Seminar: Verteilung und Integration von Informationen im Verkehrsbereich

Integration sich widersprechender Datenquellen

Prudence KOUAM

Betreuer: Heiko SCHEPPERLE

Einleitung

Informationen aus dem Verkehrsbereich

- stammen aus unterschiedlichen Quellen,
- sind deswegen mit einer gewissen Unsicherheit behaftet,
- können sich widersprechen.

→ **Grad des Vertrauens muss abgebildet werden.**

Beispielszenario

Es gibt 2km Stau auf der A5.

Es gibt kein Stau auf der A5.

Wie kann man diese Informationen integrieren?

Wem kann ich vertrauen?

Integration sich widersprechender Datenquellen

- Szenario
- Konfidenzindex
- Konfidenzindex Menge (Ciset)
- Anwendung des Ciset-Ansatzes auf das Szenario
- Ciset-Relationale Algebra
- Bewertung

Szenario

Wir haben über Radio zwei Informationen zum Straßenverkehrszustand bekommen:

- Es gibt 2km Stau auf der A5.
- Es gibt kein Stau auf der A5.

Wir wollen beide Informationen speichern und verarbeiten. → Ciset-Ansatz

Wo stehen wir?

- Szenario
- **Konfidenzindex**
- Konfidenzindex Menge (Ciset)
- Anwendung des Ciset-Ansatzes auf das Szenario
- Ciset-Relationale Algebra
- Bewertung

Ciset-Ansatz

Das Konzept Ciset dient dazu, sich widersprechende Informationen zu integrieren.

Ciset = Confidence Index Set.

Definition: Sei L ein verteilte Verband und seien α, $\beta \in L$. L ist das Einheitsintervall [0, 1].

Das Paar $a = \langle \alpha, \beta \rangle$ wird **Confidence Index** (Konfidenzindex) genannt.

Wobei α die negative Aussage und β die positive Aussage darstellen.

Beispiele

Konfidenzindizes:

$\langle 0,0\rangle, \langle 1,0\rangle, \langle 1,1\rangle, \langle 0.2, 0\rangle, \langle 1,0.4\rangle, \langle 0.5,0.7\rangle.$

Keine Konfidenzindizes

$\langle 10, 0\rangle, \langle -1, 0\rangle, \langle 0,-1.2\rangle, \langle 1.5, 0.5\rangle \notin [0, 1]$

Konfidenzindex: Supremum und Infimum

Definition: sei a = $\langle\alpha, \beta\rangle$ ein Konfidenzindex. Das **Infimum** von a sei **l (a) =** α und das **Supremum** von a, sei **u (a)** = β.

Gleichheit

- Zwei Konfidenzindizes a1 = $\langle\alpha 1, \beta 1\rangle$ und a2 = $\langle\alpha 2, \beta 2\rangle$ sind **gleich,** genau dann wenn:

l(a1) = l(a2) und u(a1) = u(a2).

Konfidenzindex

Partielle Ordnung

- $a1 \prec a2$, wenn

  $l(a1) \geq l(a2)$ und $u(a1) < u(a2)$ oder

  $l(a1) > l(a2)$ und $u(a1) \leq u(a2)$

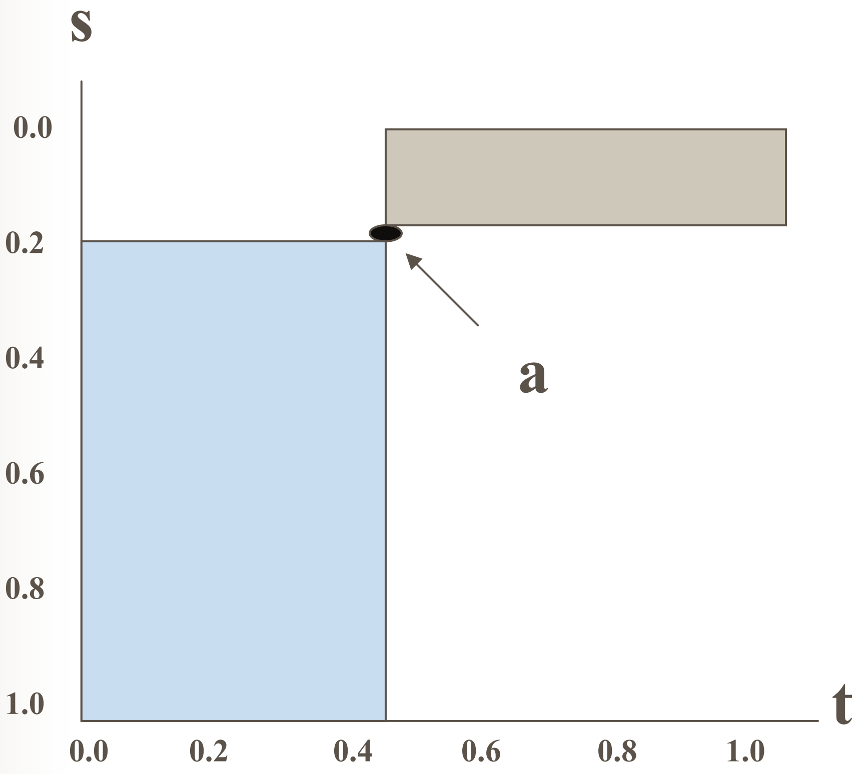
- $a1 \preccurlyeq a2$, genau dann, wenn

  $a1 \prec a2$ oder $a1 = a2$

- $a1 \succ a2$ genau dann, wenn **$a2 \prec a1$**

- $a1 \succcurlyeq a2$ genau dann, wenn **$a2 \preccurlyeq a1$**

Partielle Ordnung auf Konfidenzindex



Der schraffierte Bereich rechts von a repräsentiert die Menge alle Konfidenzindizes größer oder gleich a. Links von a steht die Menge aller Konfidenzindizes kleiner oder gleich a bezüglich $\preccurlyeq$.

Basis-Operationen auf Konfidenzindizes

Schnitt ($\cap$), Vereinigung($\cup$), Differenz(-) und Negation($\neg$) sind die Basis-Operationen auf Konfidenzindex.

Seien a1 = $\langle\alpha 1, \beta 1\rangle$ und a2 = $\langle\alpha 2, \beta 2\rangle$ zwei Konfidenzindizes.

- $a1 \cup a2 = \langle\alpha 1 \wedge \alpha 2, \beta 1 \vee \beta 2\rangle$,

wobei $\wedge$ Minimum und $\vee$ Maximum bedeuten.

- $a1 \cap a2 = \langle\alpha 1 \vee \alpha 2, \beta 1 \wedge \beta 2\rangle$,
- $\neg a1 = \langle\beta 1, \alpha 1\rangle$,
- $a1 - a2 = a1 \cap (-a2)$.

Beispiele für Basis-Operationen

Sei **Σ(L)** die Menge auf aller Konfidenzindizes aus dem kompletten Verband L.

Beispiele: Seien a1 = $\langle 0.3, 0.7\rangle$, a2 = $\langle 0.5, 0.7\rangle$, a3 = $\langle 0.4, 0.6\rangle$, a4 = $\langle 0.2, 0.8\rangle \in \boldsymbol{\Sigma}$

- a4 $\succ$ a1 $\succ$ a2 und a1 $\succ$ a3,
- a1 $\cup$ a4 = $\langle 0.3 \wedge 0.2, 0.7 \vee 0.8\rangle = \langle 0.2, 0.8\rangle$ = a4;
- a1 $\cap$ a4 = $\langle 0.3 \vee 0.2, 0.7 \wedge 0.8\rangle = \langle 0.3, 0.7\rangle$ = a1;
- a5 = a2 $\cup$ a3 = $\langle 0.4, 0.7\rangle$

  mit a5 $\succ$ a2 und a5 $\succ$ a3

Wo stehen wir?

- Szenario
- Konfidenzindex
- **Konfidenzindex Menge (Ciset)**
- Anwendung des Ciset-Ansatzes auf das Szenario
- Ciset-Relationale Algebra
- Bewertung

Confidence Index Set (Ciset)

Definitionen: Sei S eine Menge. Ein Confidence Index Set oder Ciset ist eine Abbildung F: $\mathbf{S} \rightarrow \mathbf{\Sigma}$.

Gleichheit:

Zwei Cisets F und G von S sind **gleich**, wenn :

$\mathbf{F(x) = G(x)}$ mit $x \in S$.

Man schreibt dann $\mathbf{F = G}$.

Confidence Index Set (Ciset)

Partielle Ordnung:

Seien F und G zwei Cisets auf S, sodass $F(x) \preccurlyeq G(x)$, dann ist F **Teilmenge von G** und G **Obermenge von F.**

Wenn F eine Teilmenge von G ist und es existiert ein $x \in S$ sodass $F(x) \prec G(x)$, dann ist F **echte Teilmenge Von G** und G **echte Obermenge von F**.

Basis-Operationen auf Ciset

Schnitt, Vereinigung, Differenz, Negation und **Produkt** sind die Ciset-Basis-Operationen.

Seien S eine Menge und F, G zwei Cisets auf S:

- $(F \cup G)(x) = F(x) \cup G(x)$, für alle x, Element von S.
- $(F \cap G)(x) = F(x) \cap G(x)$, für alle x, Element von S
- $(F - G)(x) = F(x) – G(x)$, für alle x, Element von S
- **$(F \times G)(x, y) = F(x) \cap G(y)$ für alle $(x, y) \in S \times S$.**
- $(-F)(x) = - F(x)$

Relationales Modell für Ciset

Ein **Modell** ist eine Abbildung einer Menge von Objekten oder Ideen der realen Welt.

Ein **Ciset relationales Modell** ist

- Teil der Datenbanken der sich widersprechender Informationen speichert.
- Hier sind die Daten meistens in Form von Tabellen dargestellt, die man „**Ciset-Relation**“ nennt.

Beispiel: Daten zum Straßenverkehrszustand können in Form einer Tabelle mit dem Namen **STÖRUNG** organisiert werden.

Ciset Relation: STÖRUNG

	S_TYPE	S_ORT	CI_ORT	S_DAUER
t1	Stau	A5	$\langle 0.3, 0.7 \rangle$	2std.
t2	Unfall	A4	$\langle 0.4, 0.6 \rangle$	1std.
t3	Sperrung	B6	$\langle 0.5, 0.7 \rangle$	24std

STÖRUNG hat vier relationale Attribute: S_TYPE, S_ORT, CI_ORT und S_DAUER.

Jede Zeile der Tabelle stellt ein Ciset-Tupel dar.

Die obige Tabelle hat drei Tupel: t_1 = {Stau, A5, $\langle 0.3, 0.7 \rangle$, 2std}, t_2 und t_3 .

Ciset-Relation

Sei U die Menge alle relationalen Attribute. Für jedes $A \in U$, repräsentiert **DOM(A)** den Definitions-bereich von A.

Sei eine Ciset Relation $R = \{A_1, \ldots, A_n\}$ eine endliche Menge von Ciset-Attributen.

Ciset Relationen können in verschiedenen Kategorien klassifiziert werden (Typ 0, Typ 1, …)

Klassifikation von Ciset-Relationen

Typ	DOM (A_i)
0	Set (Menge)
1	Ciset
2	Menge von Teilmenge einer Ciset
3	Ciset von Teilmenge einer Ciset
…	…
2_j	Menge von Teilmenge eines Bereiches der Typ (2_{j-1}), $j > 1$.
2_{j+1}	Ciset von Teilmenge eines Bereiches der Typ (2_{j-1}), $j > 1$

Beispiele

Das Beispiel STÖRUNG ist von Typ 0

S_TYP	S_ORT	CI_ORT	S_DAUER
Stau	A5	$\langle 0.3, 0.7 \rangle$	2std.
Unfall	A4	$\langle 0.4, 0.6 \rangle$	1std.
Sperrung	B6	$\langle 0.5, 0.7 \rangle$	24std

Dom(S_Typ) = Menge aller Typen;

Dom(S_ORT) = {A5, A4, B6};

Dom(CI_ORT) = $\mathbf{\Sigma}$;

Dom (S_DAUER) = Menge aller Ziffern.

Alle diese Domänen erfüllen die Definition von Typ 0.

Semantik für Ciset-Relation

Sei die folgende Tabelle eine Ciset Relation von Typ 0 :

STÖRUNG

S_TYP	S_ORT	CI_ORT	S_DAUER	CI
Stau	A5	$\langle 0.3, 0.7 \rangle$	2std.	$\langle 0.1, 0.8 \rangle$
Unfall	A4	$\langle 0.4, 0.6 \rangle$	1std.	$\langle 0.2, 0.9 \rangle$
Sperrung	B6	$\langle 0.5, 0.7 \rangle$	24std	$\langle 0.3, 0.7 \rangle$

Die Semantik des Tupels t1:

S_ORT hat ein C.I. $\langle 0.3, 0.7 \rangle$, Konfidenzindex von der ganzen Tupel t1 ist $\langle 0.1, 0.8 \rangle$.

Eigenschaften von Ciset Relationen

1. Die Reihenfolge von Attributen ist unwichtig.
2. Die Reihenfolge der Tupel in einer Ciset Relation ist unwichtig
3. Es gibt keine identischen Tupel in einer Relation
4. Ciset-Relation ist in 1.Normal-Form.

S_TYP	**S_DAUER**	**CI**	**S_ORT**	**CI_ORT**
Stau	2std.	$\langle 0.1, 0.8 \rangle$	A5	$\langle 0.3, 0.7 \rangle$
Sperrung	24std	$\langle 0.3, 0.7 \rangle$	B6	$\langle 0.5, 0.7 \rangle$
Unfall	1std.	$\langle 0.2, 0.9 \rangle$	A4	$\langle 0.4, 0.6 \rangle$

Wo stehen wir?

- Szenario
- Konfidenzindex
- Konfidenzindex Menge (Ciset)
- **Anwendung des Ciset-Ansatzes auf das Szenario**
- Ciset Relational Algebra
- Bewertung

Anwendung von Ciset auf dem Szenario

„Es gibt Stau auf der A5“ hat ein Konfidenzindex $\langle 0.3, 0.7 \rangle$.

Am Anfang hat man keine Information über der Verkehrzustand.

“Es gibt 2km Stau auf der A5” hat dann das Konfidenzindex $\langle 1.0, 0.0 \rangle$. → Kein Stau

Information von erster Quelle.

Es gibt 2km Stau auf der A5.

Vertrauensgrad in Quelle $\beta = 0,7$

CI ist $\langle 1.0, 0.7 \rangle$.

$$\langle 1.0, 0.0 \rangle \cup \langle 1.0, 0.7 \rangle = \langle 1.0, 0.7 \rangle$$

Information von zweiter Quelle.

Es gibt keinen Stau auf der A5

Vertrauensgrad in Quelle $\alpha = 0.3$ und das CI ist $\langle 0.3, 0.7 \rangle$

$$\langle 1.0, 0.7 \rangle \cup \langle 0.3, 0.7 \rangle = \langle 0.3, 0.7 \rangle$$

Wo stehen wir?

- Szenario
- Konfidenzindex
- Konfidenzindex Menge (Ciset)
- Anwendung von Ciset auf das Szenario
- **Ciset-Relationale Algebra**
- Bewertung

Ciset-Relationale Algebra

- **Vereinigung**
- **Schnitt**
- **Select**
- Project
- Join
- …

Ciset-Relationale Algebra

Definition: Zwei Relationen R = $\{A_1, \ldots, A_n\}$ und S = $\{B_1, \ldots, B_m\}$ sind **kompatibel**, wenn:

1. n = m, d.h. die Anzahl der Attribute ist gleich;
2. DOM (A_i) = DOM (B_i), für i = 1, 2, …, n.

REL_R

A	B	C	CI
a1	b1	c1	$\langle 0.3, 0.7 \rangle$
a1	b1	c2	$\langle 0.6, 0.2 \rangle$
a3	b3	c3	$\langle 0, 1 \rangle$
a4	b4	c4	$\langle 1, 1 \rangle$

REL_S

A	B	C	CI
a1	b1	c2	$\langle 0, 0.8 \rangle$
a2	b2	c1	$\langle 0.1, 0.8 \rangle$
a3	b3	c3	$\langle 1, 1 \rangle$
a4	b4	c4	$\langle 0, 0 \rangle$

Vereinigung und Schnitt

REL_R $\cup$ REL_S

A	B	C	CI
a1	b1	c1	$\langle 0.3, 0.7 \rangle$
a1	b1	c2	$\langle 0, 0.8 \rangle$
a2	b2	c1	$\langle 0.1, 0.8 \rangle$
a3	b3	c3	$\langle 0, 1 \rangle$
a4	b4	c4	$\langle 0, 1 \rangle$

REL_R $\cap$ REL_S

A	B	C	CI
a1	b1	c2	$\langle 0.6, 0.2 \rangle$
a3	b3	c3	$\langle 0, 0.7 \rangle$
a4	b4	c4	$\langle 1, 0 \rangle$

Select

REL_R

A	B	C	D
a1	b1	c1	⟨0.3, 0.7⟩
a1	b1	c2	⟨0.6, 0.2⟩
a3	b2	c1	⟨0.8, 0⟩
a1	b1	c3	⟨0.4, 0.9⟩

$\sigma_{A=a1}$ (REL_R)

A	B	C	D
a1	b1	c1	⟨0.3, 0.7⟩
a1	b1	c2	⟨0.6, 0.2⟩
a1	b1	c3	⟨0.4, 0.9⟩

Select

A	B	C	D
a1	b1	c1	⟨0.3, 0.7⟩
a1	b1	c2	⟨0.6, 0.2⟩
a3	b2	c1	⟨0.8, 0⟩
a1	b1	c3	⟨0.4, 0.9⟩

$$\sigma_{(A=a1) \wedge (D.LL<0.5) \wedge (D.UL \geq 0.7)}(REL\_R)$$

A	B	C	D
a1	b1	c1	⟨0.3, 0.7⟩
a1	b1	c3	⟨0.4, 0.9⟩

LL = Lower Index und UL = Upper Index

Wo stehen wir?

- Szenario
- Konfidenzindex
- Konfidenzindex Menge (Ciset)
- Anwendung von Ciset auf das Szenario
- Ciset Relational Algebra
- **Bewertung**

Bewertung

Ciset Ansatz

- ist zu komplex
- Vertrauensgrad ist subjektiv
- Berücksichtigt auch widersprüchliche Daten.
- statt $\langle 1.0, 0.0 \rangle$ wie im Buch, kann man auch auf $\langle 1.0, 1.0 \rangle$ oder $\langle 0, 0 \rangle$ initialisieren

  → Begründung für $\langle 1.0, 0.0 \rangle$ unklar.

ENDE!

Fragen?