

Datenmodellierung

Neue Erkenntnisse aus komplexen Daten gewinnen

Um fundierte Entscheidungen treffen zu können, müssen KMU komplexe Geschäftsdaten in relevante Visualisierungen umwandeln, um Erkenntnisse und Trends aufzudecken und so eine bessere strategische Planung und betriebliche Effizienz zu ermöglichen. Hyperbag-Graphen sind ein Tool, das es ermöglicht, solche Datenkomplexe zu modellieren und zu visualisieren.

› Gianfranco Moi, Jean-Marie Le Goff

KMU sind zunehmend auf unstrukturierte und vielschichtige Daten angewiesen wie beispielsweise beim Kundenfeedback oder bei Betriebsprotokollen. Um solche Daten effektiv auswerten zu können, benötigen sie Tools, die über tabellarische Zusammenfassungen und statische Dashboards hinausgehen. Hyperbag-Graphen sind ein solches Werkzeug. Mit dieser leistungsstarken Methode lassen sich komplexe Daten modellieren, visualisieren und neue Erkenntnisse daraus gewinnen. KMU können diesen Ansatz für strategische Entscheidungen nutzen.

Wandelnde Voraussetzungen

In der sich ständig weiterentwickelnden Datenlandschaft stehen kleine und mittlere Unternehmen vor einer doppelten Herausforderung: Sie müssen heterogene Datensätze sinnvoll interpretieren und mit beschränkten Mitteln verwertbares Wissen extrahieren. Herkömmliche Business-Intelligence-Tools eignen sich primär für strukturierte numerische Daten. Diese reichen jedoch oft nicht aus, um Fachpersonen die Modellierung komplexer Beziehungen in textuellen, seman-

tischen oder kategorialen Daten zu ermöglichen wie beispielsweise bei der Analyse von Kundenbewertungen, Servicevorfällen oder Lieferantenfeedback. Also Datenquellen, die reich an Nuancen, aber arm an Struktur sind.

Ein neuer Ansatz – das Hyperbag-Graph (hb-graph) Framework – begegnet diesen Herausforderungen. Dessen Konzept wurde am CERN von einem Doktoranden der Universität Genf entwickelt und baut auf der Hypergraphen-Theorie auf. Anders als in Standardgraphen, bei denen Knotenpaare mit einer Kante verbunden

sind, können Hypergraphen mit einer einzigen Kante – der sogenannten Hyperkante – beliebig viele Knoten verbinden. Der hb-graph-Ansatz integriert zudem Multisets, auch Bags genannt. Im Gegensatz zu einem Set, können in einem Multiset einzelne Elemente mehrmals vorkommen. Sie können also Duplikate enthalten. Dank dieser Eigenschaften lassen sich mit dem hb-graph Framework komplexe Daten und Muster darin grafisch darstellen und genauer untersuchen.

Das Framework wurde in Zusammenarbeit mit dem CERN zu einer Open-Source-Plattform für visuelle Analysen weiterentwickelt, dem Collaboration Spotting X (CSX). Diese operationalisiert die Prinzipien von Hyperbag-Graphen für die Informationsgewinnung und Wissenserkundung. Mit CSX können tabellarische Daten in dynamisch erkundbare Netzwerke umgewandelt werden. Damit lassen sich latente Assoziationen aufdecken und semantische Entdeckungen erleichtern. Durch die Kombination einer visuellen Navigation und KI-basierter Entdeckungsmethoden bietet das hb-graph Framework eine robuste Ausgangslage für die Wissensmodellierung

Stichwort: Graphentheorie

Die Graphentheorie ist ein Teilgebiet der diskreten Mathematik und der theoretischen Informatik. Betrachtungsgegenstand der Graphentheorie sind Graphen, deren Eigenschaften und ihre Beziehungen zueinander. Graphen sind mathematische Modelle für netzartige Strukturen in Natur und Technik.

Quelle: Wikipedia

besonders für vielschichtige Datenumgebungen, wie sie in KMU häufig anzutreffen sind.

Die Wissensrepräsentation

Klassische Hypergraphen stellen Kanten als Sets dar, was die Darstellung von sich wiederholenden Elementen verhindert. In vielen realen Szenarien spielen Wiederholungen jedoch eine wichtige Rolle wie zum Beispiel, wenn ein Produkt in einer einzigen Rezension mehrfach erwähnt wird. Hyperbag-Graphen erweitern klassische Hypergraphen, indem sie Kanten als sogenannte Multisets zulassen. Dabei kann jede Hyperbag-Kante wiederholte Knoten mit zugehörigen Gewichten enthalten. Dadurch lässt sich unter anderem Folgendes ausdrücken:

- › Häufigkeit eines Konzepts innerhalb eines gegebenen Kontexts, wie in der erwähnten Rezension
- › gewichtete Beteiligung von Entitäten an einer Interaktion
- › asymmetrische Beziehungen innerhalb derselben Gruppierung

In der Analyse von Kundenfeedback kann eine Hyperbag-Kante beispielsweise eine einzelne Bewertung mit Knotenpunkten wie «Lieferung» (zweimal erwähnt), «Qualität» (einmal) und «Verzögerung» (dreimal) darstellen und so die Betonung der Stimmung auf natürliche Weise modellieren.

Die Visualisierung

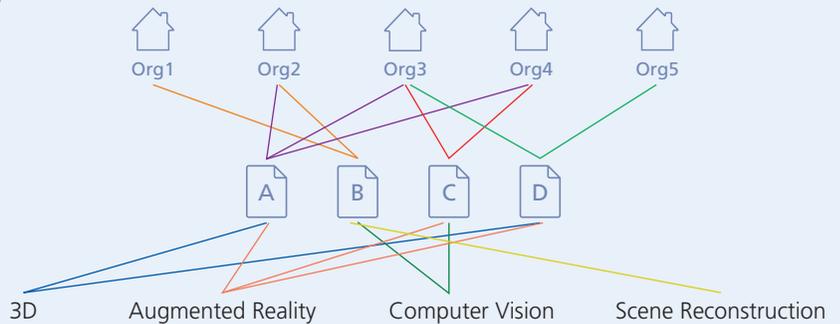
Herkömmliche Visualisierungsmethoden basieren auf Knoten-Link-Diagrammen oder Matrixansichten. Diese werden bei der Darstellung hochgradig komplexer Beziehungen oft unübersichtlich oder zu stark vereinfacht. Das hb-graph Framework führt eine facettenbasierte visuelle Analyse ein. Dabei wird der Datensatz anhand mehrerer semantischer Dimensionen – sogenannter Facetten – untersucht. Beispielsweise:

Abb. 1: Darstellung komplexer Datenmodellierung

a)

PubRef	Organisations	Keywords
PubA	['Org2', 'Org3', 'Org4']	['3D', 'Augmented Reality']
PubB	['Org1', 'Org2']	['Computer Vision', 'Scene Reconstruction']
PubC	['Org3', 'Org4']	['Augmented Reality', 'Scene Reconstruction']
PubD	['Org4', 'Org5']	['3D', 'Augmented Reality']

b)



c)

3D	{{'Org2', 'Org3', 'Org4', 'Org5'}}
Augmented Reality	{{'Org2', 'Org3', 'Org4', 'Org5'}}
Computer Vision	{{'Org1', 'Org2', 'Org3', 'Org4'}}
Scene Reconstruction	{{'Org1', 'Org2'}}

Zur Abbildung:

a) Ereignismatrix mit tabellarischen Daten eines vereinfachten Publikationsnetzwerks. Diese Matrix enthält mehr als die reinen Datenelemente. Die Datenelemente innerhalb einer Zeile sind miteinander verbunden.

b) Die Tabelle beschreibt den Zusammenhang zwischen Organisationen und

Schlüsselwörtern, die in unterschiedlichen Publikationen vorkommen. Die hochgestellte Zahl entspricht dem gleichzeitigen Auftreten von Organisationen mit dem jeweiligen Schlüsselwort als Referenz.

c) Solche Netzwerke lassen sich am einfachsten in Form von Graphen darstellen. Jedoch besteht die Gefahr, dass Informationen darin verloren gehen.

- › Themen wie Produktmerkmale,
- › Quellen wie Kundentyp,
- › Zeiträume sowie
- › Stimmungs- oder Bewertungswerte.

Jede Facette wird als separater Hyperbag-Graph dargestellt. Ein solcher weist ge-

meinsame Knotentypen wie zum Beispiel Begriffe oder Themen auf. Dadurch lassen sich mehrere Facetten vergleichen und gemeinsam untersuchen.

Eine zentrale Komponente des hb-graph Frameworks ist der sogenannte «Data-

HbEdron». Dieser fungiert als visueller Container, der diese Facetten-Graphen miteinander verknüpft. Dank diesem können Nutzende zwischen den Dimensionen wechseln, ohne den Kontext zu verlieren.

Dieses visuelle Design eignet sich besonders für KMU, da es

- › die kognitive Belastung durch die Aufschlüsselung komplexer Ansichten reduziert.
- › interaktive Filterung und semantischen Zoom unterstützt.
- › explorative Suchen ohne Datenbankabfragen erlaubt.

Die Erkenntnisgewinnung

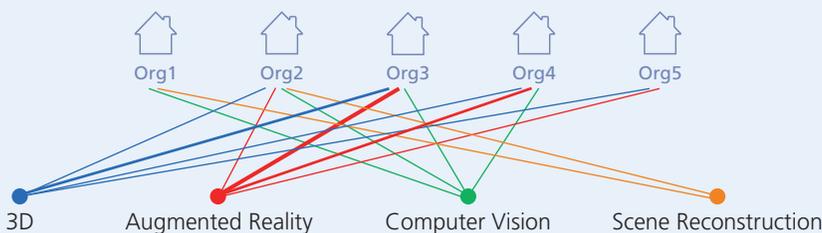
Um neue Erkenntnisse gewinnen zu können, verwendet das hb-graph Framework eine sogenannte austauschbasierte Diffusion. Das ist ein dynamischer Prozess, bei dem Bedeutung oder Relevanz basierend auf Konnektivität und Kantengewicht durch das Netzwerk fließt. Im Gegensatz zu statischen Metriken spiegelt dieser Prozess den globalen Einfluss von Knoten in bestimmten Kontexten wider.

Zusätzlich lassen sich Ergebnisse aus verschiedenen Facetten durch sogenannte Multi-Diffusion zu einem einheitlichen Ranking kombinieren. Dies ist besonders nützlich, wenn Elemente priorisiert werden sollen, die mehrere Nutzerinteressen betreffen, wie beispielsweise die Erschwinglichkeit und Qualität von Produkten oder Dienstleistungen.

Eine zusätzliche Erweiterung ist die verzerrte Diffusion. Diese ermöglicht benutzerdefinierte Gewichtungsschemata, um bestimmte Merkmale hervorzuheben. So können KMU beispielsweise die Diffusionsparameter anpassen, um Feedback von wichtigen Kundensegmenten zu priorisieren oder aktuelle Bewertungen stärker zu gewichten.

All diese Erkennungsmechanismen sind transparent und anpassbar. Wichtige Ei-

Abb. 2: Datenverbindungen durch Hyperbag-Graphen



Zur Abbildung: Mit Hyperbag-Graphen lässt sich das gemeinsame Auftreten von Schlüsselwörtern in den tabellarischen Daten genau darstellen. Die Dicke der

Kanten ist ein Mass dafür, wie häufig das jeweilige Organisation-Schlüsselwort-Paar auftritt.

Abb. 3: Die Open-Source-Plattform «Collaboration Spotting X»



Zur Abbildung: Darstellung des grafischen Userinterfaces von CSX. CSX unterstützt die dynamische Interoperabilität zwischen Grafiken, Diagrammen und Tabellen. Die Plattform bietet sowohl allge-

meine Übersichten als auch detaillierte Einblicke. Diese lassen sich auf die spezifischen Interessen und Bedürfnisse der Benutzer anpassen, um fundiertere Entscheidungen zu ermöglichen.



genschaften für das Management, wenn sie strategische Entscheide auf der Basis von interpretierbaren Datenanalysen begründen müssen.

Ein Anwendungsfall

Ein KMU, das im E-Commerce-Möbelmarkt tätig ist, erhält jedes Jahr Tausende von Produktbewertungen. Diese Bewertungen enthalten zahlreiche subjektive Informationen: Zufriedenheit mit der Lieferung, Designvorlieben, Benutzerfreundlichkeit der Produkte und viele weitere. Das Marketingteam möchte nun häufig genannte Schwachstellen und Stärken einzelner Produkte identifizieren. Ausserdem möchte es verstehen, wie diese Themen je nach Region und Saison variieren. Und schliesslich möchte es Produktverbesserungen priorisieren, die den grössten Einfluss auf die Kundschaft haben.

Mithilfe des hb-graph Frameworks lässt sich nun jede einzelne Bewertung als Hyperbag-Kante modellieren. Die Eckpunkte stehen dabei für Konzepte, die aus dem Text extrahiert wurden, beispielsweise «Montage», «Anleitung» oder «Komfort». Sogenannte Multiplizitäten kodieren die Gewichtung, also zum Beispiel wie oft ein Begriff erwähnt wird. Facetten werden erstellt für:

- › Produkttypen wie Sofa, Stuhl oder Bett,
- › Kundenstandort in Form regionaler Segmentierung und
- › Jahreszeit, beispielsweise Feedback im Winter gegenüber zu Sommer.

Mithilfe von Multi-Diffusion stellt das Unternehmen Folgendes fest:

- › «Montageanleitungen» stellen in ländlichen Regionen ein grosses Problem dar.

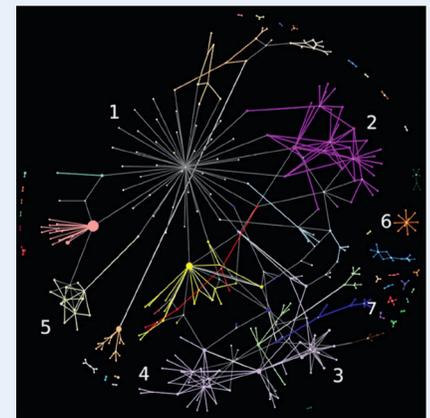
Abb. 4: Darstellung der facettenbasierten visuellen Analyse

a)



Zur Abbildung: Beide Bilder a) und b) stellen verschiedene Gruppen (1 bis 7) eines Netzwerks von Publikationen (Kanten) zu einem bestimmten Technologiethema dar, die von denselben Institutionen (Knoten) verfasst wurden. Die klassische grafische Darstellung in a) überlagert

b)



visuelle Informationen, die in der Hyperbag-Graph-Darstellung des Netzwerks in b) effektiv enthalten sind. So wird beispielsweise die Gruppe 1 in a) visuell als wichtiger wahrgenommen als Gruppe 2, während in Wirklichkeit das Gegenteil der Fall ist, wie b) korrekt darstellt.

- › Die «Stoffqualität» wird in Sommerbewertungen hoch eingestuft.
- › Die «Liefergeschwindigkeit» sorgt für positive Bewertungen bei der jüngeren Kundschaft.

Das Unternehmen nutzt diese Erkenntnisse, um die Gebrauchsanweisungen für Produkte mit hoher Priorität neu zu gestalten. Es startet eine saisonale Kampagne mit Fokus auf hochwertige Stoffe. Zudem geht es eine Partnerschaft mit einem schnelleren Kurierdienst für jugendorientiertes Marketing ein.

Implikationen für KMU

Das hb-graph Framework bietet KMU zahlreiche Vorteile. Dank einer ausdrucksstarken Modellierung lassen sich Strukturen in Daten mit Wiederholungen und Gewichten umfangreich darstellen. Die modulare Visualisierung ermöglicht navigierbare Ansichten über Facetten hinweg, wodurch die Komplexität reduziert wird. Dies legt die Basis für umsetzbare Erkenntnisse dank einer interpretierbaren Rangfolge und Erkundung durch Diffusionsmechanismen. Im Gegensatz zu Black-Box-Modellen des maschinellen Lernens legen Hyperbag-Graphen Wert auf Transparenz und Rückverfolgbarkeit, sodass KMU neue Erkenntnisse aus Daten sicher in ihre Betriebsstrategien integrieren können.

Fazit

In datenreichen und ressourcensensiblen Umgebungen benötigen KMU Tools, die nicht nur Informationen verarbeiten, sondern auch Strukturen aufzeigen und Entscheidungen leiten. Hyperbag-Graphen bieten eine mathematisch fundierte, visuelle und intuitive Lösung für die Modellierung und Untersuchung komplexer Koinzidenzdaten. Durch facettenbasierte Navigation und diffusionsbasierte Entdeckung können KMU von der Datenbeobachtung zu wissensbasierten Massnahmen übergehen – ohne den Aufwand einer komplexen Datenwissenschaftsinfrastruktur. ‹‹



Weiterführende Links

www.kmu-magazin.ch/datenmodellierung



Porträt



Gianfranco Moi

Berater, Doktorand, Universität Genf

Gianfranco Moi ist Berater für digitale Strategie an der Universität Genf und Doktorand in Sozialwissenschaften mit Schwerpunkt Informationssysteme. Als ehemaliger Leiter der Abteilung für digitale Transformation des Kantons Genf leitete er Initiativen in den Bereichen Blockchain, E-Government und Smart-City-Innovation. Seine aktuelle Forschung befasst sich mit Strategien zur Ressourcenallokation, wobei er künstliche Intelligenz zur Unterstützung resilienter komplexer Systeme nutzt. Moi ist als Direktor für Forschung und Entwicklung bei IDE4 tätig und moderiert regelmässig Podiumsdiskussionen zu KI und inklusivem digitalem Handel. Er hat einen Masterabschluss in Elektrotechnik und einen Executive Master in E-Governance von der EPFL.



Jean-Marie Le Goff

Co-CEO, Dtangle

Jean-Marie Le Goff, ehemaliger Physiker und Informatiker am CERN, ist heute Co-CEO von Dtangle und Präsident des Vereins IDE4. Am CERN leitete Le Goff die Entwicklung von Collaboration Spotting X (CSX), einem Software-Tool, das es einfacher macht, Muster, Trends und Chancen in grossen oder verstreuten Datenmengen zu erkennen. Es kombiniert Hochgeschwindigkeitsrechner mit benutzerfreundlichen Visualisierungen und hilft Teams, schnell zu verstehen, was die Daten wirklich aussagen, ohne dass ein Team von Datenwissenschaftlern erforderlich ist. CSX ist jetzt unter der MIT-2.0-Lizenz als Open Source verfügbar. Le Goff und sein Team helfen KMU dabei, CSX für reale Anforderungen wie Marktanalysen, Lieferanteninformationen, Wettbewerbsbeobachtung und Innovationsscouting einzusetzen.

Die Autoren sind SAIROP-Partner. Mit der Swiss Artificial Intelligence Research Overview Platform SAIROP bringt die Schweizerische Akademie der Technischen Wissenschaften SATW wichtige Akteure des Schweizer KI-Ökosystems zusammen und bietet einen Überblick über KI-Dienstleistungen, Forschungsprojekte und Kooperationspartner. SAIROP soll die Zusammenarbeit zwischen Forschung, Industrie und Verwaltung fördern und den Wissenstransfer in der Schweiz unterstützen. Die Schweizerische Akademie der Technischen Wissenschaften SATW ist das Experten-Netzwerk im Bereich Technikwissenschaften in der Schweiz und im Kontakt mit den massgeblichen Schweizer Gremien für Wissenschaft, Politik und Industrie.



Kontakt

gianfranco.moi@unige.ch, www.unige.ch, www.sairop.swiss

jean.marie.henri.le.goff@proton.me, www.proton.me, www.satw.ch