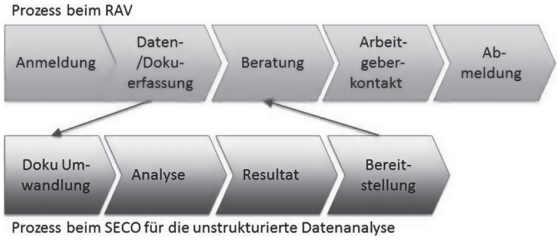
Bis heute werden in der Arbeitsmarktstatistik ausschliesslich strukturierte Daten ausgewertet. Mit der Analyse von unstrukturierten Daten, wie zum Beispiel Text in Lebensläufen, kann zusätzliches Wissen generiert werden. Damit soll die Effizienz in der Verwaltung sowie die Betreuung und Unterstützung von arbeitslosen und stellensuchenden Personen optimiert werden.

Ausgangslage

Das Staatssekretariat für Wirtschaft SECO ist das Kompetenzzentrum des Bundes für alle Kernfragen der Wirtschaftspolitik.Das SECO hat auch die Rolle der Schweizer Arbeitsmarktbehörde. Damit ist es verantwortlich für das Arbeitsvermittlungs- und Arbeitslosenversicherungsgesetz.
Bis in die 1990-er Jahre wurden arbeitslose Personen eher verwaltet als betreut. Dies änderte sich mit dem Anstieg der Arbeitslosenzahlen. Die regionalen Arbeitsvermittlungszentren (RAV) wurden gegründet, eine Vielzahl an Kursen und Programmen für Arbeitslose werden seither angeboten. Damit stieg auch der Bedarf an statistischen Daten. Aus diesem Grund wurde die Datenaufbereitung in einer statistischen Datenbank zentralisiert. In der Datenbank werden nicht nur die strukturierten Daten gespeichert, auch Dokumente wie Lebensläufe, Bewerbungsschreiben und Arbeitszeugnisse sollen hinterlegt werden können.

Ziel

Mit den Technologien zur Analyse von unstrukturierten Daten haben wir jetzt die Möglichkeit, neue Wege in der Arbeitsmarktstatistik zu gehen. Zeitnah können wir sowohl strukturierte, als auch unstrukturierte Daten analysieren. Resultate aus beiden Analysen können zusammen verknüpft einen Mehrwert für die Betreuung der Arbeitslosen sowie für die Entwicklung neuer arbeitsmarktlicher Massnahmen generieren. Ausserdem kann mit maschinellen Analysen viel rascher auf Marktveränderungen reagiert werden als mit den herkömmlichen Studien.



Zusammenspiel der Prozesse

Diese Analysen sollen dazu dienen, arbeitslose Personen rascher und/oder nachhaltiger in den Arbeitsmarkt zu integrieren. Die Folge davon ist, dass weniger Kosten bei der Arbeitslosenversicherung anfallen.

Vorgehen

Als Beispiel für eine Einführung von unstrukturierten Datenanalysen wurden Lebensläufe von Hand analysiert. Nach einer ersten Sichtung wurde ein Kriterienkatalog erstellt. Nach einer erneuten Prüfung der Lebensläufe wurden die Kriterien konkretisiert. Als Massstab galt nicht die Aussteuerung, sondern das Finden eines Jobs. Die Frage «wie soll ein Lebenslauf optimaler Weise aussehen, damit ein Stellensuchender einen Job findet?» soll beantwortet werden.
Folgende Kriterien wurden identifiziert:

- Wenig Jobwechsel
- Letzte Anstellung dauerte länger als zwei Jahre
- Korrekte Rechtschreibung
- Grafische Elemente
- Vollständig: Personalien, Ausbildung, Berufserfahrung, Sprachkenntnisse
- Foto vorhanden
- Plausibel, ohne Lücken oder Überschneidungen

In einem nächsten Schritt wurden die technischen Möglichkeiten erörtert, mit denen diese Kriterien maschinell gesucht werden können. Die einzelnen Kriterien können nur mit einem erheblichen Programmieraufwand herausgefiltert werden. Dieser entsteht in erster Linie dadurch, die Dokumente in ihrer eigenen Struktur zu erkennen. Das heisst, die Textblöcke, welche zusammengehören, müssen als solche erkannt werden.

Resultat und Diskussion

Es erweist sich als empfehlenswert, ein grösseres Pilotprojekt zu initiieren. Für die Organisation und die Einbindung des Projektes werden verschiedene Varianten diskutiert. In Anbetracht der weiteren Möglichkeiten, unstrukturierte Daten für Analysen und Prognosen zu verwenden wird empfohlen, dieses Know-how im Betrieb aufzubauen.

Franziska Winkler