

Multimedia Retrieval

6 Similarity Measures

Prof. Dr. Marcin Grzegorzek

Research Group for Pattern Recognition
www.pr.informatik.uni-siegen.de

Institute for Vision and Graphics
University of Siegen, Germany



Table of Contents

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Introduction

1.1 Fundamental Concept

1.2 Search in a MMDBS

1.3 Applications of MMDBMS

2 Fundamentals of Information Retrieval

2.1 Introduction

2.2 Information Retrieval Models

2.3 Relevance Feedback

2.4 Evaluation of Retrieval Systems

2.5 User Profiles

Table of Contents

Einführung
Dist. vs.
Ähnlichkeit
Grenzen
Konkrete
Aggregation
Umwandlung
Partielle

3 Fundamentals of Multimedia Retrieval

- 3.1 Characteristics of MM Management and Retrieval
- 3.2 Processing Pipeline of a Multimedia Retrieval Systems
- 3.3 Data of a Multimedia Retrieval System
- 3.4 Features
- 3.5 Applicability of Different Retrieval Models
- 3.6 Multimedia Similarity Model

4 Transforms for Feature Extraction

- 4.1 Fourier Transform
- 4.2 Wavelet Transform
- 4.3 Principal Component Analysis
- 4.4 Singular Value Decomposition

Table of Contents

5 Distance Functions

- 5.1 Properties and Classification
- 5.2 Distance Functions for Points
- 5.3 Distance Functions for Binary Data
- 5.4 Distance Functions for Sequences
- 5.5 Distance Functions for Sets

► 6 Similarity Measures

- 6.1 Introduction
- 6.2 Distance versus Similarity
- 6.3 Range of Similarity Measures
- 6.4 Concrete Similarity Measures
- 6.5 Aggregation of Similarity Values
- 6.6 Conversion of Distances into Similarity Values
- 6.7 Partial Similarity

Einführung
Dist. vs.
Ähnlichkeit
Grenzen
Konkrete
Aggregation
Umwandlung
Partielle

Table of Contents

Einführung
Dist. vs.
Ähnlichkeit
Grenzen
Konkrete
Aggregation
Umwandlung
Partielle

7 Efficient Algorithms and Data Structures

- 7.1 High-Dimensional Index Structures
- 7.2 Algorithms for Aggregation of Similarity Values

8 Query Processing

- 8.1 Introduction
- 8.2 Concepts of Query Processing
- 8.3 Database Model
- 8.4 Languages

9 Summary and Conclusions

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Einführung

2 Distanz versus Ähnlichkeit

3 Grenzen von Ähnlichkeitsmaßen

4 Konkrete Ähnlichkeitsmaße

5 Aggregation von Ähnlichkeitsmaßen

6 Umwandlung von Distanzen in Ähnlichkeitswerte

7 Partielle Ähnlichkeit

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Einführung

- 2 Distanz versus Ähnlichkeit
- 3 Grenzen von Ähnlichkeitsmaßen
- 4 Konkrete Ähnlichkeitsmaße
- 5 Aggregation von Ähnlichkeitsmaßen
- 6 Umwandlung von Distanzen in Ähnlichkeitswerte
- 7 Partielle Ähnlichkeit

Allgemeines

- Objekte werden von Menschen als ähnlich wahrgenommen, wenn sie bei Menschen zu ähnlichen Reizen führen.
- Wir gehen hier davon aus, dass die relevanten Reize den extrahierten und aufbereiteten Feature-Werten entsprechen.
- Eine allgemein akzeptierte und exakte Definition des Begriffs Ähnlichkeit gibt es nicht.
- In Gebieten wie z. B. Psychologie, Mathematik, Statistik, Bildverarbeitung, und Mustererkennung wird zu diesem Thema geforscht und es wurden viele unterschiedliche Ähnlichkeitsmodelle entwickelt.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Definition

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Definition

Ein Ähnlichkeitsmaß ist eine Funktion, die einem Paar von Objekten eine Zahl aus dem reellen Intervall $[0, 1]$ zuordnet. Dabei korrespondiert der Wert 1 zur maximalen Ähnlichkeit und der Wert 0 zur maximalen Unähnlichkeit.

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Einführung

2 Distanz versus Ähnlichkeit

3 Grenzen von Ähnlichkeitsmaßen

4 Konkrete Ähnlichkeitsmaße

5 Aggregation von Ähnlichkeitsmaßen

6 Umwandlung von Distanzen in Ähnlichkeitswerte

7 Partielle Ähnlichkeit

Probleme mit den Distanzfunktionen

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- In der Psychologie wurden Untersuchungen durchgeführt, die belegen, dass die Distanzeigenschaften im Allgemeinen für das menschliche Ähnlichkeitsempfinden zu restriktiv sind.
- *Selbstidentität*. Diese Eigenschaft, formuliert als $d(A, A) = 0$, gilt für das menschliche Empfinden nicht grundsätzlich.
- *Positivität*. Diese Eigenschaft wurde für Menschen widerlegt.
- *Symmetrie*. Beim Ähnlichkeitsempfinden zwischen einem Suchbild und einem Datenbild macht es einen Unterschied, wenn die beiden Bilder ihre Rollen tauschen.
- *Dreiecksungleichung*. Menschen neigen oft dazu, Unterschiede zwischen Vergleichsobjekten zu hoch zu bewerten, wenn kein drittes Referenzobjekt zum Vergleich erfahrbar ist.

Beispiel 1

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Sucht man in einer Bilddatenbank ein Bild A mit einem Haus, dann akzeptiert man in der Regel ein Datenbankbild B, das neben dem Haus weitere Objekte zeigt (z. B. ein Baum).
- Sucht man jedoch anhand des Datenbankbildes B und erhält das ursprüngliche Suchbild A, wird der Baum vermisst.

Beispiel 2

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Vergleicht man ein Kreis (Objekt A) mit einem Kreuz (Objekt B), stellt man kaum Gemeinsamkeiten fest.
- Beide Objekte haben aber Gemeinsamkeiten zum dritten Objekt - einem Kreuz in einem Kreis (Objekt C)
- Der Mensch neigt zu der Bewertung:
 $d(A, B) > d(A, C) + d(B, C)$

Distanz vs. Ähnlichkeit - Schlussfolgerung

- Die geschilderten Probleme zeigen, dass Distanzeigenschaften nicht generell auf Ähnlichkeitsmaße übertragbar sind.
- In vielen Anwendungen wurden trotzdem gute Erfahrungen mit Ähnlichkeitsberechnungen auf der Grundlage von Distanzfunktionen gemacht.
- Da die Distanzeigenschaften im Allgemeinen als zu restriktiv erachtet werden, wird im Folgenden statt einer Distanz der Begriff des Ähnlichkeitsabstandes verwendet.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Ähnlichkeitsabstand nach Tversky und Gati

Einführung
Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- *Dominanz.* Der Ähnlichkeitsabstand, der mehrere Dimensionen berücksichtigt, kann nicht kleiner als der maximale Ähnlichkeitsabstand aller einzelner Dimensionen sein.
- *Konsistenz.* Die einzelnen Dimensionswerte wirken unabhängig voneinander.
- *Transitivität.* Mögliche Reihenfolgen von Objekten müssen pro Dimension transitiv wirken. Eine Reihenfolge von drei Objekten $x_1|x_2|x_3$ gilt, wenn x_2 zwischen x_1 und x_3 liegt. Die Transitivität fordert, dass wenn $x_1|x_2|x_3$ und $x_2|x_3|x_4$ gelten, dass dann auch $x_1|x_2|x_4$ und $x_1|x_3|x_4$ erfüllt sein müssen.

Ähnlichkeitsabstand nach Tversky und Gati

- Man kann nachweisen, dass diese Eigenschaften allgemeiner als die Distanzeigenschaften sind.
- Die Symmetrieeigenschaft wird für einen Ähnlichkeitsabstand nicht gefordert.
- Die drei Eigenschaften besitzen eine wichtige Eigenart. Wendet man auf den Werten eines solchen Abstandmaßes eine monoton steigende Funktion an, bleiben dadurch die Eigenschaften erhalten.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Einführung

2 Distanz versus Ähnlichkeit

3 Grenzen von Ähnlichkeitsmaßen

4 Konkrete Ähnlichkeitsmaße

5 Aggregation von Ähnlichkeitsmaßen

6 Umwandlung von Distanzen in Ähnlichkeitswerte

7 Partielle Ähnlichkeit

Allgemeines

- Maschinell berechenbaren Ähnlichkeitswerten sind im Vergleich zum menschlichen Ähnlichkeitsempfinden enge Grenzen gesetzt. Diese Grenzen sollen hier diskutiert werden.
- Ähnlichkeitswerte werden automatisch aus Feature-Werten berechnet.
- Semantische Eigenschaften in Feature-Daten sind oft holistischer Natur. Dies bedeutet, dass die interessierenden Eigenschaften nicht exakt bestimmten Feature-Werten zugeordnet werden können, sondern durch die Gesamtheit der Daten ausgedrückt werden.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Beispiel 3

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Ein MM-System soll auf beliebigen Rasterbildern Hunde erkennen. Dies ist jedoch nicht zuverlässig möglich, da die Eigenschaft "Hund" weder an einzelne Pixel noch an bestimmte Feature-Werte exakt gebunden werden kann, sondern durch die Gesamtheit der Daten ausgedrückt wird.
- Bei der menschlichen Ähnlichkeitsempfindung spielt bewusst oder unbewusst immer ein bestimmtes Weltwissen eine Rolle.

Beispiel 4

Einführung
Dist. vs.
Ähnlichkeit
Grenzen
Konkrete
Aggregation
Umwandlung
Partielle

- Wird einem geschichteinteressierten Deutschen ein Photo von Erich Honecker und eine Abbildung von einem Trabi präsentiert, ergibt sich für ihn eine gewisse Ähnlichkeit aufgrund der DDR-Zugehörigkeit.
- Ohne Weltwissen ist ein solcher Zusammenhang nicht erkennbar.
- Ein großes Problem ist daher die Subjektivität menschlich wahrgenommener Ähnlichkeit.

Beispiel 5

- Ein weit verbreiteter “Volkssport” ist das Vergleichen von Gesichtszügen eines Kindes mit denen der Eltern. Kommt das Kind mehr nach dem Vater oder nach der Mutter?
- Dabei stellt sich heraus, dass die Meinungen weit auseinanderliegen.
- Oft ist das Ergebnis solcher Ähnlichkeitstests abhängig davon, wie gut derjenige die Elternteile kennt.
- Weiterhin werden oft unterschiedliche Merkmale für den Vergleich herangezogen und unterschiedlich gewichtet.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Drei Ebenen der Inhaltsverarbeitung

Bezüglich der Verwendung von Weltwissen unterscheidet man drei Ebenen, wie der Inhalt von Medienobjekten verwaltet und eine Ähnlichkeit berechnet werden könnte:

- *Syntaktische Ebene*: Hier erfolgt die Verarbeitung rein syntaktisch, ohne dass die Bedeutung der Medienobjekte berücksichtigt wird.
- *Semantische Ebene*: Auf dieser Ebene wird die Bedeutung von Medienobjekten verwaltet und für den Ähnlichkeitsvergleich verwendet.
- *Pragmatische Ebene*: Auf der pragmatischen Ebene werden Medienobjekte interpretiert und thematisch Kategorien zugeordnet.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Pre-Attentive vs. attentive Wahrnehmung

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Bei der menschlichen Wahrnehmung von Reizen wird die sog. pre-attentive von der attentiven Wahrnehmung unterschieden (nur visuelle Reize).
- Die pre-attentive Wahrnehmung erfolgt in den ersten 250ms. In dieser Zeit ist ein Mensch noch nicht in der Lage, sein Weltwissen zur Interpretation zu nutzen.
- Dies erfolgt in der attentiven Phase.

Nachbildung der pre-attentiven Wahrnehmung

Ein Ziel der Feature-Extraktion und beim Ähnlichkeitsvergleich ist es, die pre-attentive Phase nachzubilden. Es handelt sich um Folgende Features:

- Linienorientierung,
- Länge, Breite, Größe von Objekten,
- Krümmung,
- Anzahl von Objekten, und
- Farbe und Intensität von Objekten.

Leider sind solche Features nicht immer leicht algorithmisch berechenbar.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Einführung

2 Distanz versus Ähnlichkeit

3 Grenzen von Ähnlichkeitsmaßen

4 Konkrete Ähnlichkeitsmaße

5 Aggregation von Ähnlichkeitsmaßen

6 Umwandlung von Distanzen in Ähnlichkeitswerte

7 Partielle Ähnlichkeit

Eine Auswahl von Ähnlichkeitsmaßen

- Feature-Kontrast-Modell nach Tversky
- Fuzzy-Feature-Kontrast-Modell von Santini und Jain
- **Histogrammschnitt**
- **Kosinusmaß**
- Ähnlichkeitsmaße aus der Taxonomie

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Histogrammschnitt

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Der Histogrammschnitt berechnet einen Ähnlichkeitswert zwischen zwei Histogrammen. Wir gehen von einem normalisierten Histogramm aus:

$$\sum_{i=1}^n h_a[i] = 1$$

- Der Histogrammschnitt zwischen zwei Histogrammen h_a und h_b berechnet sich nach folgender Formel:

$$S_{nH}(h_a, h_b) = \sum_{i=1}^n \min(h_a[i], h_b[i])$$

Histogrammschnitt

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Aufgrund der Normierung der Histogramme ist garantiert, dass die Ähnlichkeitswerte auf das Intervall $[0, 1]$ abgebildet werden.
- Eine Umwandlung zu einem Ähnlichkeitsabstand erfolgt mit folgender Formel:

$$d_{S_{nH}}(h_a, h_b) = 1 - S_{nH}(h_a, h_b)$$

- Der Histogrammschnitt als Ähnlichkeitsabstand ist eine Distanzfunktion.

Kosinusmaß

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Das Kosinusmaß ist auf Vektoren eines Vektorraums definiert und beschreibt den Kosinus des eingeschlossenen Winkels zwischen zwei Vektoren.
- Der Name dieses Ähnlichkeitsmaßes ergibt sich aus der Tatsache, dass über den Skalarprodukt der Kosinus des Winkels berechnet wird.

Kosinusmaß

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Wenn a und b zwei Vektoren darstellen, dann berechnet sich das Kosinusmaß folgendermaßen:

$$S_{\cos}(a, b) = \frac{\langle a, b \rangle}{\|a\| * \|b\|}$$

$$\langle a, b \rangle = a^T * b - \text{Skalarprodukt}$$

$$\|a\| = \sqrt{a^T * a} - \text{Betrag}$$

- Das Kosinusmaß ermittelt zwischen zwei beliebigen Vektoren einen Wert aus dem Intervall $[-1, 1]$. Eine Abbildung auf das Ähnlichkeitsintervall $[0, 1]$ erhält man durch eine Halbierung gefolgt von der Addition mit $\frac{1}{2}$.

Kosinusmaß

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Um aus einem Ähnlichkeitswert aus dem Intervall $[0, 1]$ einen Ähnlichkeitsabstand zu erzeugen, wird dieser Wert von 1 abgezogen:

$$d_{\cos}(a, b) = 1 - S_{\cos}(a, b) = 1 - \frac{\langle a, b \rangle}{\|a\| * \|b\|}$$

- Diese Abstandsfunktion ist eine Semi-Pseudo-Distanzfunktion bzw. eine Semi-Distanzfunktion, wenn von längennormierten Vektoren ausgegangen wird.

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- 1 Einführung
- 2 Distanz versus Ähnlichkeit
- 3 Grenzen von Ähnlichkeitsmaßen
- 4 Konkrete Ähnlichkeitsmaße
- 5 Aggregation von Ähnlichkeitsmaßen
- 6 Umwandlung von Distanzen in Ähnlichkeitswerte
- 7 Partielle Ähnlichkeit

Beispiel

Einführung
Dist. vs.
Ähnlichkeit
Grenzen
Konkrete
Aggregation
Umwandlung
Partielle

- In einer Bilddatenbank, die Abbildungen von Stoffen für die Produktion von Kleidungsstücken enthält, soll nach einem bestimmten Stoff gesucht werden.
- Vorgabe dabei ist ein bestimmtes Muster und eine bestimmte Farbe.
- Diese beiden Eigenschaften führen pro Stoffabbildung der Datenbank zu zwei Ähnlichkeitswerten, die zu einem endgültigen Ähnlichkeitswert kombiniert werden müssen.

Anforderungen

An eine Aggregation agg , die Ähnlichkeitswerte für ein Objekt aggregiert, werden bestimmte Forderungen gestellt:

1. *Ähnlichkeitswerte.* Die Funktion muss mehrere Ähnlichkeitswerte aus dem Intervall $[0, 1]$ auf einen Wert aus dem Intervall $[0, 1]$ abbilden:

$$\text{agg} : [0, 1]^n \rightarrow [0, 1]$$

2. *Monotonie.* Wenn die Eingangswerte nicht sinken, dann sinkt auch der aggregierte Ähnlichkeitswert nicht:

$$x_1 \leq y_1 \wedge \dots \wedge x_n \leq y_n \Rightarrow \text{agg}(x_1, \dots, x_n) \leq \text{agg}(y_1, \dots, y_n)$$

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Anforderungen

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

3. *Strikte Monotonie.* Wenn alle Eingangswerte wachsen, dann muss auch der entsprechende, aggregierte Ähnlichkeitswert wachsen:

$$x_1 < y_1 \wedge \dots \wedge x_n < y_n \Rightarrow \text{agg}(x_1, \dots, x_n) < \text{agg}(y_1, \dots, y_n)$$

4. *Stetigkeit.* Die Aggregatfunktion soll bezüglich der Eingangswerte stetig sein, also keine abrupten Sprünge aufweisen.

Anforderungen

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

5. *Idempotenz.* Eine Aggregation derselben Werte muss diesen Wert selbst ergeben:

$$\text{agg}(a, \dots, a) = a$$

6. *Unabhängigkeit von der Reihenfolge.* Das Resultat einer Aggregation ist unabhängig von der Reihenfolge der zu aggregierenden Ähnlichkeitswerte:

$$\text{agg}(x_1, x_2, \dots, x_n) = \text{agg}(x_{p_1}, x_{p_2}, \dots, x_{p_n})$$

wobei $[p_i]$ eine beliebige Permutation der Werte $[i]$ darstellt.

Generalisiertes Mittel

Einführung
Dist. vs.
Ähnlichkeit
Grenzen
Konkrete
Aggregation
Umwandlung
Partielle

- Das generalisierte Mittel ist definiert wie folgt:

$$\text{agg}_{gm}^{\alpha}(x_1, \dots, x_n) = \left(\frac{x_1^{\alpha} + \dots + x_n^{\alpha}}{n} \right)^{\frac{1}{\alpha}}$$

- Der Parameterwert α muss ungleich 0 sein.
- Folgende Spezialfälle ergeben sich:
 - $\alpha = 1$: arithmetisches Mittel
 - $\alpha = \infty$: maximaler Ähnlichkeitswert
 - $\alpha = -\infty$: minimaler Ähnlichkeitswert

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Einführung

2 Distanz versus Ähnlichkeit

3 Grenzen von Ähnlichkeitsmaßen

4 Konkrete Ähnlichkeitsmaße

5 Aggregation von Ähnlichkeitsmaßen

6 Umwandlung von Distanzen in Ähnlichkeitswerte

7 Partielle Ähnlichkeit

Allgemeines

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Hier beschäftigen wir uns kurz mit der Umwandlung von Distanzen in Ähnlichkeitswerte.
- Eine Umwandlungsfunktion ist eine Funktion f , die nichtnegative, reelle Werte auf das Intervall $[0, 1]$ abbildet und eine Reihe von Eigenschaften erfüllen muss.

Eigenschaften der Umwandlungsfunktion

Einführung
Dist. vs.
Ähnlichkeit
Grenzen
Konkrete
Aggregation
Umwandlung
Partielle

1. *Grenzbedingung maximale Ähnlichkeit.* Eine maximale Ähnlichkeit entspricht dem Ähnlichkeitswert 1 und einem Distanzwert 0:

$$f(0) = 1$$

2. *Grenzbedingung minimale Ähnlichkeit.* Eine minimale Ähnlichkeit entspricht dem Ähnlichkeitswert 0 und einem maximalen Distanzwert. Häufig existiert zu einer Distanzfunktion ein maximaler Distanzwert d_{\max}

$$f(d_{\max}) = 0;$$

Ansonsten fordern wir:

$$\lim_{d \rightarrow \infty} f(d) = 0$$

Eigenschaften der Umwandlungsfunktion

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

3. *Streng monoton fallend.* Eine korrekte Abbildung erfordert eine streng monoton fallende Umwandlungsfunktion:

$$x_1 > x_2 \Rightarrow f(x_1) < f(x_2)$$

4. *Stetigkeit.* Stetigkeit der Funktion f ist gefordert, um unerwünschte Sprünge zu vermeiden.

Linearkombination

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Die einfachste Variante ergibt sich aus der linearen Kombination beider Grenzbedingungen:

$$f(x) = 1 - \frac{x}{d_{\max}}$$

- Die Sensibilität (der absolute Betrag des Anstiegs) ist hier konstant hinsichtlich variierender Distanzwerte x .

Dynamische Sensibilität

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- Eine Distanzfunktion, die sensibler bei geringeren Distanzen als bei hohen Distanzen reagiert kann folgendermaßen definiert werden:

$$f(x) = e^{-x}$$

- Diese Funktion ist nicht an einen maximalen Distanzwert gebunden, da sie sich asymptotisch dem Ähnlichkeitswert 0 nähert.

Overview

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

1 Einführung

2 Distanz versus Ähnlichkeit

3 Grenzen von Ähnlichkeitsmaßen

4 Konkrete Ähnlichkeitsmaße

5 Aggregation von Ähnlichkeitsmaßen

6 Umwandlung von Distanzen in Ähnlichkeitswerte

7 Partielle Ähnlichkeit

Partielle Ähnlichkeit - Allgemeines

- Bis jetzt sind wir davon ausgegangen, dass zwei Medienobjekte als jeweils zwei ganze Einheiten miteinander verglichen werden sollen.
- Es gibt jedoch Anwendungen, bei denen die Grundannahme nicht zutrifft.
- Einige Feature-Werte des “größeren” Medienobjekts gehen in die Berechnung der Ähnlichkeit nicht ein.

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

Beispiel

Einführung

Dist. vs.
Ähnlichkeit

Grenzen

Konkrete

Aggregation

Umwandlung

Partielle

- In einer Anwendung verwaltet ein Bilddatenbanksystem Rasterbilder von Landschaftsaufnahmen.
- Eine partielle Ähnlichkeitsanfrage liegt hier vor, wenn nach Bildern gesucht wird, bei denen die Farbe rot möglichst stark vertreten ist.
- Die Häufigkeiten anderer Farben sollen ignoriert werden.
- Die Anfrage wird also durch einen Vektor repräsentiert, der weniger Dimensionen aufweist als die Medienobjekte in der Datenbank.