Fuzzy Logik

Christian Wurm
cwurm@phil.hhu.de

April 22, 2024

Abstract

Skript zum gleichnamigen Seminar, ist noch nicht fertig und ändert sich jede Woche.

Contents

1 Einleitung, Motivation etc.

Vagheit ist ein Phnomen das *extrem* geläufig ist in natürlicher Sprache. Es fällt in der Tat schwer, ein Wort zu finden was einen Gegenstand/Eigenschaft unseres Alltags bezeichnet, das nicht vag ist. Adjektive wie `blau`, `groß`, `breit`, `sympathisch`, weiterhin Substantive wie `Stuhl`, `Tisch`, Verben wie `lieben`, `klopfen` können alle mit gutem Grund als vage bezeichnet werden. Wir verstehen uns aber trotzdem sehr gut!

Fuzzy Logik ist eine Familie von Logiken, die mit Wahrheitswerten in $[0, 1]$ operieren; man kann mit ihnen zeigen, dass wir auch mit vagen Prädikaten valide Argumente machen können. Hierbei gilt: wir operieren mit gradueller Wahrheit, wobei graduelle Wahrheit (in $[0, 1]$) eine Generalisierung ist von klassischer Wahrheit (in $\{0, 1\}$). Wir generalisieren also zum nicht-binären Fall, und daraus folgt unmittelbar: Fuzzy Logik **generalisiert** klassische Logik.

Anders gesagt:

- Jede Inferenz, die in Fuzzy Logik gültig ist, ist auch gültig in klassischer Logik. Denn als Spezialfall in Fuzzy Logik haben wir immer auch die Wahrheitswerte $\{0, 1\}$, und was in diesem Fall gilt, soll in jedem Fall auch klassisch gelten!

- Das Gegenteil gilt nicht: es gibt gewisse klassische Inferenzen, die nicht gltig sind, wenn wir beliebige Wahrheitswerte in $[0, 1]$ annehmen!

- Also: wenn wir CL die Menge der klassischen Inferenzen nehmen (klassische Logik), und FL eine (beliebige) Fuzzy Logik (deren Inferenzen), dann gilt: FL $\subseteq$ CL.

Das ist sehr wichtig: zur Erinnerung:

Lemma 1 *Klassische Logik (kurz (CL) ist* ***maximalkonsistent****, das bedeutet: nimm an, wir haben eine Logik L so dass $CL \subsetneq L$, und L ist abgeschlossen unter Substitution. Dann ist L inkonsistent.*

Um das zu verstehen (und zu beweisen), folgt man am besten diesem Gedankengang:

- Jede Implikation in klassischer Logik entspricht einer (Un)gleichung in Booleschen Algebren.

- Eine (Un)gleichung ist wahr in allen Booleschen Algebren, gdw. sie wahr ist in der eindeutigen Booleschen Algebra über $\{0,1\}$
- Wenn man nun eine (Un)gleichung hinzufügt, die nicht gültig ist in der Algebra über $\{0,1\}$, dann folgt daraus $0=1$ (sonst wäre sie ja bereits gültig)
- Die Algebra wird also trivial (hat nur ein Element), d.h. *jede* (Un)gleichung ist darin wahr.
- Die logische Entsprechung dieser Tatsache ist *Inkonsistenz*.

Nachdem wir diese formale Seite geklärt haben, betrachten wir einige Beispiele. Das Zeichen $\because$ steht hierbei für Schlussfolgerung.

Wenn Essen schwarz ist, dann ist es angebrannt
Wenn Essen verbrannt ist, dann soll man es nicht mehr essen
Das Essen hier ist schwarz
$\because$ Das Essen hier soll man nicht essen

Dieses Argument ist klassisch sicher gültig (wir vereinfachen etwas):

$$\{p \to q, q \to r, p\} \models r$$

Das bedeutet soviel wie: wenn alle Prämissen links von $\models$ den Wert 1 haben, dann hat auch r den Wert 1. Nun sind aber alle Prämissen nicht *kategorisch* wahr:

- *Wenn Essen schwarz ist, dann ist es angebrannt* – stimmt nur, wenn das Essen nicht schon vor dem Kochen schwarz war (schwarze Bohnen)
- *Wenn Essen verbrannt ist, dann soll man es nicht mehr essen* – auch das stimmt nur bedingt: viele Sachen (Steak, Braten, Bratkartoffeln) werden immer leicht angebrannt. Das mag ungesund sein, aber es schmeckt!
- *Das Essen hier ist schwarz* – diese Aussage wird oft gemacht, wenn das Essen nicht vollstndig schwarz ist, sondern nur in bedeutenden Teilen. Wenn ich also die Kartoffeln anbrennen lasse, bleiben sie oft in weiten Teilen gelb (es sei denn sie verkohlen).

Unser Argument muss also anders aussehen:

	Wenn Essen schwarz ist, dann ist es angebrannt	0.7
	Wenn Essen verbrannt ist, dann soll man es nicht mehr essen	0.9
	Das Essen hier ist schwarz	0.65
$\because$	Das Essen hier soll man nicht essen	x

Wir haben also numerische Werte von Prämissen, und möchten wissen, welchen numerischen Wert wir der Konsequenz zuweisen können – das ist es was Fuzzy Logik leisten soll! Welchen Wert sollte x also intuitiv haben? Schwer zu sagen, wir werden das nachher prüfen!

2 Was ist Vagheit? Kriterien und verwandte Phänomene

Was ist Vagheit? Wir haben oben bereits einige Beispiele gesehen. Aber reicht das um ein vollständiges Verständnis zu haben? Es gibt insbesondere drei Phänomene, die verwandt sind:

1. Ambiguität
2. Allgemeinheit
3. Unsicherheit

Ambiguität Der große Unterschied von Ambiguität und Vagheit besteht darin, dass Ambiguität **diskret** ist: die ambigen Lesarten eines Wortes sind *endlich viele* und *klar getrennt*. Klassische Beispiele sind `Flügel`, `Schimmel`, `Bank`. Man nennt die getrennten Teilbedeutungen dann **Lesarten**. Ambiguität und Vagheit treten oft zusammen auf: z.B. ist die `Bank`-Lesart "Sitzgelegenheit" durchaus vage. Das Wort `Gruppe` ist ambig, mit seiner umgangssprachlichen und mathematischen Bedeutung (Algebra mit assoziativer Operation, neutralen und inversen Elementen); eine Lesart ist vage, eine nicht.

Allgemeinheit Die Bedeutung von `Fahrzeug` ist vielleicht vage, aber sicher allgemein. Die beiden Eigenschaften sind hier wieder verschränkt. Der Unterschied lässt sich wieder am besten mit mathematischen Beispielen zeigen: der Begriff `Algebra` ist extrem allgemein, es fallen darunter alle Gruppen, Monoide, Boolesche Algebren, Körper, etc. Aber: er ist kein bisschen vage. Er hat klar definierte Ränder. Was darin ist, ist voll und ganz darin. Für ein Fahrzeug kann man das nicht unbedingt sagen: auf ein Bobby Car mit festen Rädern appliziert sich der Begriff sicher weniger als auf ein Auto.

Unkenntnis Nimm das bekannte Wort `bärschnell`. `bärschnell` bedeutet soviel wie: schneller als jeder Bär bis zum Jahr 2000. Die zeitliche Festlegung ist wichtig, denn so kann sich die Bedeutung nicht über die Zeit verändern. Das ist sicherlich kein vager Begriff: er ist eindeutig definiert! Das Problem ist: wir keinen seine (eindeutig definierte) Bedeutung nicht. Es handelt sich hier durchaus um ein grundsätzliches Problem: wir kennen sie nicht, werden sie niemals kennen und können sie auch niemals kennen. Dennoch ist sie nicht vage im engeren Sinn: denn wir kennen die Bedeutung von `rot` – sie ist eben vage! Man muss also unterscheiden zwischen a) *vagen Bedeutungen die wir kennen*, und b) *nicht-vagen Bedeutungen die wir nicht kennen*. Um ein etwas lebensnahes Beispiel zu bringen: nur weil Sie womöglich eine etwas vage Vorstellung von Algebra haben, ist Algebra *kein* vager Begriff.

Versuch einer Definition von Vagheit Wir versuchen nun das Phänomen zu definieren. Es gibt drei Hauptkriterien für Vagheit. Sei P ein beliebiges Prädikat; wir sagen P ist vage falls es alle folgenden Bedingungen erfüllt:

1. Es gibt Grenzfälle, also Objekte a wo nicht klar ist, ob $P(a)$ oder $\neg P(a)$ wahr ist.

   Z.B. gibt es große Menschen, nicht große Menschen, und Menschen, bei denen es nicht klar ist.

2. Es gibt keine klare Grenze auch nicht für die Menge der Objekte a, für die $P(a)$ wahr ist.

   Das bedeutet auch die klaren Fälle sind von den Grenzfällen nicht klar abgegrenzt.

3. P erlaubt die Konstruktion von Sorites Reihen.

   Das bedeutet: es gibt einen klaren Fall wo $P(a)$ wahr ist; ich kann eine minimale Änderung an a vornehmen zu a', wobei die Änderung $[-]'$ (scheinbar)die Wahrheit präserviert. Es gibt aber auch ein $a^{'\cdots'}$

Wir schauen uns die Kriterien nochmal im Detail an.

1 Das sollte klar sein: es gibt die unklaren Ränder. Z.B.: `Algebra` ist sehr allgemein, aber es ist vollkommen klar was eine Algebra ist und was nicht. Man kann nicht sagen: eine Boolesche Algebra ist mehr eine Algebra als ein freier Monoid. Ein solches Gefälle gibt es bei vagen Prädikaten immer: es gibt typische Fälle und weniger typische Fälle (z.B.: das ist sicher rot, das etwas weniger)

2 Dieses Kriterium ist ebenfalls wichtig und von 1 zu unterscheiden. Man könnte z.B. sagen: wir erfinden ein Wort **Palgebra**. Das trifft voll zu auf jede Algebra. Wenn wir nun eine partielle Algebra nehmen (also mit partiellen Funktionen), dann wird es immer unwahrer, je partieller die Funktion ist. Das Ergebnis ist nicht vage nach diesem Kriterium (wohl aber nach dem ersten): denn ein vages Prädikat wie `rot` hat die Eigenschaft, dass es selbst unklar ist, ab wann etwas nicht mehr eindeutig rot ist. Also die Grenze des Grenzbereichs selber ist unklar. Ein anderes Beispiel ist `groß`: Ab wieviel cm ist ein Mensch nicht mehr eindeutig groß? Hierauf gibt es keine Antwort. Dieses Kriterium können wir z.B. anwenden auf `bärschnell`: es gibt gemessene Geschwindigkeiten von Bären. Jenseits dieser Geschwindigkeit wird es fraglich.

3 Sorites Reihen sind benannt nach dem berühmten *Sorites Paradox*. Das besagt: 10.000 Körner sind ein Haufen. Wenn ich nun ein Korn von einem Haufen wegnehme, bleibt es ein Haufen. Also, mit iteriertem Modus Ponens, bekomme ich: 9.999 Körner sind ein Haufen, ..., 1 Korn ist ein Haufen. Das ist offensichtlich falsch, also wahr hier ein Fehler aber wo? Die Antwort von Fuzzy Logik lautet: der Satz *Wenn ich nun ein Korn von einem Haufen wegnehme, bleibt es ein Haufen* ist nicht ganz wahr: er hat nur einen hohen Wahrheitswert, aber nicht 1. Dementsprechend sinkt mit jeder Applikation von Modus Ponens der Wahrheitswert der Konklusion! Eine Sorites-Reihe ist nun einfach eine Reihe $a_1, ..., a_n$, so dass

1. $P(a_1)$ sicher wahr ist.

2. a_i und a_{i+1} sich nur minimal unterscheiden, nach welchem Kriterium auch immer. Wichtig ist: egal wie wir das Kriterium setzen, wir finden immer eine Reihe! Und wichtig ist: es ist bis auf weiteres wahr, dass ein minimaler Unterschied von a_i zu a_{i+1} die Wahrheit von P **präserviert**.

3. $P(a_n)$ sicher falsch ist.

Solche Reihen gehören also mit zur Definition von Vagheit. Wir können z.B. eine Reihe von Männern uns vorstellen, von denen der erste groß ist, es immer nur einen minimalen Unterschied von zwei Nachbarn gibt, und der letzt ist nicht groß. Oder: eine Reihe von Autos, das erste ist rot, das letzte nicht etc.

Für `bärschnell` gilt das Kriterium nicht: die Aussage:

> Falls a bärschnell ist, a' nur minimal langsamer (um wie wenig auch immer), dann ist auch a' bärschnell.

ist ohne jeden Zweifel vollkommen falsch; man kann ein klares Gegenbeispiel konstruieren.

Frage Wenn ich einen Haufen Körner habe und dann ein Korn wegnehme und wieder drauf lege, wie sehr ist es dann ein Haufen?

Das ist ganz interessant: denn ein Korn wegzunehmen reduziert die Haufenhaftigkeit des Haufens, formal:

$$h(x) \to h(x-1) \text{ hat Wahrheitswert } 0.999$$

(oder so), aber nicht 1. Das Gegenstück gilt nicht: $h(x) \to h(x+1)$ ist zwar sicher wahr (hat Wahrheitswert 1), aber damit wird der Haufen nicht haufenhafter, sondern seine Haufenhaftigkeit bleibt nur bestehen: die Prämisse verlangt ja bereits einen Haufen. Das relevante Axiom wäre:

$$\neg h(x) \to \neg h(x+1) \text{ hat Wahrheitswert } 0.999$$

Sprich: wenn ich ein Korn dazulege, dann reduziere ich die Nicht-Haufenhaftigkeit des (Nicht-)Haufens. Da ein Haufen ein Nicht-Nicht-Haufen ist, haben wir damit die Haufenhaftigkeit erhöht.

Soweit die Theorie. Funktioniert das auch in der Praxis? Müssen wir nachher ausprobieren, bitte erinnern!

3 Alternative Ansätze zu Vagheit

Wahrscheinlichkeit ist kein wirklich vernünftiger Ansatz zu Vagheit. Es gibt einen prominenten Ansatz, der zumindest in philosophisch-linguistischen Kreisen große Beliebtheit hat: Supervaluation. Ich bin kein Experte darin (bzw. hab schon wieder alles vergessen), aber die Grundidee ist relativ einfach.

- α ist *superwahr* in M, falls für alle σ gilt: $M, \sigma \models \alpha$
- α ist *superfalsch* in M, falls für alle σ gilt: $M, \sigma \not\models \alpha$
- α ist super-undefiniert in M, falls es weder superwahr noch superfalsch ist.

Hier betrifft σ normalerweise die Interpretation von Prädikaten. Es gibt also eine Menge Σ von Interpretationen, und

(1) $$\bigcap_{\sigma \in \Sigma} \sigma(P)$$

ist die Menge der Objekte, auf die das Prädikat P (z.B. `rot`) in jedem Fall zutrifft;

(2) $$\bigcup_{\sigma \in \Sigma} \sigma(P)$$

ist die Menge der Objekte, auf die das Prädikat P (z.B. `rot`) möglicherweise zutrifft.

Daraus folgt dann z.B.: $P(x) \vee \neg P(x)$ ist superwahr, aber $P(x) \wedge \neg P(x)$ ist nicht superfalsch. Das ganze funktioniert einigermaßen, hat aber zwei Probleme:

- Es scheint dass es Kriterium 2 von unserer Definition von Vagheit verletzt – nämlich genau der Begriff von superwahr macht das!
- Es gibt eine wunderschöne mathematische, entscheidbare und effektive Theorie von Fuzzy Logiken. Für Supervaluationen gibt es meines Wissens nach nichts vergleichbares. Es bleibt hier bislang ein philosophisches Konzept, dass man nicht effektiv anwenden kann.

4 Fuzzy Logik und Wahrscheinlichkeit

Es ist hier vielleicht sinnvoll, sich klar zu machen, wie sich FL von Wahrscheinlichkeit unterscheidet. Die Unterschiede sind im wesentlichen zwei:

1. Sei P eine Wahrscheinlichkeitsverteilung über eine Menge M. Das bedeutet:

   (3) $P : M \to [0, 1]$

   (das wäre auch korrekt für ein fuzzy Prädikat), v.a. aber:

   (4) $\sum_{x \in M} P(x) = 1$

   Sprich: die Summe aller Wahrscheinlichkeiten ergibt 1! Das würde in unserem Fall bedeuten:

   - wenn wir P als fuzzy Prädikat lesen wie `rot`, M als die Domäne (unser Universum),
   - dann könnte es *höchstens ein* Objekt geben, das wirklich rot ist – und alle anderen wären es nicht!
   - Außerdem: je roter ein Objekt wird, desto weniger rot werden die anderen!

   Fuzzy Prädikate verhalten sich anders; nimm an R ist so ein Prädikat. Es kann beliebig viele $x \in M$ geben so dass $R(x) = 1$. Es gibt keine Beschränkung an $\sum_{x \in M} R(x)$ – warum auch? Dadurch dass viele Dinge rot sind, werden rote Dinge nicht weniger rot!

2. Wahrscheinlichkeiten sind **nicht wahrheitsfunktional**: Dadurch, dass wir die Wahrscheinlichkeit $P(x), P(y)$ kennen, kennen wir nicht die Wahrscheinlichkeit

   (5) $P(x \cap y) = P(x)P(y|x)$

   (6) $P(x \cup y) = P(x) + P(y) - P(x \cap y)$

   Ich brauche in beiden Fällen zusätzliche Informationen, ohne die ich die gefragt Wahrscheinlichkeiten nicht errechnen kann. Also:

> $P(x \cap y)$ kann niemals aufgefasst werden als eine reine Funktion von $P(x), P(y)$, ebensowenig wie $P(x \cup y)$.

Als Beispiel: nimm an, $P(x) = P(y)$, aber $x \neq y$. Dann gilt natürlich: $P(x \cap x) = P(x) \neq P(x \cap y)$.

Fuzzy Logik, auf der anderen Seite, **ist wahrheitsfunktional**. Sprich: wir haben Funktionen

$$\wedge, \vee, \rightarrow [0,1] \times [0,1] \rightarrow [0,1],$$

die interessieren sich nicht für die "Inhalte" einer Proposition, sondern nur für deren Wahrheitswert. Das hat insbesondere zwei Konsequenten:

(a) Fuzzy Logik ist eher heuristischer Natur als der Wahrscheinlichkeitskalkül (der ja nichts weniger ist als eine komplette Theorie von Rationalität)

(b) Fuzzy Logik ist aber viel einfacher und effektiver berechenbar. Während selbst die Berechnung einfacher Wahrscheinlichkeit eine große Zahl von Annahmen und Weltwissen benötigen, können Inferenzen in Fuzzy Logik mechanisiert werden. Das macht es so interessant für viele Anwendungen.

5 Fuzzy Logik

Fuzzy logic is neither a poor man's logic, nor a poor man's probability.
Petr Hajek

5.1 Einleitung

Fuzzy Logik beschreibt Logiken, die entstehen, wenn man nicht mehr davon ausgeht, dass Aussagen wahr (1) oder falsch (0) sind, sondern einen beliebigen Wahrheitswert in [0,1] annehmen können. Der Name kommt eben daher: unsere Wahrheitswerte sind nicht mehr hart und scharf, sondern eben unscharf oder "fuzzy". Man darf sich aber nicht täuschen:

- Fuzzy Logik hat unscharfe, "weiche" Wahrheitsverhältnisse zum Gegenstand;
- die mathematische Theorie der Fuzzy Logik ist aber scharf und präzise wie die Theorie jeder anderen Logik auch.

Insbesondere werden wir (vielleicht) sehen, dass sich Fuzzy Logiken in den größeren Zusammenhang der *substrukturellen Logiken* einordnen lassen. Das bedeutet, dass Fuzzy Logiken "ganz gewöhnliche" Logiken sind, die mit den gewöhnlichen Werkzeugen der Metalogik analysiert werden können (Beweiskalküle, algebraische Semantik etc). Es gibt mehrere Ausgangspunkte für die Fuzzy Logik; einer ist die Arbeit des polnischen Logikers Łukasiewicz über mehrwertige Logik aus den 1930er Jahren. Eine andere Wurzel ist die Arbeit über fuzzy Mengen, die in die 1960er Jahre zurückgeht (Lotfi Zadeh). Das Standardwerk zum Thema, an das ich mich im Zweifelsfall halte, ist "Metamathematics of Fuzzy Logic" von Petr Hajek.

Fuzzy Logiken haben mittlerweile eine sehr große Bandbreite von Anwendungen. Die Hauptanwendung ist aber nach wie vor die **Kontrolle von physischen Systemen** (damit ist gemeint: Systemen, die in der "echten Welt" funktionieren müssen), so etwa Klimatisierungen, U-Bahnen (erste große Anwendung in der Nanboku-Linie, Sendai, Jp.), Roboter, Fabriken etc.

5.2 Ein Beispiel: Klimatisierung

Die Aufgabe einer Klimatisierung ist vor der Hand einfach:

- Sie geben eine Zieltemperatur ein;
- das System soll heizen/kühlen, um die Zieltemperatur zu erreichen.

Seltsamerweise ist die Sache dann doch wesentlich komplizierter. Das liegt unter anderem an folgenden Punkten:

1. Tagüber sollte es allgemein wärmer sein als nachts.
2. Dieselbe Raumtemperatur wird als wärmer empfunden, wenn die Sonne scheint.
3. Wenn jemand die Temperatur (deutlich) hinuntersetzt, möchte er einen Kühlungseffekt. Aber: üblicherweise wird die Senkung übertrieben, und später wieder korrigiert. In der Zwischenzeit wird aber relativ viel Energie verschwendet!
4. Wenn jemand die Temperatur minimal korrigiert, dann ist derjenige interessiert an einer exakten Temperatur, keiner schnellen Korrektur. Mit diesem Wissen lässt sich Energie sparen.
5. Wird die Temperatur häufig geändert, sollte sie sensibler reagieren als andernfalls.
6. Gibt es starke Variationen in der gemessenen Temperatur, deutet das auf häufige Nutzung hin – die Kontrolle sollte also sensibler reagieren als andernfalls!

NB: es handelt sich hier nicht um Kleinigkeiten: Klimatisierung ist ein wichtiger Faktor des globalen Energieverbrauchs. Die Klimatisierung eines großen Bürogebäudes ist ein bedeutender Kostenfaktor für jede Firma!

Wir haben nun folgende Situation: alle 6 Faktoren spielen eine Rolle, und lassen sich relativ leicht als Eingaben für das System nutzen. Aber: keine der Eingaben ist binär; sie treffen alle mehr oder weniger zu. Außerdem *addieren* sich Effekte teilweise, teilweise heben sie einander auf. Das bedeutet: für unser Kontrollsystem ist jeweils wichtig, nicht nur ob, sondern auch **in welchem Maße** eine Bedingung erfüllt ist. Jede Handlung soll berücksichtigen, in welchem Maße alle Bedingungen erfüllt sind. Das ist aber gar nicht einfach, denn eine normale Logik kommt dafür nicht in Betracht!

6 Zwei Begriffe von Fuzzy Logik

Es gibt zwei Bedeutungen von Fuzzy Logik, nämlich einen engeren Sinn und einen weiteren Sinn.

1. Fuzzy Logik im weiteren Sinn behandelt Fuzzy Mengen, und wie man aus gewissen Prämissen (in diesem Sinne) gewisse Folgerungen ziehen kann.

2. Fuzzy Logik im engeren Sinne behandelt tatsächlich eine Logik, also eine formale Sprache mit 1. einer Semantik (formale Interpretation) und 2. einer Beweistheorie, also einem Kalkül der erlaubt, alle validen Inferenzen aufzulisten.

1. ist relativ interessant für viele praktische Anwendungen. Allerdings sind die of sehr spezifisch und erfordern massgefertigte Systeme. Dagegen haben die Logiken in 2. eine sehr allgemeine, schöne Theorie (und auch hier gibt es Anwendungen).

Wir gehen deswegen wie folgt vor: zunächst betrachten wir Fuzzy Mengenlehre im Allgemeinen, da das ein relativ übersichtliches Feld ist. Als nächstes betrachten wir Fuzzy Logik im engeren Sinne als ein Feld der Logik. Zuletzt, wenn noch Zeit ist, betrachten wir konkrete Anwendungen; dabei werden wir wieder eher Fuzzy Mengenlehre nutzen.

7 Fuzzy Mengenlehre

7.1 Gundlagen

Fuzzy Mengen sind eine natürliche Verallgemeinerung der Mengenlehre. Mengen sind darüber bestimmt, welche Objekte ein Element der Menge sind und welche nicht. Ein nützlicher Begriff ist hier der der **charakteristischen Funktion** einer Menge:

Definition 2 *Sei M eine Menge; dann ist die charakteristische Funktion χ_M definiert durch*

$$\chi_M(x) = \begin{cases} 1, \textit{ falls } x \in M \\ 0 \textit{ andernfalls} \end{cases}$$

Wenn man nun Mengen zu Fuzzy Mengen erweitert, erweitert man einfach die Menge der Werte: anstelle $\{0, 1\}$ nimmt man die Menge $[0, 1]$, also eine Menge von Wahrheitsgraden. In diesem Sinne generalisieren fuzzy Mengen (diskrete) Wahrscheinlichkeiten: man bildet eine Menge nach $[0, 1]$ ab, allerdings ohne sich weiter um Konsistenz zu kümmern. Also formal gesprochen: eine **fuzzy Menge** M ist eine Funktion

$$M : X \to [0, 1]$$

Man kann sich dass in etwa so vorstellen (intuitiv):

"$M(x) = \alpha \in [0, 1]$" entspricht "$x \in M$ hat den **Wahrheitsgrad** α"

Hier sieht man, dass fuzzy Mengenlehre durchaus etwas mit Logik im engeren Sinne zu tun hat: klassische Mengenlehre entspricht ja der klassischen Logik; fuzzy Mengenlehre entsteht, wenn man annimmt dass Aussagen wie $x \in M$ Wahrheitsgrade in $[0, 1]$ haben anstatt Werte in $\{0, 1\}$.

Eine interessante Frage ist: was ist X? Wir wollen ja den Definitionsbereich von M nicht von vornherein beschränken; wir müssen also verlangen, dass X die universelle Klasse ist (denn es gibt keine universelle Menge). Wir verlangen weiterhin noch dass

$$M^{-1}(0, 1] = \{x : M(x) > 0\}$$

eine *Menge* im technischen Sinn ist. Man kann dieses Problem umgehen, indem man eine **Referenzmenge** festlegt, und fuzzy Mengen nur für die Referenzmenge definiert.

Wir kommen nun zur ersten wichtigen Definition: fuzzy Mengen sind intendiert als Generalisierung von normalen Mengen; in der fuzzy Mengenlehre nennen wir normale Mengen **knackig**. Also gilt: eine fuzzy Menge M ist knackig, wenn

$$M[X] = \{0, 1\};$$

d.h. jedes Objekt wird auf 0 oder 1 angebildet. Damit haben wir, was man eine **charakteristische Funktion** ist:

$$x \text{ ist ein Element von } M, \text{ gdw. } M(x) = 1.$$

Man kann knackige Mengen mit ihren charakteristischen Funktionen eindeutig charakterisieren (und umgekehrt); aus diesem Grund werden wir öfter von einer Charakterisierung zur anderen wechseln, ohne das explizit zu sagen.

Fuzzy Mengen generalisieren die Element-Relation zu verschiedenen Graden von Element-Sein. Die Intuition dahinter ist einfach: in der formalen Semantik sind Eigenschaften Mengen; wenn wir nun sagen etwas hat eine gewisse Eigenschaft (z.B. "groß"), dann kann das mehr oder weniger wahr sein, nicht nur wahr oder falsch (dass viele andere Probleme dabei ignoriert werden: "große Ameise", "kleiner Elefant", nehmen wir in Kauf).

Fuzzy Mengen generalisieren Wahrscheinlichkeiten Sei X unser Definitionsbereich; dann ist jede Wahrscheinlichkeitsverteilung

$$P : X \to [0, 1]$$

auch eine Fuzzy Menge. Umgekehrt gilt das natürlich nicht: wir haben ja im diskreten Fall

(7) $$\sum_{x \in X} P(x) = 1$$

Das bedeutet: in Wahrscheinlichkeitsverteilungen "konkurrieren" die Objekte um Wahrscheinlichkeitsmasse; in fuzzy Mengen gibt es nichts dergleichen; wir sind also viel freier, wenn es darum geht, Grade zuzuweisen. Dasselbe gilt ähnlich auch für stetige Wahrscheinlichkeitsdichtefunktionen

$$p : \mathbb{R} \rightarrow \mathbb{R}_0^+$$

die ja erfüllen müssen

$$(8) \qquad \int_a^b p(x)dx \in [0,1] : \; a < b$$

und

$$(9) \qquad \int p(x)dx = 1$$

7.2 Operationen und Relationen

Die erste große Frage ist:

> was ist mit den klassischen Relationen und Operationen der Mengenlehre, wie können wir die benutzen?

Wir werden im folgenden die wichtigsten Konzepte der Reihe nach durchgehen.

Die fuzzy Teilmengenrelation ist wie folgt definiert:

$$\text{wir sagen } A \subseteq B, \text{ falls f.a. } x \in X,\ A(x) \leq B(x).$$

Sei M eine fuzzy Menge; die **fuzzy Potenzmenge** von M ist die Menge aller fuzzy Teilmengen, geschrieben

$$\mathcal{F}(M) := \{A : A \subseteq M\}.$$

NB: die fuzzy Potenzmenge selbst ist eine knackige Menge, und zwar von fuzzy Mengen! Die **Kardinalität** einer fuzzy Menge errechnet man mit der einfachen Formel

$$|M| = \textstyle\sum_{x \in X} M(x).$$

Ein wichtiges Konzept, dass fuzzy mit knackigen Mengen verbindet, ist das Konzept des α-Schnittes, wobei $\alpha \in [0,1]$. Wir definieren den α-Schnitt von M als

(10) ${}^{\alpha}M := \{x : M(x) \geq \alpha\}$

Es gibt auch eine schärfere Version

(11) ${}^{\alpha+}M := \{x : M(x) > \alpha\}$

wir nennen das den scharfen Schnitt. Es ist klar dass

$${}^{\alpha_1}M \subseteq {}^{\alpha_2} M \text{ gdw. } \alpha_2 \leq \alpha_1$$

Wir haben also, für alle $\alpha \in [0,1]$, eine knackige Menge, die immer kleiner wird indem α größer wird. Zwei Begriffe sind besonders hervorzuheben, nämlich zunächst der **Kern** von M, definiert als

(12) $K(M) := {}^1M$

und der **Umfang** von M, definiert als

(13) $U(M) := {}^{0+}M$

Ein weiterer wichtiger Begriff ist die **Höhe** von M,

(14) $H(M) := max_{x \in X} M(x)$

und der Wertebereich von M,

$$\{\alpha \in [0,1] : \exists x \in X : M(x) = \alpha\}.$$

Diese Konzepte sind offensichtlich; was etwas weniger offensichtlich ist folgendes: jede fuzzy Menge wird eindeutig charakterisiert durch die Menge ihrer α-Schnitte, denn wir haben

(15) $M(x) = sup\{\alpha \cdot ({}^\alpha M)(x) : \alpha \in [0,1]\}$

Diese Formel sollte klar sein, wenn wir bedenken dass ${}^\alpha M(x) \in \{0,1\}$; das heißt wir wählen in der obigen Formel einfach das größte α so dass ${}^\alpha M(x) = 1$. Wir benutzen übrigens das Supremum, nicht das Maximum (was war nochmal der Unterschied??), und so bekommen wir

(16) $sup\{\alpha \cdot^\alpha M(x) : \alpha \in [0,1]\} = sup\{\alpha \cdot^{\alpha+} M(x) : \alpha \in [0,1]\}$

Wir können also eine fuzzy Menge M eindeutig darstellen als eine knackige Menge von Paaren

$$\{\langle M_\alpha, \alpha\rangle\}, \text{ wobei } \alpha \in [0,1]$$

wobei auch jedes A_α knackig ist und eben den (scharfen) α-Schnitt von M darstellt.

Schnittwürdigkeit Das bringt uns zu dem zentralen Konzept der Schnittwürdigkeit. Wir haben gesagt dass fuzzy Mengen knackige Mengen generalisieren; wir stehen also oft vor der Frage: gegeben

- eine Operation (wie Schnitt, Vereinigung),
- Eigenschaft (wie Konvexität im Hinblick auf R)
- oder Relation (wie Teilmenge)

auf knackigen Mengen, wie sollen wir diese Operation auf beliebige fuzzy Mengen erweitern? Oft gibt es eine Vielzahl von Möglichkeiten hierfür, aber es gibt nur wenige (oft nur eine) davon, die schnittwürdig ist.

Definition 3 *Sei F eine n-äre Operation, P eine Eigenschaft, R eine m-äre Relation auf knackigen Mengen, und F', P', R' jeweils ihre Erweiterung auf fuzzy Mengen.*

1. *F' ist schnittwürdig, falls f.a. fuzzy Mengen $M_1, ..., M_n$, $\alpha \in [0,1]$ gilt: ${}^{\alpha}F'(M_1, ..., M_n) = F({}^{\alpha}M_1, ..., {}^{\alpha}M_n)$.*
2. *P' ist schnittwürdig, falls f.a. fuzzy Mengen M gilt: M hat Eigenschaft P', genau dann wenn f.a. $\alpha \in [0,1]$ ${}^{\alpha}M$ die Eigenschaft P hat.*
3. *R' ist schnittwürdig, falls f.a. fuzzy Mengen $M_1, ..., M_m$ gilt: $R'(M_1, ..., M_m)$ gilt genau dann wenn f.a. $\alpha \in [0,1]$ gilt $R({}^{\alpha}M_1, ..., {}^{\alpha}M_m)$.*

Die Schnittwürdigkeit ist der Ritterschlag für eine fuzzy Operation/Eigenschaft/Relation.

Hausaufgabe 1 Zu bearbeiten bis zum 26.4.21, Abgabe per email vor dem Seminar.

Ist die Operation $\cup'$, definiert durch

(17) $(A \cup' B)(x) = max(A(x), B(x))$

eine schnittwürdige Erweiterung des klassischen $\cup$? Zeigen Sie!

Übung Zeigen Sie, dass die fuzzy Teilmengenrelation, die wir oben beschrieben haben, ist schnittwürdig ist.

Als kleine Hilfestellung: Wir schreiben $\subseteq'$ für die (oben definierte) fuzzy Teilmengenrelation. Wir müssen zeigen dass für alle fuzzy Mengen M_1, M_2, $M_1 \subseteq' M_2$ genau dann wenn für alle $\alpha \in [0,1]$ gilt: ${}^{\alpha}M_1 \subseteq^{\alpha} M_2$. Die Aussage zerfällt also in zwei Teile:

Für beliebige M_1, M_2 gilt:
a) $M_1 \subseteq' M_2$ impliziert für alle $\alpha \in [0,1]$ gilt: ${}^{\alpha}M_1 \subseteq^{\alpha} M_2$.
b) ${}^{\alpha}M_1 \subseteq^{\alpha} M_2$ für alle $\alpha \in [0,1]$ impliziert $M_1 \subseteq' M_2$.

Die beiden sind zu zeigen.

Erinnerung: $A \subseteq' B$ gdw. f.a. x gilt: $A(x) \leq B(x)$.

Beweis:

a) Nimm an, $M_1 \subseteq' M_2$, also f.a. x gilt: $M_1(x) \leq M_2(x)$. Sei $\alpha \in [0,1]$, beliebig.

Nimm an, $x \in {}^{\alpha}M_1$. Das bedeutet: $M_1(x) \geq \alpha$. Also: $M_2(x) \geq M_1(x) \geq \alpha$. Da $M_2(x) \geq \alpha$, gilt: $x \in {}^{\alpha}M_2$.

b) Kontraposition: nimm an, $M_1 \not\subseteq' M_2$. Das bedeutet: es gibt ein x so dass: $M_1(x) > M_2(x)$. Setze $\alpha_1 := M_1(x)$. Es gilt: $x \in {}^{\alpha_1}M_1$. Aber: $x \notin {}^{\alpha_1}M_2$, denn $M_2(x) < \alpha_1$. Also: ${}^{\alpha_1}M_1 \not\subseteq {}^{\alpha_1}M_2$. $\dashv$

Beispiel 2 Die fuzzy Schnittmenge $(M \cap_{min} N)(x) = min(M(x), N(x))$ ist schnittwürdig, wie sich leicht zeigen lässt.

Beispiel 3 Die fuzzy Schnittmenge $(M \cap_{prod} N)(x) = M(x) \cdot N(x)$ ist *nicht* schnittwürdig (setze $M(x) = N(x) = 0.5 = \alpha$).

Falls $M(x), N(x) \in \{0,1\}$, dann gilt: $(M \cap_{min} N)(x) = (M \cap_{prod} N)(x)$

Zu Beispiel 3 Schnittwürdig hiesse:

(18) ${}^{\alpha}(M \cap_{prod} N) = {}^{\alpha}M \cap {}^{\alpha}N$

Sei $M(x) = N(x) = 0.5 = \alpha$. Also: $x \in {}^{\alpha}M \cap {}^{\alpha}N$. Allerdings: $(M \cap_{prod} N)(x) = 0.5^2 = 0.25$, also: $x \notin {}^{\alpha}(M \cap_{prod} N)$. Also: ${}^{\alpha}(M \cap_{prod} N) \neq {}^{\alpha}M \cap {}^{\alpha}N$.

Zu Beispiel 2 Schnittwürdig heisst:

(19) ${}^{\alpha}(M \cap_{min} N) = {}^{\alpha}M \cap {}^{\alpha}N$

Heisst soviel wie: $x \in {}^{\alpha}(M \cap_{min} N) \Longleftrightarrow x \in {}^{\alpha}M \cap {}^{\alpha}N$
$\Longrightarrow$ Nimm an, $x \in {}^{\alpha}(M \cap_{min} N)$. Also: $min(M(x), N(x)) \geq \alpha$. Also: $M(x) \geq \alpha$, $N(x) \geq \alpha$. Also: $x \in {}^{\alpha}M$, $x \in {}^{\alpha}N$, also: $x \in {}^{\alpha}M \cap {}^{\alpha}N$.

$\Longleftarrow$ Nimm an, $x \in {}^{\alpha}M \cap {}^{\alpha}N$. Also: $x \in {}^{\alpha}M$, ... Also: $M(x) \geq \alpha$,.... Also: $min(M(x), N(x)) \geq \alpha$. Also: $x \in {}^{\alpha}(M \cap_{min} N)$.

Übung

Zu bearbeiten 3.5.2021, Abgabe bitte vor dem Seminar per email.

Wir sagen eine (knackige) Menge M ist konvex für eine Relation R, falls gilt: wenn $(x, y) \in R$, $(y, z) \in R$, $x \in M$, $z \in M$, dann ist auch $y \in M$.

Wir erweitern nun diesen Begriff auf fuzzy Mengen:

Wir sagen eine fuzzy Menge M ist konvex′ für eine Relation R, falls gilt: wenn $(x, y) \in R$, $(y, z) \in R$, dann ist $M(y) \geq min(M(x), M(z))$.

Zeigen Sie, ob/dass dies eine schnittwürdige Erweiterung ist des Begriffes der Konvexität!

Beispiel Wir sagen eine fuzzy Menge M ist konvex″ für eine Relation R, falls gilt: wenn $(x, y) \in R$, $(y, z) \in R$, dann ist $M(y) = 1$.

Der folgende Satz ist fundamental, und ergibt sich aus Gleichung (10) und den dazgehörigen Erwägungen:

Theorem 4 *Jede Operation, Eigenschaft, Relation der klassischen Mengenlehre hat höchstens eine fuzzy Erweiterung, die schnittwürdig ist.*

Das folgt, da die Eigenschaften in Definition (nach (1)) die fuzzy Erweiterung eindeutig bestimmen. Wir können diese Erweiterung auch **kanonisch** nennen. Zwei Anmerkungen sind wichtig:

1. Im Lemma steht *höchstens*; der Grund dafür ist: es gibt nicht immer eine sinnvolle fuzzy Erweiterung von fuzzy Relationen. Z.B. sehe ich nicht, wie man die Relation $\in$ sinnvoll auf fuzzy Mengen erweitern soll.

2. Schnittwürdigkeit ist wichtig, aber nicht entscheidend: schnittwürdige Erweiterungen sind, wie man sieht, eindeutig durch klassische Mengenlehre bestimmt, und damit zwar kanonisch, aber auch manchmal etwas "langweilig": wenn wir nur mit schnittwürdigen Erweiterungen arbeiten würden, wir würden das beste verpassen; denn es sind die nicht schnittwürdigen Eigenschaften, in denen fuzzy Mengen ihre genuinen Eigenschaften entfalten.

Gleichzeitig liefert uns Theorem 4 und Gleichung (10) eine Methode, um Operationen, Relationen etc. über knackigen Mengen effektiv auf fuzzy Mengen zu erweitern. Wir sagen daher auch: wir fuzzifizieren die Operation.

Fuzzyfizierung von Funktionen Eine wichtige Technik, die damit zusammenhängt, ist die **Fuzzyfizierung von Funktionen**. Seien M, N knackige Mengen, und $f : M \to N$ eine Funktion. Die Funktion wird fuzzifiziert zu einer Funktion

$$F : \mathcal{F}(M) \to \mathcal{F}(N),$$

also einer Funktion der fuzzy Potenzmengen. Die Funktion F wird bestimmt durch das **Erweiterungsprinzip**:

(20) $F(A) = B \iff$ f.a. $y \in N,\ B(y) = \underset{x:y=f(x)}{max} A(x)$

D.h. $B(y)$ wird bestimmt durch den maximalen Wert $A(x)$, vorausgesetzt dass $f(x) = y$. Dadurch ist die Funktion F *eindeutig* bestimmt. Die inverse Funktion F^{-1} ist bestimmt durch

(21) $(F^{-1}(N))(x) = N(y)$, wobei $y = f(x)$

Damit haben wir $F^{-1}(F(M)) \supseteq M$, und falls f eine Bijektion ist, haben wir $F^{-1}(F(M)) = M$.

7.3 Modifikatoren

Wir kommen nun zu den sog. **Modifikatoren**. Wenn eine fuzzy Menge eine Eigenschaft wie "groß" beschreibt, dann beschreibt ein Modifikator ein Adverb wie "sehr" – und damit meine ich, dass die jetzige Analogie problematisch ist wie die vorige, sich aber (unter Nicht-Linguisten) eingebürgert hat. Ein Modifikator ist – nach üblicher Verwendung – eine monotone, stetige Funkton

$$m : [0, 1] \to [0, 1].$$

Monoton bedeutet dass falls $x \leq y$, dann $m(x) \leq m(y)$, stetig bedeutet, dass minimale Veränderungen des Arguments minimale Veränderungen des Wertes implizieren. Sei m ein Modifikator; er modifiziert eine fuzzy Menge M zu einer fuzzy Menge $m(M)$ durch folgende Gleichung:

(22) $m(M)(x) = m(M(x))$

Der Modifikator interessiert sich also nicht für x selber, sondern nur für den Wert $M(x)$. Im Hinblick auf unsere späteren logischen Konzepte könnten wir sagen: Modifikatoren sind "Wahrheitsfunktional"; sie interessieren sich nur für die Wahrheitwerte, nicht für die Proposition selber. Das wird für die meisten fuzzy Operatoren gelten, denn es ist meist der einzige weg, sie allgemein zu definieren. Eine sehr offensichtliche Familie von Modifikatoren lässt sich wie folgt beschreiben:

(23) $m_\lambda(x) = x^\lambda$,

wobei $\lambda \in \mathbb{R}^+$ ein beliebiger Parameter aus den (streng) positiven reellen Zahlen ist. Es ist leicht zu prüfen dass für alle $\lambda \in \mathbb{R}^+$ m_λ ein Modifikator im obigen Sinne ist. Der Fall

(24) $\lambda = 0$

ist natürlich pathologisch denn wir haben $x^0 = 1$ f.a. $x \in \mathbb{R}$; deswegen kann er durchaus Probleme bereiten (wenn wir mit universellen Klassen arbeiten).

Falls $\lambda > 1$, dann $x^\lambda \leq x$
Falls $\lambda < 1$, dann $x^\lambda \geq x$
Falls $x = 1$, dann ist $x^\lambda = x$

7.4 Komplemente

Knackige Komplemente sind definiert für Mengen i.H.a. eine Referenzmenge: sei X die Referenzmenge, M eine (knackige) Menge; dann ist das Komplement von A

(25) $\overline{A} := \{x : x \in X, x \notin A\}$

Fuzzy Komplemente sind eine Verallgemeinerung davon, und sind *nicht eindeutig*, d.h. es gibt viele Arten, ein fuzzy Komplement zu bilden. Ein fuzzy Komplement ist ebenfalls eine Funktion

$$c : [0,1] \to [0,1],$$

die zusätzlich folgende Bedingungen erfüllt:

1. $c(1) = 0$,
2. $c(0) = 1$,
3. für $x, y \in [0,1]$, falls $x \leq y$, dann $c(y) \leq c(x)$ (antiton),
4. (fakultativ) c ist eine stetige Funktion,
5. (fakultativ) f.a. $x \in [0,1]$ gilt: $c(c(x)) = x$

Das Komplement von M $c(M)$ wird wiederum definiert durch

$$c(M)(a) = c(M(a)).$$

- Bedingungen 1 und 2 Stellen sicher dass c das klassische, knackige Komplement generalisiert;
- 3 verlangt dass es es die fuzzy Teilmengenrelation umkehrt: wenn a stärker zu M gehört als b, dann gilt für $c(M)$ das Gegenteil.
- Bedingung 4 ist klar;
- Bedingung 5 ist die sogenannte **Involutivität** (doppelte Negation hebt sich auf), eine wichtige logische Eigenschaft, die aber normalerweise nicht notwendig ist.

Wie gesagt gibt es für fuzzy Mengen viele Komplemente, da es viele Funktionen gibt, die Bedingung 1-5 erfüllen. Eine Familie von Komplementen, die besonders leicht zugänglich ist, sind die sog. ***Yager*-Komplemente**:

(26) $c_\lambda(x) = (1 - x^\lambda))^{1/\lambda}$,

wobei $\lambda \in \mathbb{R}^+$ (diesmal streng positiv). Hier bekommen wir für jedes λ ein Komplement. Ein besonderer Fall ist der Fall $\lambda = 1$; dann reduziert sich die Gleichung zu

(27) $c(x) = 1 - x$

Das liefert uns das sog. kanonische fuzzy Komplement, und wir schreiben $c_1(M) \equiv \overline{M}$, wobei $\overline{M}(x) = 1 - M(x)$.

Lemma 5 *c_1 ist nicht schnittwürdig.*

Beweis Nehmen wir $\alpha = 0.3$, $M(a) = 0.6$. Dann ist $\overline{M}(a) = 0.4$, also $a \in^\alpha \overline{M}$. Allerdings ist $a \in^\alpha M$, also $a \notin \overline{^\alpha M}$, also $^\alpha\overline{M} \neq \overline{^\alpha M}$. ⊣

Allgemeiner funktioniert diese Konstruktion immer wenn $M(a) > 0.5$ und $M(a) + \alpha \leq 1$. Also ist nicht einmal das kanonische fuzzy Komplement schnittwürdig. Man kann das wie folgt verallgemeinern:

Lemma 6 *Es gibt keine schnittwürdige Erweiterung des mengentheoretischen Komplementes.*

Beweis Mit Widerspruchsbeweis: nimm an es gibt ein schnittwürdiges fuzzy Komplement $\hat{c}$. Wir nehmen die Fuzzy Menge $A : [0,1] \to [0,1]$, so dass $A(x) = x$ f.a. $x \in [0,1]$. Nimm nun $\hat{c}(A)$. Dann gilt, nach Annahme,

$$^{0.2}\hat{c}(A) = \overline{^{0.2}A}$$

Da $0.3 \notin \overline{^{0.2}A}$, gilt also $0.3 \notin^{0.2} \hat{c}(A)$, also:

$$\hat{c}(A)(0.3) < 0.2$$

Weiterhin:

$$^{0.5}\hat{c}(A) = \overline{^{0.5}A}$$

Da $0.3 \in \overline{^{0.5}A}$, gilt auch $0.3 \in^{0.5} \hat{c}(A)$, also

$$\hat{c}(A)(0.3) \geq 0.5$$

Daraus folgt: $0.2 > \hat{c}(A)(0.3) \geq 0.5$ – Widerspruch! ⊣

Man sieht also: schnittwürdig ist hinreichend, um eine fuzzy Operation zu rechtfertigen, aber nicht notwendig: das wäre ein zu enges Kriterium!

Übung

Zu bearbeiten 3.5.2021, Abgabe bitte vor dem Seminar per email.

Wir haben in der klassischen Mengenlehre eine Relation, die eine Menge mit ihren Partitionen verbindet; $P \subseteq \wp(M)$ ist eine Partition von M, genau dann wenn gilt:

1. $\bigcup P = M$
2. falls $N, N' \in P$, dann ist $N \cap N' = \emptyset$.

Wir sagen also $(P, M) \in Part$, genau dann wenn P eine Partition von M ist.

Wir definieren nun die Fuzzy Partition P einer Fuzzy Menge M wie folgt: $P \subseteq \mathcal{F}(M)$ ist eine Fuzzy Partition genau dann wenn gilt:

1. $M(x) = max_{N \in P} N(x)$
2. falls $N, N' \in P$, $N \neq N'$, dann ist $min(N(x), N'(x)) = 0$.

Wir definieren entsprechend $Part'$: $(P, M) \in Part'$ genau dann wenn P eine Fuzzy Partition von M ist.

Ist diese Erweiterung schnittwürdig?

8 Erweiterung von Booleschen Operationen – Grundlage von Fuzzy Logik

Wir haben folgendes gesehen (nicht hier aber anderswo):

- Klassische Logik korrespondiert mit klassischer Mengenlehre, nämlich über Boolesche Algebra:

- jede endliche Boolesche Algebra ist isomorph zu einer Potenzmengenalgebra; und so entsprechen sich klassische Mengenlehre und klassische Logik.

Wir fangen nun hier mit Fuzzy Logik im engeren Sinne, und dabei mit ihrer Semantik an. Unsere Semantik wird (halbwegs) algebraisch, also:

- Wir generalisieren die Operationen der Booleschen Algebren

Wichtig: Wir reden also an dieser Stelle erstmal nicht über eine logische Sprache, sondern über deren Bedeutung!

Unsere Erweiterung der klassischen Mengenlehre war sehr einfach:

- die Funktion max generalisiert $\vee$ (wahrheitsfunktional und $\cup$ (auf charkteristischen Funktionen)

- die Funktion min generalisiert $\wedge$ (wahrheitsfunktional und $\cap$ (auf charkteristischen Funktionen)

Aber hier müssen wir kurz vorsichtig sein: ist das wirklich die einzige Art, diese Funktionen auf $\{0,1\}^2$ nach $[0,1]^2$ zu generalisieren? Natürlich nicht, z.B. gibt es auch

- $s(p \wedge q) = s(p) \cdot s(q)$

- $s(p \vee q) = min(1, s(p) + s(q))$

Tatsächlich gibt es unendlich viele Funktionen, die diese Eigenschaften erfüllen, also unsere Booleschen Funktionen erweitern. Wir möchten diese Funktionen irgendwie sinnvoll einschränken .

Es ist daher sinnvoll, eine Reihe von weiteren Kriterien aufzustellen, die jede solche Generalisierung erfüllen muss. Das sind die folgenden:

1. Assoziativ
2. Kommutativ
3. Monoton: wenn man den Wert in den Argumenten erhöht, dann sollte sich auch der Wert der Funktion erhöhen (nicht streng).

Das sind alles Eigenschaften, die gleichermaßen $\cup$ and $\cap$ betreffen. Wir konzentrieren uns zunächst auf $\cap$. Hier ist der entscheidende Begriff derjenige der t-norm.

8.1 Fuzzy $\wedge$ - die t-Norm

Wir erinnern uns: ein entscheidende Unterschied von Fuzzy Logiken zu Wahrscheinlichkeitstheorie ist, dass Fuzzy Logiken **wahrheitsfunktional** sind; das bedeutet:

> Der Wahrheitsgrad von $p \wedge q$ sollte berechenbar sein aus den Wahrheitsgraden von p und q.

(Das ist gerade das, was z.B. in der Wahrscheinlichkeitstheorie nicht gegeben ist.)

Da wir Wahrheitsfunktionen suchen (also Funktionen auf Wahrheitsgraden in $[0,1]$!), definieren zunächst einmal zwei Funktionen:

$$s, v : [0,1] \times [0,1] \to [0,1],$$

Diese Funktionen liefern dann die Grundlage für Schnitt und Vereinigung:

$$M \cup_v N(a) = v(M(a), N(a)) \tag{28}$$

$$M \cap_s N(a) = s(M(a), N(a)) \tag{29}$$

Wir basieren also Schnitt und Vereinigung auf gewisse Funktionen. Die Funktion selbst ist noch nicht bestimmte; wir hatten aber oben eine Liste von Bedingungen die sie erfüllen muss. In der Fuzzy Logik hat sich eine noch strkere Bedingung als grundlegend erwiesen, nämlich:

> Die Funktion s muss eine sog. **Dreiecksnorm** (t-Norm) sein, die Funktion v muss eine **Dreieckskonorm** (t-Konorm) sein.

Die Bedingungen sind wie folgt:

Definition 7 $s : [0,1] \times [0,1] \to [0,1]$ *ist eine t-Norm, falls sie folgende Bedingungen erfüllt:*

1. $s(x,1) = x$; *(1 neutral)*

2. $s(x,y) = s(y,x)$; *(kommutitativ)*

3. *aus* $x \leq y$ *folgt* $s(z,x) \leq s(z,y)$ *(monoton);*

4. $s(x, s(y,z)) = s(s(x,y), z)$ *(assozativ).*

Was man vielleicht vermisst (um sicherzustellen dass s den klassischen Schnitt generalisiert) ist

$$s(x,0) = 0.$$

Das lässt sich aber wie folgt ableiten: wir haben $s(0,1) = 0$, und da $x \leq 1$, folgt

(30) $s(x,0) = s(0,x) \leq s(0,1) = 0$

Ebenso gilt:

(31) $s(x,x) \leq s(x,1) = x$

Die umgekehrte Ungleichgung können wir aber nicht aus den obigen Prinzipien ableiten!

Es gibt noch weitere wichtige Eigenschaften von t-Normen, die oft gewünscht, aber nicht obligatorisch sind:

- Eine t-Norm ist stetig, falls eine minimale Änderung im Argument nur eine minimale Änderung im Wert zeitigt (der Begriff ist bekannt aus der Analysis).

- Eine t-Norm $*$ ist archimedeisch, falls gilt: für alle $x, y \in (0,1)$ gibt es ein n so dass $\overbrace{x * ... * x}^{n \text{ mal}} \leq y$. Das bedeutet: je öfter wir etwas assertieren, desto unwahrer wird es.

- Eine t-Norm $*$ ist streng, wenn 0 der einzige Nilpotent ist, d.h. falls es ein n gibt mit $\overbrace{x * ... * x}^{n \text{ mal}} = 0$, dann ist $x = 0$.

- nilpotent bedeutet: nicht streng.

Wir kommen nun zu den fundamentalen t-Normen.

Bekannte t-Normen Grob kann man sagen: **Jede t-Norm entspricht einer Logik.**

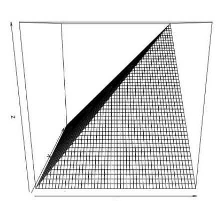
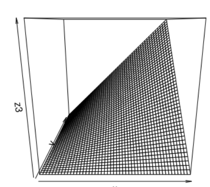
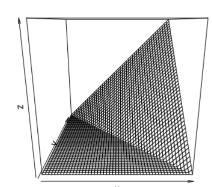
Es gibt folgende bekannte t-Normen:

- Kanonischer fuzzy Schnitt: $s(x,y) = min(x,y)$ (Gödel-Norm); ist stetig und jedes Objekt ist idempotent
- Produkt Norm: $s(x,y) = x \cdot y$; ist stetig, archimedeisch, streng
- beschränkte Differenz: $s(x,y) = max(0, x+y-1)$ (Łukasiewicz-Norm); ist stetig, nilpotent und archimedeisch
- drastischer Schnitt: $s_{min}(x,y) = \begin{cases} x, \text{ falls } y = 1 \\ y, \text{ falls } x = 1 \\ 0 \text{ andernfalls} \end{cases}$

  Ist nicht stetig aber archimedeisch.
- Nilpotente minimum t-Norm: $x * y = \begin{cases} min(x,y) \text{ falls } x + y > 1 \\ 0 \text{ andernfalls} \end{cases}$

  Ist nicht stetig!

Die 3 wichtigsten sind Gödel, Produkt und Łukasiewicz Norm. Daher folgende Grafiken:

Insbesondere gilt, f.a. t-Normen s, beliebige $x, y \in [0,1]$,

(32) $s_{min}(x,y) \leq s(x,y) \leq min(x,y)$

Wir haben also minimale und maximale t-Normen, und alle anderen liegen dazwischen. Eine Methode, t-Normen zu konstruieren, die den Zwischenraum ausfüllen, sind die sog. Yager-Schnitte. Das ist wiederum eine Familie von Schnitten i_λ, die definiert ist durch

(33) $i_\lambda(x,y) = 1 - min(1, [(1-x)^\lambda + (1-y)^\lambda]^{1-\lambda})$

wobei $\lambda \in \mathbb{R}^+$. Es ist nicht leicht zu sehen dass i_λ, für $\lambda \to \infty$, gegen min konvergiert. min ist nicht nur die maximale t-Norm, es ist auch die einzige fuzzy Schnittmengenbildung, die schnittwürdig ist; um Mißverständnisse zu vermeiden, nennen wir diese Norm ks

Lemma 8 *ks ist schnittwürdig: f.a. $\alpha \in [0,1]$ gilt:* ${}^\alpha(M \cap_{ks} M) ={}^\alpha M \cap^\alpha N$.

Proof. ${}^\alpha(M \cap_{ks} M) \subseteq^\alpha M \cap^\alpha N$: sei $a \in^\alpha (M \cap_{ks} M)$; dann ist $min(M(a), N(a)) \geq \alpha$, also $M(a) \geq \alpha, N(a) \geq \alpha$, also $a \in^\alpha M \cap^\alpha N$.

${}^\alpha M \cap^\alpha N \subseteq^\alpha (M \cap_{ks} M)$: sei $a \in^\alpha (M \cap_{ks} M)$, dann ist $a \in^\alpha M, a \in^\alpha N$, also $M(a) \geq \alpha, N(a) \geq \alpha$, also $M \cap_{ks} N(a) \geq \alpha$. ⊣

sk hat also eine besondere Bedeutung. ks hat noch eine weitere Eigenschaft hat: es ist die einzige t-Norm die **idempotent** ist, also

$$s(x,x) = x$$

erfüllt. Ich habe aber keinen Beweis dafür. t-Normen sind nicht nur wichtig für fuzzy Mengen, sondern für fuzzy Logiken: wir werden sie benutzen, um unsere Konjunktion zu interpretieren.

Übung

Nehmen wir folgende Operation:

(34) $$s_c(x,y) = \frac{1}{x+y}$$

Ist das eine t-Norm? Falls nein, welche der 4 Axiome erfüllt die Operation, welche nicht?

Übung

Beweisen Sie, dass die Gödel und Łukasiewicz t-Normen tatsächlich die 4 Kriterien einer t-Norm erfüllen!

Hier nur eine Musterlösung: Łukasiewicz t-Norm ist assoziativ.
Wir unterscheiden 4 Fälle:

Fall 1: $x + y > 1$, $y + z > 1$: dann ist

$$
\begin{aligned}
s(x, s(y, z)) &= max(0, x + max(0, y + z - 1) - 1) \\
&= x + (y + z - 1) - 1 && \text{(durch Annahme)} \\
&= (x + y - 1) + z - 1 && \text{(Arithmetik)} \\
&= max(0, max(0, x + y - 1) + z - 1) && \text{(durch Annahme)} \\
&= s(s(x, y), z)
\end{aligned}
$$

Fall 2a: $x + y > 1$, $y + z \leq 1$.

$$
\begin{aligned}
s(x, s(y, z)) &= max(0, x + max(0, y + z - 1) - 1) \\
&= max(0, x + 0 - 1) && \text{(Annahme)} \\
&= 0 && \text{(Arithmetik)} \\
&= max(0, (x - 1) + (y + z - 1)) && \text{(Annahme)} \\
&= max(0, x + y + z - 1 - 1) && \text{(Arithmetik)} \\
&= max(0, (x + y - 1) + z - 1) && \text{(Arithmetik)} \\
&= max(0, max(0, x + y - 1) + z - 1) && \text{(Annahme)} \\
&= s(s(x, y), z)
\end{aligned}
$$

Fall 2b: $x + y \leq 1$, $y + z > 1$.
Parallel zu 2a

Fall 3: $x + y \leq 1$, $y + z \leq 1$.

$$
\begin{aligned}
s(x, s(y, z)) &= max(0, x + max(0, y + z - 1) - 1) \\
&= max(0, x + 0 - 1) && \text{(Annahme)} \\
&= 0 \\
&= max(0, 0 + z - 1) && (z \in [0, 1]) \\
&= max(0, max(0, x + y - 1) + z - 1) && \text{(Annahme)} \\
&= s(s(x, y), z)
\end{aligned}
$$

QED

8.2 Das Mostert-Shields Theorem

Hier präsentieren wir das zentrale Theorem über stetige t-Normen (es ist ein bisschen schwer das richtig zu platzieren; um die Bedeutung zu verstehen muss man jedenfalls wissen, dass nur stetige t-Normen für die FL von Bedeutung sind, denn nur Stetigkeit garantiert, dass die Residua wohldefiniert sind (das weiter unten).

Der nächste wichtige Begriff ist derjenige der ordinalen Summe. Sei I zunächst eine beliebige Indexmenge, mit der wir eine Menge $(s_i, a_i, b_i) : i \in I$ indizieren, wobei

- jedes s_i eine t-Norm ist,
- jedes $(a_i, b_i) \subseteq [0, 1]$ (offene Intervalle!)
- alle $(a_i, b_i) : i \in I$ sind paarweise disjunkt
- $\bigcup\{(a_i, b_i) : i \in I\}$ liegt dicht in $[0, 1]$, d.h. zu jedem Punkt in $[0, 1]$ finden wir einen beliebig nahen Punkt in der Menge $\bigcup\{(a_i, b_i) : i \in I\}$

Nehmen wir also eine solche Menge. Dann definieren wir uns eine t-Norm $\star$ durch folgende Gleichung:

(35) $$x \star y = \begin{cases} a_i + (b_i - a_i) \cdot s_i(\frac{x - a_i}{b_- a_i}, \frac{y - a_i}{b_- a_i}), \text{ falls } x, y \in (a_i, b_i) \text{ für ein } i \in I \\ min(x, y) \text{ andernfalls} \end{cases}$$

Wir sagen $\star$ ist die ordinale Summe von $(s_i, a_i, b_i) : i \in I$; allgemeiner sagen wir $\star$ ist eine ordinale Summe der t-Normen $s_i : i \in I$.

Es ist nich ganz leicht zu sehen dass $\star$ auch eine t-Norm bildet, da die Definition etwas verwickelt ist; wir glauben das an dieser Stelle einmal. Folgendes Theorem ist grundlegend für die Semantik der FL:

Theorem 9 *Jede stetige t-Norm ist isomorph zu einer ordinalen Summe aus Łukasiewicz, Gödel und Produkt Norm.*

Wir erinnern uns an den Begriff der Isomorphie, auf deutsch etwa soviel wir "Strukturgleichheit". Wir sagen also: 2 t-Normen s, s' sind isomorph, wenn es eine bijektive Abbildung $i : [0, 1] \to [0, 1]$ gibt, so dass für alle $x, y \in [0, 1]$ gilt:

(36) $$i(s(x, y)) = s'(i(x), i(y))$$

Die Bijektivität von i impliziert übrigens dass eine Umkehrfunktion i^{-1} gibt so dass

(37) $i^{-1}(s'(x,y)) = s(i^{-1}(x), i^{-1}(y))$

i nennt man einen Isomorphismus. Isomorphie bedeutet also, dass s, s' in allen wesentlichen Teilen gleich sind.

Somit bedeutet Mostert-Shields Theorem dass wir für alle stetigen t-Normen nur unsere drei bekannten brauchen.

8.3 Vereinigung

Vereinigung ist eine duale Operation zu Schnitt; wir benutzen hier statt *t*-Normen sog. t-Konormen:

Definition 10 $v[0,1] \times [0,1] \to [0,1]$ *ist eine t-Konorm, falls sie folgende Bedingungen erfüllt:*

1. $v(x,0) = x$*; (0 neutral)*
2. $v(x,y) = v(y,x)$*; (kommutitativ)*
3. *aus* $x \leq y$ *folgt* $v(z,x) \leq v(z,y)$ *(monoton);*
4. $v(x, v(y,z)) = v(v(x,y), z)$ *(assozativ).*

Aus diesen Regeln können wir – nach demselben Muster wie oben – ableiten, dass $v(x,1) = 1$; umgekehrt haben wir $v(x,x) \geq x$, aber nicht die umgekehrte Ungleichung. Wichtige t-Konormen sind

1. kanonische Vereinigung: $kv(x,y) = max(x,y)$;
2. algebraische Summe: $v(x,y) = x + y - xy$ (vgl. Wahrscheinlichkeiten!)
3. begrenzte Summe: $v(x,y) = min(1, x+y)$;
4. drastische Vereinigung: $v_{max}(x,y) = \begin{cases} x, \text{ falls } y = 0 \\ y, \text{ falls } x = 0 \\ 1 \text{ andernfalls} \end{cases}$

Wiederum kann man recht leicht verifizieren dass f.a. t-Konormen v gilt:

(38) $kv(x,y) \leq v(x,y) \leq v_{max}(x,y)$;

wir haben also eine minimale und eine maximale Vereinigung. Widerum ist kv idempotent, $kv(x,x) = x$, und die einzige idempotente t-Konorm (nach dem was ich lese). Außerdem ist kv die (einzige) schnittürdige fuzzy Vereinigung; der Beweis läuft wie oben ab. Eine Familie von t-Konormen, mit denen man die Lücke zwischen minimalen und maximalen Konormen füllen kann, sind die Yager-Vereinigungen, die wie folgt definiert sind:

(39) $v_\lambda(x, y) = min(1, (x^\lambda y^\lambda)^{1/\lambda})$

wobei $\lambda \in \mathbb{R}^+$. Wiederum konvergiert diese Operation auf kv für $\lambda \to \infty$. Eine weitere Anmerkung sollte ich machen: es ist bekannt, dass klassischer Schnitt, Vereinigung und Komplement durch eine Reihe von Gesetzen miteinander verbunden sind (das wird oft unter Booleschen Algebren behandelt). Diese Gesetze gelten *nicht* für die fuzzy Erweiterungen, und zwar in keinem Fall. Man kann zwar unter gewissen Umständen manche Gesetze erfüllen, aber niemals alle.

8.4 Intermezzo: Fuzzyfizierung der logischen Konsequenz

In klassischer Logik schreiben wir

$$\alpha \models \beta$$

wenn gilt: falls $M \models \alpha$, dann $M \models \beta$. Auch dieser Begriff muss abgewandelt werden, wenn wir ihn auf Fuzzy Logik beziehen wollen.

Zunächst nehmen wir einmal **wahrheitsfunktionale Semanik**; das bedeutet, unsere Modelle haben die Form $\overline{\sigma}$, wobei

$$\sigma : Var \to \{0, 1\},$$

und $\overline{\sigma}$ ist die Erweiterung auf $Form(Var)$, nach den bekannten Wahrheitstabellen.

Diese Semantik lässt sich nun generalisieren: wir definieren eine fuzzy Valuation als eine Funktion

$$\tau : Var \to [0, 1]$$

atomare Propositionen bekommen also Wahrheitsgrade. τ wird erweitert zu $\overline{\tau}$, indem logische Konnektoren als binäre Funktionen interpretiert werden; so

- $\wedge$ als t-Norm,
- $\vee$ als t-Konorm etc.

– die Details folgen.

Die Frage Was bedeutet nun $\alpha \models \beta$ in *Fuzzy Logik*?

$M \models \alpha$ ist ja nun keine binäre Angelegenheit mehr, sondern kann jeden "Wahrheitsgrad" in $[0, 1]$ annehmen. Wir kommen aber den Begriff der Konsequenz wie folgt reformulieren:

$$\text{Wahrheitsfunktional: } \alpha \models \beta \text{ gdw. für alle } \sigma,\ \overline{\sigma}(\alpha) \leq \overline{\sigma}(\beta)$$

Und genau diesen Begriff kann man 1 zu 1 auf Fuzzy Logik applizieren:

$$\text{Fuzzy: } \alpha \models \beta \text{ gdw. für alle } \tau,\ \overline{\tau}(\alpha) \leq \overline{\tau}(\beta)$$

Das bedeutet: die Relation $\leq$ (auf $[0,1]$) ist entscheidend für die logische Konsequenz.

Wichtig ist aber, das wir hier vorgegriffen haben: τ ist eine beliebige Abbildung $Var \to [0,1]$, aber wie genau $\overline{\tau}$ definiert ist, haben wir nicht festgelegt. Wir wissen nur:

$$\overline{\tau}_s(\alpha \wedge \beta) = s(\overline{\tau}(\alpha), \overline{\tau}(\beta)),\ s \text{ eine } t\text{-Norm.}$$

Wie genau $\overline{\tau}$ für die anderen Konnektoren definiert ist, steht noch aus. Ebenso sollte klar sein: für jede t-Norm s wird sich eine andere Relation $\models$, sprich: eine andere Logik ergeben!

8.5 Residua und Implikationen

Wir kennen die klassische Implikation:

(40) $\alpha \to \beta \equiv \neg\alpha \lor \beta$

Diese Definition macht in allgemeinerer Logik keinen Sinn, sondern nur im Spezialfall der Wahrheitsfunktionen. Stattdessen nimmt man allgemeine Eigenschaften, die eine Implikation erfüllen muss. Der algebraische Begriff, der einer Implikation entspricht, ist der Begriff des **Residuums**.

Was macht eine Implikation aus? Zunächst folgende Eigenschaft: wir möchten dass Modus Ponens gilt:

(MP) $\alpha \land (\alpha \to \beta)$ impliziert β

Modus Ponens ist als Regel derart grundlegend, dass es die Implikation sozusagen definiert. Allerdings ist $\alpha \to \beta$ durch (MP) natürlich unterspezifiziert: z.B. würde (MP) auch gelten, falls allgemein gelten würde

(41) $\alpha \to \beta \equiv \bot$

was natürlich Unsinn ist.

Stattdessen nehmen wir an, dass (MP) sowohl eine notwendige als auch eine hinreichende Bedingung enthält:

- ► $\alpha \to \beta$ ist die *schwächste Annahme*, so dass $\alpha \land (\alpha \to \beta)$ impliziert β.

Das bedeutet:

(MP2) Falls $\alpha \land \gamma$ impliziert β, dann gilt: γ impliziert $\alpha \to \beta$.

Und schon haben wir die Definition des Residuums: $\alpha \to \beta$ wäre soz. das Residuum von α für β.

Ich schreibe "sozusagen", weil das Residuum eigentlich ein algebraischer Begriff ist, ich es aber mit Logik erklärt habe. In Algebren wird aus "impliziert" die Relation $\leq$. Nehmen wir also eine Algebra mit einem kommutativen Operator $\land$ und einer partiellen Ordnung $\leq$. Dann kann man das Residuum wie folgt definieren:

(Res) $a \land b \leq c$ genau dann wenn $b \leq a \to c$

Übung: Residuen in BA Nehmen Sie Boolesche Algebren, alles wie bekannt. Definieren Sie das Residuum nach (Res); wir haben also einen zusätzlichen Operator $\rightarrow$. Zeigen Sie dass $a \rightarrow b = {\sim}a \vee b$.

Lösung Wir haben zwei Richtungen in (Res):

1. $a \wedge (a \rightarrow b) \leq b$.

Wir haben

$$\begin{aligned} a \wedge (a \rightarrow b) &= a \wedge ({\sim}a \vee b) \\ &= (a \wedge {\sim}a) \vee (a \wedge b) \\ &= a \wedge b \\ &\leq b \end{aligned}$$

2. Falls $a \wedge c \leq b$, dann ist $c \leq a \rightarrow b$.

Wir zeigen als Zwischenschritt: $a \wedge b \leq c$ impliziert $b \leq {\sim}a \vee c$.

$$\begin{aligned} a \wedge b \leq c &\Leftrightarrow a \wedge b \wedge c = c \\ &\Rightarrow c \vee {\sim}a = (a \wedge b \wedge c) \vee {\sim}a \end{aligned}$$

Weiterhin:

$$\begin{aligned} c \vee {\sim}a &= (a \wedge b \wedge c) \vee {\sim}a \\ &= 1 \wedge (b \vee {\sim}a) \wedge (c \vee {\sim}a) \\ &= (b \vee {\sim}a) \wedge (c \vee {\sim}a) \\ &\Leftrightarrow b \vee {\sim}a \leq c \vee {\sim}a \\ &\Rightarrow b \leq c \vee {\sim}a \end{aligned}$$

Hiermit ist nun einfach zu zeigen: $a \wedge c \leq b$ impliziert $c \leq {\sim}a \vee b = a \rightarrow b$ QED.

Übung: Residuen in der Arithmetik Nehmen Sie die ganzen Zählen $(\mathbb{Z}, +)$. Was wäre das Residuum für die Operation $+$ in dieser Algebra ($\leq$ wie bekannt)?

$a + b \leq c$ gdw. $b \leq a \rightarrow c$. Was ist $a \rightarrow c$? $c - a$ natürlich! Denn $a + c - a = c$

Übung: Residuen in der Analysis Nehmen Sie die reellen Zählen $(\mathbb{R}^+, \cdot)$. Was wäre das Residuum für die Operation $\cdot$ in dieser Algebra ($\leq$ wie bekannt)?

$a \rightarrow c$: das größte x so dass $a * x \leq c$? c/a, denn $a \cdot c/a = c$

8.6 Residuen in Fuzzy Logik

Wir haben nun also folgende Begriffe: wir

1. die t-Norm, entsprechend der Konjunktion
2. die t-Konorm, entsprechend der Disjunktion
3. das Residuum, entsprechend der Implikation.

Genau das letztere werden wir uns jetzt etwas genauer anschauen für den Spezialfall der t-Normen. Wir lesen

- '$\wedge'$ als '$*'$,
- '$\models'$ als '$\leq'$;
- die semantische Übersetzung von '$\to'$ nennen wir '$\Rightarrow'$, ohne zu wissen was genau sie bedeutet – aber sie wird definiert über (Res).

Dann bekommen wir für unsere Fuzzy-Logik:

$$\tau(\alpha) \leq (\tau(\beta) \Rightarrow \tau(\gamma)) \text{ gdw. } \tau(\alpha) * \tau(\beta) \leq \tau(\gamma).$$

Da es sich hier um Funktionen auf Werten in $[0,1]$ handelt, ist klar, dass

$$\Rightarrow: [0,1]^2 \to [0,1]$$

auch eine solche Funktion sein muss. Was viell. weniger klar ist, ist dass $\Rightarrow$ durch das Gesetz der Residua bereits eindeutig bestimmt ist durch $*$:

$$x \Rightarrow y = max\{z : x * z \leq y\}.$$

Damit dieses Maximum existiert, muss $*$ stetig sein; damit lässt sich eindeutige Existenz einfach zeigen. Hier sehen wir also wir, dass wir stetige t-Normen brauchen. Als Gegenbeispiel nehmen wir einmal die drastische t-Norm. Hier haben wir, zur Erinnerung,

$$(42) \quad s_{min}(x,y) = \begin{cases} x, \text{ falls } y = 1 \\ y, \text{ falls } x = 1 \\ 0 \text{ andernfalls} \end{cases}$$

Definieren wir hier das Residuum $x \Rightarrow_{min} y$, bekommen wir:

(43) $x \Rightarrow_{min} y = max\{z : s_{min}(x,z) \leq y\}$

Sei nun $y := 0.5$, $x \in (0,1)$. Dann ist $s_{min}(x,z) \leq 0$ gdw. $z < 1$. Das bedeutet: $max\{z : s_{min}(x,z) \leq y\}$ existiert nicht!

Da die drastische t-Norm wegfällt, haben wir noch 3 Interpretationen von $*$, deren Residua wir einzeln betrachten werden.

1. $* = min$,
2. $* = \cdot$, und
3. $* = max(0, x+y-1)$.

Es ist leicht zu sehen dass alle diese Funktionen stetig sind. Wir schauen uns nun an, wie die Implikationen hierfür definiert sind. Wichtig ist zu beachten: wenn wir max schreiben, dann meinen wir den maximalen Wert in $[0,1]$, denn andere Werte stehen uns gar nicht zur Verfügung.

1. $x \Rightarrow_G y := max\{z : min(x,z) \leq y\} = \begin{cases} 1, \text{ falls } x \leq y; \\ y \text{ andernfalls.} \end{cases}$
2. $x \Rightarrow_{\cdot} y := max\{z : x \cdot z \leq y\} = \begin{cases} 1, \text{ falls } x \leq y; \\ \frac{y}{x} \text{ andernfalls.} \end{cases}$
3. $x \Rightarrow_L y := min\{1, z : x + z - 1 \leq y\} = \begin{cases} 1, \text{ falls } x \leq y; \\ (1-x) + y \text{ andernfalls.} \end{cases}$

Wir nennen die Implikation 1. die Gödel, 3. die Łukasiewicz Implikation. Łukasiewicz Implikation muss man kurz erklären: falls $x \leq y$, dann ist $x + 1 - 1 \leq y$; andernfalls haben wir $x + z - 1 = y$ gdw. $z = 1 + y - x$. Wichtig ist folgendes: wir haben für alle Implikationen

$$\text{falls } x \leq y, \text{ dann } x \Rightarrow y = 1.$$

Das ist kein Zufall: $\leq$ entspricht $\models$ entspricht logischer Konsequenz; also heißt das:

$$x \leq y \text{ gdw. “aus } x \text{ folgt } y\text{” wahr ist gdw. } x \Rightarrow y = 1.$$

Natürlich kann $\tau(\phi) \leq \tau(\chi)$ auch rein zufällig sein für ein gewisses v; uns interessiert aber immer was gilt *für alle* Valuationen v.

8.7 Negation

Uns fehlt nun nur noch ein Konnektor, nämlich die Negation. Sie kommt zum Schluss, denn in Fuzzy Logik wird die Negation definiert aus Implikation und Falsum (wie übrigens auch manchmal in klassischer Logik). Wir brauchen also zunächst eine logische Konstante 0, mit

$$\tau(0) = 0 \text{ (alternativ: } 0 : [0,1]^0 \to \{0\}).$$

Wir definieren damit nun:

(44) $\neg\phi := \phi \to 0$

Das bedeutet – zur Illustration:

(45) $\tau(\neg\phi) = \tau(\phi \to 0) = max\{x : \tau(\phi) * x \leq 0\}$

Für min, die maximale t-Norm, folgt also

$$\tau(\neg(\phi)) = 0 \text{ gdw. } \tau(\phi) > 0.$$

Wenn wir z.B. den drastischen Schnitt, die minimale t-Norm, nehmen, folgt (für $\tau(\phi) \in (0,1)$)

(46) $\tau(\neg\phi) = max\{x : \tau(\phi) * x \leq 0\} = max\{x : x < 1\}$,

was natürlich nicht existiert. Hier sehen wir wieder, dass wir stetige t-Normen brauchen.

Wir bekommen für unsere drei "prominenten" t-Normen nun also wiederum drei Negationen:

1. $* = min$,
2. $* = \cdot$, und
3. $* = max(0, x + y - 1)$.

Macht folgende Residua:

1. $x \Rightarrow y := max\{z : min(x,z) \leq y\} = \begin{cases} 1, \text{ falls } x \leq y; \\ y \text{ andernfalls.} \end{cases}$

2. $x \Rightarrow y := max\{z : x \cdot z \leq y\} = \begin{cases} 1, \text{ falls } x \leq y; \\ \frac{y}{x} \text{ andernfalls.} \end{cases}$

3. $x \Rightarrow y := max\{0, z : x + z - 1 \leq y\} = \begin{cases} 1, \text{ falls } x \leq y; \\ (1-x) + y \text{ andernfalls.} \end{cases}$

Macht folgende Negationen:

1. Gödel: $\neg x = \begin{cases} 1, & \text{falls } x = 0 \\ 0 & \text{andernfalls} \end{cases}$

2. Produkt: $\neg x = \begin{cases} 1, & \text{falls } x = 0 \\ 0 & \text{andernfalls} \end{cases}$

3. Łukasiewicz: $\neg x = 1 - x$

Wir sehen also dass die Produkt- und Gödel-Negation zusammenfallen! Diese beiden Negationen sind *drastisch* in dem Sinne, dass sie fast Wahrheitswerte auf 0 abbilden – alle bis auf 0. Die Łukasiewicz Negation dagegen ist *kanonisch* im Sinne der Fuzzy Mengenlehre.

Wir haben nun die (numerische) Semantik von Fuzzy Logiken besprochen. Wir werden nun anfangen die Logiken selber zu betrachten.

8.8 Das Gesamtbild

Wir möchten nun in der Lage sein, alle logischen Konnektoren zu definieren, und das sind wir im Prinzip auch. Es gibt nur noch ein paar Kleinigkeiten die Fehlen. Die großen Punkte sind:

- Konjunktion wird interpretiert als die t-Norm; wir nennen das auch die **starke Konjunktion**, üblicherweise * geschrieben.
- $\rightarrow$ wird interpretiert als das Residuum zur t-Norm.

Wir haben also soweit die Korrespondenz:

$$\text{eine } t\text{-Norm} \cong \text{eine Fuzzy Logik}$$

Um diese Korrespondenz aufrechtzuhalten, werden nun auch die anderen Konnektoren aus $*$ und $\rightarrow$ definiert:

- $\neg\alpha \equiv \alpha \rightarrow 0$
- (!) $\alpha \wedge \beta \equiv \alpha * (\alpha \rightarrow \beta)$
- $\alpha \vee \beta \equiv ((\alpha \rightarrow \beta) \rightarrow \beta) \wedge ((\beta \rightarrow \alpha) \rightarrow \alpha)$
- $1 \equiv \neg 0$
- $\alpha \leftrightarrow \beta \equiv (\alpha \rightarrow \beta) * (\beta \rightarrow \alpha)$

Diese Definitionen scheinen teilweise etwas willkrlich zu sein; um zu verstehen warum Konnektoren gerade so und nicht anders definiert sind, braucht man wohl ein tieferes Verständnis von Fuzzy Logik. Man kann vielleicht erstmal sagen: diese Definitionen sind diejenigen, die am besten funktionieren.

Wichtig sind hier zwei Punkte:

1. Es handelt sich hierbei immer noch um eine Generalisierung von klassischer Logik, d.h. alle diese Definitionen sind klassische Äquivalenzen
2. Durch diese Definitionen haben wir nach wie vor die Korrespondenz von t-Norm und Logik; das macht das Feld der möglichen Fuzzy Logiken sehr geordnet und übersichtlich.

Eine besondere Eigenschaft von Fuzzy Logiken ist die Unterscheidung von starker und schwacher Konjunktion. Die starke Konjunktion ist die t-Norm, die schwache Konjunktion wird daraus abgeleitet.

Eine Eigenschaft die jede Fuzzy Logik erfüllen muss ist die Gültigkeit von **Modus Ponens**, d.h.

$$\alpha, \alpha \to \beta \models \beta$$

Daraus und aus der Definition des $\wedge$ folgt dass:

$$\alpha, \alpha \to \beta \models \alpha \wedge \beta \qquad (= \alpha * (\alpha \to \beta))$$

Andere klassische Tautologien fallen aber weg, wie wir leicht prüfen können! Wenn wir aber prüfen wollen, ob eine Formel eine Tautologie oder eine Implikation gültig ist, müssen wir uns für eine Interpretation entscheiden. Wir sprechen daher von drei Logiken, nämlich

1. Gödel Logik (min-Norm)
2. Produkt Logik ($\cdot$-Norm)
3. Łukasiewicz Logik (Łukasiewicz Norm)
4. Basic Logic: Es gibt noch eine vierte interessante Logik, in der eine Formel eine Tautologie ist wenn Sie unter *jeder* stetigen t-Norm eine Tautologie ist.

8.9 Interpretation von Formeln, Tautologien etc.

Wir betrachten nun die Interpretation von Formeln. Wir nehmen an:

- Formeln werden konstruiert über $prop = \{p_1, p_2, p_3, ...\}$
- Binäre Konnektoren sind $*, \to, \wedge, \vee$
- Unärer Konnektor ist $\neg$

v ist eine sogenannte **Valuation**. Sie basiert auf einer Abbildung

$$v : prop \to [0, 1]$$

Sei s eine beliebige, aber festgelegte t-Norm, $\Rightarrow_s$ das zugehörige Residuum.

v wird dann kanonisch erweitert (für s) von atomaren auf beliebige Formeln durch v_s:

$$\begin{array}{rcl} v_s(\alpha * \beta) & = & s(v_s(\alpha), v_s(\beta)) \\ v_s(\alpha \to \beta) & = & v_s(\alpha) \Rightarrow_s v_s(\beta) \\ v_s(\neg\alpha) & = & v_s(\alpha) \Rightarrow_s 0 \\ v_s(\alpha \wedge \beta) & = & s(v_s(\alpha), v_s(\alpha \to \beta)) = min(v_s(\alpha), v_s(\beta) \\ (!)v_s(\alpha \vee \beta) & = & min(v_s((\alpha \to \beta)) \to \beta), v_s((\beta \to \alpha) \to \alpha)) \end{array}$$

Wir sagen dann: α ist eine Tautologie in der Fuzzy Logik basierend auf s, falls gilt: für alle $v : prop \to [0, 1]$ gilt:

$$v_s(\alpha) = 1$$

Weiterhin: sei Γ eine Menge von Formeln. Wir sagen

$$\Gamma \models_s \alpha$$

falls gilt: Wenn $v_s(\gamma) = 1$ f.a. $\gamma \in \Gamma$, dann $v_s(\alpha) = 1$.

8.10 Einige Abkürzungen/Ergebnisse

Die Definitionen der Konnektoren $\wedge, \vee$ sind einigermaßen kompliziert; wir werden zeigen, dass es auch einfacher geht:

(47) $v.(\alpha \wedge \beta) = min(v.(\alpha), v.(\beta))$

Beweis Siehe Stunde!

(48) $\tau_L(\alpha \wedge \beta) = min(\tau_L(\alpha), \tau_L(\beta))$

Beweis Es gilt:

$$\begin{aligned} \tau_L(\alpha \wedge \beta) &= \tau_L(\alpha) * \tau_L(\alpha \Rightarrow \beta) \\ &= max(0, \tau_L(\alpha) + \tau_L(\alpha \Rightarrow \beta) - 1) \end{aligned}$$

Nimm nun an, wir haben $\tau_L(\alpha) \leq \tau_L(\beta)$. Dann ist $\tau_L(\alpha \Rightarrow \beta) = 1$, also $\tau_L(\alpha) + \tau_L(\alpha \Rightarrow \beta) - 1 = \tau_L(\alpha)$.

Nimm dagegen an, $\tau_L(\alpha) > \tau_L(\beta)$. Dann gilt: $\tau_L(\alpha \Rightarrow \beta) = 1 - \tau_L(\alpha) + \tau_L(\beta)$, also

$$\begin{aligned} \tau_L(\alpha) + \tau_L(\alpha \Rightarrow \beta) - 1 &= \tau_L(\alpha) + (1 - \tau_L(\alpha) + \tau_L(\beta)) - 1 \\ &= \tau_L(\alpha) - \tau_L(\alpha) + \tau_L(\beta) + 1 - 1 \\ &= \tau_L(\beta) \end{aligned}$$

QED

Das bedeutet soviel wie: die Gödel-Norm ist in der Produkt-Norm definierbar! Nun zum $\vee$.

(49) $v.(\alpha \vee \beta) = max(v.(\alpha), v.(\beta))$

Beweis Siehe Stunde!

(50) $\tau_L(\alpha \vee \beta) = max(\tau_L(\alpha), \tau_L(\beta))$

Beweis Wir haben

$$\begin{aligned}\tau_L(\alpha \vee \beta) &= (\tau_L(\alpha) \Rightarrow_L \tau_L(\beta)) \Rightarrow_L \tau_L(\beta)) \wedge ((\tau_L(\beta) \Rightarrow_L \tau_L(\alpha)) \Rightarrow_L \tau_L(\alpha)) \\ &= min((\tau_L(\alpha) \Rightarrow_L \tau_L(\beta)) \Rightarrow_L \tau_L(\beta)),\ ((\tau_L(\beta) \Rightarrow \tau_L(\alpha)) \Rightarrow_L \tau_L(\alpha))\end{aligned}$$

Nun eine Fallunterscheidung:

Fall 1 Sei $x = \tau_L(\alpha) \leq \tau_L(\beta) \leq y$. Dann gilt: $x \Rightarrow_L y = 1$, also $(x \Rightarrow_L y) \Rightarrow_L y = 1 - 1 + y = y$.

Weiterhin: $y \Rightarrow x = 1 - y + x \geq x$ (da $y \geq 0$), also $(y \Rightarrow x) \Rightarrow x = 1 - (1 - y + x) + x = 1 - 1 + y - x + x = y$. Also:

$$\begin{aligned}(\tau_L(\alpha) \Rightarrow_L \tau_L(\beta)) \Rightarrow_L \tau_L(\beta)) \wedge ((\tau_L(\beta) \Rightarrow_L \tau_L(\alpha)) \Rightarrow_L \tau_L(\alpha)) &= min(\tau_L(\beta), \tau_L(\beta)) \\ &= \tau_L(\beta)\end{aligned}$$

Fall 2 Genau parallel.
QED

Das zusammen (mit einigen anderen Dingen) beweist folgendes Lemma:

Lemma 11 *In Gödel, Łukasiewicz und Produkt-Logik gilt:*

1. $\tau(\alpha \wedge \beta) = min(\tau(\alpha), \tau(\beta))$
2. $\tau(\alpha \vee \beta) = max(\tau(\alpha), \tau(\beta))$

Übung

Bitte bearbeiten bis zum 31.5.21

Berechnen Sie die Wahrheitswerte folgender Formeln in allen drei Logiken unter folgender Belegung, für Produkt, Łukasiewiecz und Gödel-Logik!

- $val(p_1) = 0.6$
- $val(p_2) = 0.3$
- $val(p_3) = 0.1$
- $val(p_4) = 0$

1. $(p_1 \wedge p_2) \to (p_3 \vee p_4)$
2. $(p_1 * \neg p_4) \to p_3$

Übung

Bitte bearbeiten bis zum 31.5.21

Wir haben oben gezeigt, dass für Łukasiewicz und Produkt-Logik das logische $\vee$ dem numerischen Maximum entspricht. Zeigen Sie das dasselbe für Gödel-Logik gilt!

Lösung Wir haben

$$\tau(\alpha \vee \beta) = \tau(((\alpha \to \beta) \to \beta) \wedge ((\beta \to \alpha) \to \alpha)) \tag{51}$$

$$= min(\tau((\alpha \to \beta) \to \beta), \tau((\beta \to \alpha) \to \alpha)) \tag{52}$$

Wir machen wie üblich die Fallunterscheidung.

Fall 1: $\tau(\alpha) < \tau(\beta)$. Dann ist $\tau(\alpha \to \beta) = 1$, also $\tau((\alpha \to \beta) \to \beta)) = 1 \Rightarrow_G \tau(\beta) = \tau(\beta)$.

Weiterhin: $\tau(\beta \to \alpha) = \tau(\alpha)$, also $\tau((\beta \to \alpha) \to \alpha) = \tau(\alpha) \Rightarrow_G \tau(\alpha) = 1$. Also wird

$$\begin{aligned} min(\tau((\alpha \to \beta) \to \beta), \tau((\beta \to \alpha) \to \alpha)) &= min(\tau(\beta), 1) \\ &= \tau(\beta) = max(\tau(\beta), \tau(\alpha)) \end{aligned}$$

(qua Fallannahme)x

Fall 2: $\tau(\beta) < \tau(\alpha)$. Genau parallel.

Fall 3: $\tau(\beta) = \tau(\alpha)$. In dem Fall ist $\tau(\alpha \to \beta) = 1$, also $\tau((\alpha \to \beta) \to \beta)) = 1 \Rightarrow_G \tau(\beta) = \tau(\beta)$; ebenso: $\tau(\beta \to \alpha) = 1$, also $\tau((\beta \to \alpha) \to \alpha)) = 1 \Rightarrow_G \tau(\alpha) = \tau(\alpha)$, also:

$$\begin{aligned} min(\tau((\alpha \to \beta) \to \beta), \tau((\beta \to \alpha) \to \alpha)) &= min(\tau(\alpha), \tau(\beta)) \\ &= \tau(\beta) = \tau(\alpha) \\ &= max(\tau(\beta), \tau(\alpha)) \end{aligned}$$

QED

Übung

Prüfen Sie ob folgende Formeln Tautologien sind in den drei Logiken (Gödel, Produkt, Łukasiewicz).

1. $p \vee \neg p$
2. $p \to p$
3. $p \to (q \to p)$
4. $(p \wedge q) \to q$
5. $p \to \neg\neg p$
6. $\neg\neg p \to p$
7. $(\neg p) \vee (\neg\neg p)$
8. $(p \to q) \to ((\neg q) \to (\neg p))$

Übung : Prüfen Sie ob folgende Implikationen gültig sind. Vorsicht, wir nehmen hier eine etwas andere Relation an: $\Gamma \models \alpha$ soll bedeuten: falls $\overline{\tau}(\gamma) = 1$ f.a. $\gamma \in \Gamma$, dann ist $\overline{\tau}(\alpha) = 1$. Also Vorsicht!

- $p, p \to q \models q$
- $p \vee q, \neg q \models p$
- $p \to (q \to r) \models (p \to q) \to (p \to r)$
- $(p \vee q) \to r \models (p \to r) \wedge (q \to r)$

Übung : Zeigen Sie, dass $\alpha \wedge \beta \equiv \beta \wedge \alpha$. Erinnern wir uns dass die Definition des schwachen $\wedge$ ja erstmal nicht symmetrisch ist!

Übung

Beweisen bzw. widerlegen Sie, dass 4., 7. und 8: Tautologien sind, und zwar für Gödel Logik und Łukasiewicz Logik!

8.11 Modus Ponens und die Bedeutung der Implikation

Wir haben gesagt, dass $p, p \to q \models q$ eine valide Inferenz ist in allen Fuzzy Logiken. Gleichzeitig gilt: alle Fuzzy Logiken sind **wahrheitsfunktional**, also ob wir etwas über p oder eine beliebige Formel α sagen, macht keinen Unterschied. Also:

Lemma 12 *In allen Fuzzy Logiken gilt:* $\alpha, \alpha \to \beta \models \beta$.

Das bedeutet: wir können MP als universelle Regel nutzen. Man kann das noch ein wenig stärker machen:

Lemma 13 *Sei* $\tau(\alpha) = x \in [0,1]$, $\tau(\alpha \to \beta) = 1$. *Dann gilt für alle Fuzzy Logiken:* $\tau(\beta) \geq x$.

Eine wahre Implikation $\alpha \to \beta$ ist also eine Aussage der Form: α ist mindestens so wahr wie β.

Das lässt sich nun wiederum verallgemeinern: nimm an (Łukasiewicz),

(53) $\tau_L(\alpha \to \beta) = 0.9 = 1 - \tau(\alpha) + \tau(\beta)$

Das lässt sich umformen zu:

(54) $\tau(\alpha) = \tau(\beta) + 0.1$

Ebenso für Produkt:

$$\tau_L(\alpha \to \beta) = 0.9 = \tau(\beta)/\tau(\alpha) \quad (55)$$
$$\Leftrightarrow \tau(\beta) = 0.9 \cdot \tau(\alpha) \quad (56)$$

Das bedeutet: die einzelnen Wahrheitswerte sind in diesen Fällen funktional bestimmt: kennt man $\tau(\alpha)$, dann auch $\tau(\beta)$ und umgekehrt. (In Gödel-Logik gilt das nicht).

9 Zurück zu Sorites

Wir kommen nun zurück auf das Sorites Paradox, und schauen ob wir mit unserer Theorie das Paradox lösen bzw. umgehen können. Wir haben folgende Prämissen:

a. Tausend Körner sind ein Haufen.

b. Wenn n Körner ein Haufen sind, dann sind auch $n-1$ Körner ein Haufen.

Das Problem war:

c. 1 Korn ist ein Haufen.

soll *nicht* aus diesen Prämissen folgen. Wenn wir diese Aussagen in unsere Logik übersetzen wollen, müssen wir etwas vorsichtig sein, denn wir haben ja nur eine Aussagenlogik, sprich, eine Formel wie

(57) $\forall n.H(n+1) \to H(n)$

ist *nicht* in Teil unserer Sprache (das wäre Prädikatenlogik). Stattdessen machen wir folgendes:

- Aussage a. übersetzen wir als Formel h_{1000}. Diese Formel ist wahr, also $\tau(h_{1000}) = 1$
- Aussage b. übersetzen wir als eine **Menge** von Formeln:

  $$\{h_{n+1} \to h_n : n < 1000\}$$

  Alle diese Formeln haben für uns denselben Wahrheitswert, aber nicht 1: denn wenn wir ein Korn wegnehmen, nimmt die Haufenhaftigkeit ab. Also sagen wir: die Formeln haben den Wahrheitswert 0.999.
- Wir nehmen also an, es gibt für jedes $n \in \mathbb{N}$ eine Formel h_n. Was der Wahrheitswert von h_1 ist, werden wir sehen. Andere Wahrheitswerte legen wir jedenfalls (erstmal) nicht fest.

Jetzt stehen wir also vor der Frage:

Wie ist der Wahrheitswert von h_1?

Das hängt natürlich davon ab, welche Logik wir betrachten.

Łukasiewicz Wir haben $\tau(h_{1000}) = 1$, $\tau(h_{1000} \to h_{999}) = 0.999$. Also:

$$\begin{aligned}
0.999 &= \tau(h_{1000} \to h_{999}) \\
&= \tau(h_{1000}) \Rightarrow_L \tau(h_{999}) \\
&= 1 \Rightarrow_L \tau(h_{999}) \\
&= 1 - 1 + \tau(h_{999}) \\
&= \tau(h_{999})
\end{aligned}$$

Der Wahrheitswert sinkt hier also. Diese Prozedur können wir jetzt abkürzen (denn in der Implikation sind 1000,999 ja beliebig!):

$$\begin{aligned}
0.999 &= 1 - \tau(h_{1000-k}) + \tau(h_{1000-k-1}) \\
\Leftrightarrow \tau(h_{1000-k-1}) &= 0.999 - 1 + \tau(h_{1000-k}) \\
&= \tau(h_{1000-k}) - 0.001
\end{aligned}$$

Wir haben also $\tau(h_{1000-k-1}) = \tau(h_{1000-k}) - 0.001$; jedesmal wenn wir ein Korn wegnehmen, sinkt der Wahrheitswert der Haufenhaftigkeit um 0.001 (ein tausendstel).

Wir können also hier sehr leicht ausrechnen dass gerade $\tau(h_1) = 0.001$. Übrigens nehmen wir natürlich an, dass $\tau(h_n) = 1$ für $n > 1000$. Dementsprechend gilt z.B. $\tau(h_{1001} \to h_{1000}) = 1$. Insgesamt würde ich sagen: das funktioniert also sehr gut!

Produkt Es gilt:

$$\begin{aligned}
0.999 &= \tau(h_{1000} \to h_{999}) \\
&= \tau(h_{1000}) \Rightarrow_P \tau(h_{999})) \\
&= \frac{\tau(h_{999})}{\tau(h_{1000})} \\
&= \frac{\tau(h_{999})}{1} \\
&= \tau(h_{999})
\end{aligned}$$

h_{999} hat also den gleichen Wahrheitswert wie vorher, 0.999. Die Prozedur können wir wieder leicht verallgemeinern zu:

$$\begin{aligned}
0.999 &= \frac{\tau(h_{1000-k-1})}{\tau(h_{1000-k})} \\
\Leftrightarrow \tau(h_{1000-k}) \cdot 0.999 &= (h_{1000-k-1})
\end{aligned}$$

Sprich: anstatt dass wir 0.001 substrahieren, multiplizieren wir mit 0.999. Also bekommen wir:

(58) $\tau(h_1) = 0.999^{999} = 0.3680635$

Also: der Wert sinkt langsamer. Er kann offensichtlich auch niemals 0 werden, aber er wird gegen 0 konvergieren, wenn wir die Zahlen richtig wählen. Funktioniert also, aber etwas schlechter (für mein Gefühl).

Gödel Wir haben hier

$$\begin{aligned}
0.999 &= \tau(h_{1000} \to h_{999}) \\
&= \tau(h_{1000}) \Rightarrow_G \tau(h_{999})) \\
&= \tau(h_{999})
\end{aligned}$$

Das können wir einfach festhalten, da falls $x \Rightarrow_G y \neq 1$, dann $x \Rightarrow_G y = y$. Soweit so gut, jetzt kommt aber ein Problem:

$$\begin{aligned}
0.999 &= \tau(h_{999} \to h_{998}) \\
&= \tau(h_{999}) \Rightarrow_G \tau(h_{998})) \\
&= 0.999 \Rightarrow_G \tau(h_{998}))
\end{aligned}$$

Jetzt haben wir ein Problem:

- Falls $\tau(h_{998}) < 0.999$, dann ist auch $\tau(h_{999} \to h_{998}) = \tau(h_{998})) < 0.999$ – Das steht im Widerspruch zu dem oben gegebenen Wahrheitswert $\tau(h_{999} \to h_{998}) = 0.999$!
- Falls $\tau(h_{998}) \geq 0.999$, dann ist auch $\tau(h_{999} \to h_{998}) = 1$ – das steht ebenfalls im Widerspruch zur Annahme!

Wir haben also bereits einen Widerspruch. Was bedeutet das? Unsere Annahmen waren bereits inkonsistent! Insbesondere dass es eine Menge gibt von Formeln $\{h_{n+1} \to h_n : n < 1000\}$, die alle einen Wahrheitswert haben von 0.999 (oder einen beliebigen anderen, für alle identischen Wahrheitswert) ist inkonsistent in Gödel Logik! Das funktioniert also nicht so gut...man kann das retten indem man die Wahrheitwerte stetig sinken lässt, aber das ist nicht im Einklang mit unserer ursprünglichen Intuition.

Zusammenfassung Wir haben also gesehen, dass Sorites Paradoxien sich relativ gut lösen lassen, v.a. mit Łukasiewicz. Ich würde das also mal als Erfolg verbuchen. Übrigens sehen wir hier auch: Gödel Logik ist etwas problematisch. Diese Logik wird auch sonst eher selten verwendet, denn: wir haben gesehen, die Operation *min* (Gödel Norm) lässt sich in den anderen beiden Logiken bereits definieren (als $\wedge$ aus &). Produkt und Łukasiewcz Logik sind also echt mächtiger als Gödel Logik.

Übung

Nehmen Sie folgende Valuation v:

$$\tau(p) = 0.5$$

$$\tau(q) = 0.8$$

$$\tau((p * q) \to r) = 0.9$$

Was folgt hieraus für den Wahrheitswert von r, also was können wir über $\tau(r)$ sagen? Bitte geben Sie die Antwort für Produkt, Gödel und Łukasiewicz-Logik.

10 Schätzung altertümlicher Meeresspiegel: eine linguistisch-geologische Anwendung

(entnommen aus: Ancient Sea Level Estimation in: Fuzzy Logic in Geology)

10.1 Das geologische Problem

Wir versuchen uns einmal an einer praktischen Anwendung; leider kommt die eher aus der Geologie als aus der Linguistik. Es geht darum: wir möchten schätzen, wie in lange vergangenen Zeiten sich der Meeresspiegel entwickelt hat. Hier ist ein gewisses geologisches Hintergrundwissen unabdingbar, das wir jetzt sehr kurz besprechen (ohne Gewähr, ich bin kein Geologe).

Es gibt verschiedene Arten von Gesteinen; für uns interessant sind die sog. sedimentären Gesteine, hauptsächlich Kalkstein (aber auch Dolomit). Diese Gesteine entstehen durch Ablagerungen von Kalk aus kleinstlebewesen, wie z.B. kleinen Muscheln und Schnecken *unter Wasser*. Auch grössere Muscheln kommen natürlich hierein, spielen aber eine untergeordnete Rolle. Das bedeutet: was wir als Kalkstein sehen, war früher einmal, nämlich zum Zeitpunkt seiner Entstehung, unter Wasser.

Nun gibt es folgendes Phänomen: nehmen wir an, Kalkstein entsteht, kommt aber dann, durch ein Sinken des Meerespiegels, ans Trockene. Hier gibt es eine Pause, denn an Land "wächst" kein sedimentäres Gestein. Später wiederum ist der Stein aber wieder unter Wasser, wächst also weiter. Die Pause des Wachstums macht sich jedoch bemerkbar als eine **geologische Schicht**; und je nach Mächtigkeit einer Schicht können wir gewisse geologische Perioden erschließen. (Wie man weiß hängt der Meeresspiegel ja eng mit dem Klima zusammen). Mächtige Schichten stehen also grob für lange warme Perioden; schwache Schichten für kurze Warmperioden.

Umgekehrt bedeutet ein großer Altersunterschied zwischen den beiden Schichten, dass es eine lange kühle Periode mit niedrigem Meeresspiegel gegeben hat; ein kurzer Altersunterschied deutet eine kurze Periode an (natürlich sind "lang" und "kurz" relativ hier).

Soweit, so unproblematisch. Nun das Problem: wir sind hier mit vielen Messungen konfrontiert, die an vielen verschiedenen Stellen durchgeführt werden. Es ist also sehr schwer, hier von "harten Zahlen" zu sprechen, denn Sedimentierung ist per se ein sehr unregelmäßger Prozess, der selbst an *einer einzigen* Stelle sehr unterschiedlich ausfallen kann; Annahmen über den Meeresspiegel sind aber per se *global*, beziehen sich also auf Messungen auf der gesamten Erde.

Hinzu kommt folgendes Problem: es ist ja nun gar nicht so einfach, ein Stück Stein zu finden, das aus dem Jahr 3.000.000 v.Chr. stammt; man ist hier also auf Zufallsfunde und unabhängig voneinander stattfindende Untersuchungen angewiesen. Es gibt also keine oder wenig Systematik in den Daten.

Nun also das Problem: wie kann man aus einer Anzahl disparater Beschreibungen allgemeine Schlüsse über den Meeresspiegel ziehen? Dieses ist unser Problem.

Die Methode Der Kern der Methode besteht in der numerischen Evaluation natürlichsprachlicher Ausdrücke. Z.B. haben wir Ausdrücke wie `sehr tief`,`mehr oder weniger dick`,`ungefähr mitteldünn`,`um die 200`,`seicht` usw.

Mit diesen Beschreibungen wollen wir Inferenzen ziehen!

10.2 Linguistische Beschreibungen

Es gibt eine Typologie linguistischer Beschreibungen, die im Prinzip eine sehr simple grammatische Klassifizierung ist. Wir fangen an mit

Atomare Wertbeschreibungen Das sind die Grundlagen für alle linguistischen Beschreibungen. Das sind typischerweise Adjektive wie

- "klein",
- "mittel",
- "groß"

ebenso dick, mittel, dünn etc, jung, alt. In diese Kategorien gehören auch *Fuzzy Quantitäten*, wie etwa

- "annähernd x" (x eine Zahl),
- "tausende",
- "eine Millionen".

Auch die letztere Quantität ist normalerweise als vage Mengenangabe verstanden.

Atomare Wertbeschreibungen kommen üblicherweise in **Antonymen**:

Adjektive – Antonym

Manchmal gibt es noch einen Mittelwert, wie für "groß" – "klein" den Term "mittelgroß"; das nennt man eine **Trichotomie**. Man spricht dementsprechend von **atomaren Paaren** und **atomaren Trichotomien**.

Einfache Wertbeschreibungen Einfache Wertbeschreibungen haben die Form

$$\langle\text{skalares Adverb}\rangle\ \langle\text{Atomare Wertbeschreibungen}\rangle$$

Typische skalare Adverben sind: sehr, äußerst, ungefährt, annaähernd, kaum, um die; dementsprechend haben wir einfache Wertbeschreibungen wie äusserst dick, ungefähr tausend, sehr dünn etc.

Man unterscheidet bei skalaren Adverben zwischen zwei Kategorien.

1. Diejenigen der ersten Kategorie haben einen **verengenden Effekt**, wie z.B. “sehr”, äußerst”. Sie verengen also die Bedeutung der Wertbeschreibung.

2. Diejenigen der zweiten Kategorie haben einen **erweiternden Effekt**, wie z.B. “mehr oder weniger”, “etwa”, denn sie erweitern die Bedeutung der atomaren Wertbeschreibung die sie modifizieren.

Ein sehr spezielles skalares Adverb ist das **leere Adverb**; die Bedeutung ist die Identitätsfunktion, und damit ist jede atomare Wertbeschreibung auch eine einfache Wertbeschreibung. Es gibt in der Fuzzy Logik auch eine Tradition der skalaren Adverbe, wo deren Anordnung festgelegt ist:

(59) $\mathsf{extremely}\ \langle$ atomarer Ausdruck $\rangle$

(60) $\mathsf{significantly}\langle$ atomarer Ausdruck $\rangle$

(61) $\mathsf{very}\langle$ atomarer Ausdruck $\rangle$

(62) $\epsilon\langle$ atomarer Ausdruck $\rangle$

(63) $\mathsf{more\ or\ less}\langle$ atomarer Ausdruck $\rangle$

(64) $\mathsf{roughly}\langle$ atomarer Ausdruck $\rangle$

(65) $\mathsf{quite\ roughly}\langle$ atomarer Ausdruck $\rangle$

(66) $\mathsf{very\ roughly}\langle$ atomarer Ausdruck $\rangle$

Es gibt aber auch hier einige Einschränkungen. Z.B. kann man verengende Adverben nicht mit mittleren Wertbeschreibungen nutzen. Also hat ein Ausdruck wie “sehr mitteldick” keine Bedeutung.

Komplexe Wertbeschreibungen Einfache Wertbeschreibungen können verknüpft werden mittels linguistischen Konnektoren. Wir haben hier nur zwei Konnektoren: “und” und “oder”. Die Semantik dieser Konnektoren werden wir noch besprechen.

Prädikationen Sei A eine Wertbeschreibung, X ein Nomen. Eine Prädikation hat die Form

$$X \ \mathsf{ist} \ A$$

Man nennt das auch eine einfache Prädikation. Z.B. etwas wie: “Felsmächtigkeit ist sehr groß”, oder “Meerespiegel Steigung ist sehr hoch”.

Fuzzy Implikationen Jetzt können wir endlich definieren, was eine Fuzzy Implikation ist. Fuzzy Implikationen sind unsere zentrale Methode zur Inferenz. Die syntaktische Form ist die folgende:

$$\mathsf{IF}\ \langle\text{Prädikation}\rangle\ \mathsf{THEN}\ \langle\text{Prädikation}\rangle$$

10.3 Fuzzy Implikationen

Wir möchten Inferenzen ziehen von Gesteinsschichten zu Meeresspiegel- Schwankungen. Dafür gibt es sog. *fuzzy* IF-THEN Regeln:

$\mathcal{C}$:=IF sequence is very thin THEN sea level change is rather small

Wie man sieht, ist die Regel zusammengesetzt aus einem festen Schema: Dinge haben Eigenschaften, und das eine impliziert das andere. Wir nennen

- die linke Seite die unabhängige Variable,
- die rechte Seite die **abhängige Variable**.

Einfachheit halber nennen wir die linke Seite X, die rechte Seite Y.

Eine Menge von Fuzzy Implikationen nennt man eine **linguistische Beschreibung**. Fuzzy Implikationen sind das Herzstück unseres Problems: mit ihnen kommen wir von geologischen Beschreibungen zu Schätzungen des Meeresspiegels.

Die Kernfragen sind aber:

1. wie interpretieren wir Fuzzy Implikationen?
2. woher kommen unsere Fuzzy Implikationen?

Frage 1 : Die Antwort ist recht einfach, auch wenn die technischen Gründe etwas komplexer sind: die Implikation selbst wird interpretiert als **Łukasiewcz-Implikation**. Die Wahrheitswerte der linguistischen Beschreibungen sind aber wiederum komplex und werden im nächsten Kapitel besprochen.

Frage 2 Werden wir weiter unten besprechen.

10.4 Die Semantik

Wir haben nun einige rein syntaktische Begriffe eingeführt, wir könnten sagen Typen oder Kategorien von Ausdrücken. Die Basis für diese Semantik ist die Unterscheidung von **Intension** und **Extension**. Die Begriffe ind konzeptuell eindeutig, werden aber in verschiedenen technischen Sinnen benutzt. Intuitiv bedeutet die Intension eines Wortes dasjenige, was es bedeuten kann oder könnte; die Extension die Menge der Objekte, auf die es tatsächlich referiert. Hier benutzen wir die Begriffe im Kontext der modalen Logik:

- Die Intension eines Begriffs B ist eine Funktion von Welten zu Extensionen.
- Die Extension eines Begriffs ist nur definiert an einer Welt, und ist, je nachdem, eine Menge, ein Intervall, ein Wahrheitswert...

Das klingt philosophisch, ist es auch, hat aber hier ganz konkrete Anwendung. Wir lösen auf diese Art das Problem der Kontextabhängigkeit von Wertbeschreibungen.

Beispiel "tief" kann z.B. ganz verschiedene Dinge bedeuten (aber immer eine antitone Funktion $\mathbb{R} \to \mathbb{R}$); im Kontext von Meeren und Flüssen z.B. Das sind genau die verschiedenen Extensionen in diesem Ansatz.

Wie funktioniert das? Wir definieren zunächst die **Menge der möglichen Welten** W als eine Menge von Tripeln:

$$W := \{\langle v_l, v_s, v_r\rangle :\in [0, \infty) \text{ und } v_l < v_s < v_r\}$$

Die drei Zahlen denotieren

- die linke Grenze,
- Mitte und
- rechte Grenze der Werte,

die ein Ausdruck annehmen kann. Sei nun $x \in \mathbb{R}$. Wir sagen dass x zu $w = \langle v_l, v_s, v_r\rangle$ gehört gdw. $x \in [v_l, v_r]$. Wir schreiben dann einfach $x \in w$, auch wenn w eigentlich keine Menge ist.

Ebenso schreiben wir das kartesische Produkt von Mengen:

$$w_1 \times ... \times w_n = [v_{1,l}, v_{1,r}] \times ... \times [v_{n,l}, v_{n,r}]$$

Die **Intension** eines Ausdrucks ist also definiert als eine Funktion

$$A : W \to \mathcal{F}(V),$$

wobei V das Universum ist. Das bedeutet also wir haben

$$A(w) : V \to [0, 1]$$

$A(w)$ ist also seinerseits eine Funktion. Da es etwas seltsam ist zu schreiben $A(w)(v)$, schreiben wir stattdessen $A_w(v)$.

Beispiel: klein Wir beschreiben nun die Bedeutung eines Ausdrucks wie "klein". Die Welt ist in diesem Fall das Referenzintervall, auf dem die Werte liegen können. Die Extension ist eine Funktion, die monoton fällt, wobei ihr Maximum am linken Rand liegt. Wir können uns das auch vorstellen als einen Beobachter, der am linken Rand des Intervalles steht und das Intervall betrachtet.

Mathematisch definieren wir klein(w) mit einer linearen Funktion, die wir L nennen (L wie links). Wir bekommen

(67) $$L_w(x) = \left(\frac{v_s - x}{v_s - v_l}\right)^\star$$

wobei $[-]^\star$ eine Funktion ist, die den Wertebereich auf $[0, 1]$ beschränkt.

Beispiel: groß Das ist der linke Horizont; jetzt kommt das Gegenstück, der rechte Horizont, der dem Begriff "groß" entspricht:

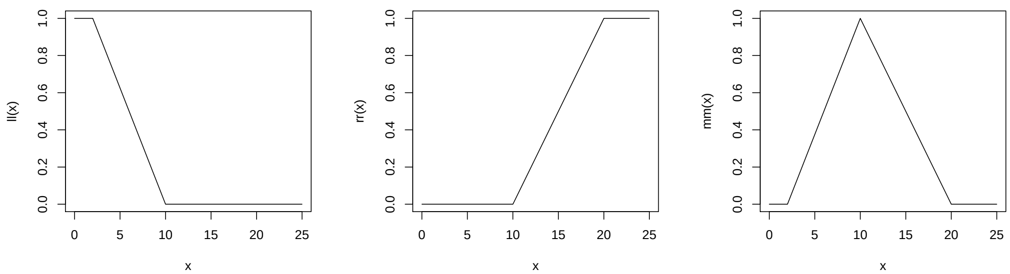
(68) $$R_w(x) = \left(\frac{x - v_s}{v_r - v_s}\right)^\star$$

Mit diesen beiden Funktionen können wir nun auch den mittleren Horizont definieren:

(69) $$M_w(x) = (1 - L_w(x)) \wedge (1 - R_w(x)) = \left(\frac{x - v_l}{v_s - v_l}\right)^\star \wedge \left(\frac{v_r - x}{v_r - v_s}\right)^\star$$

wobei $\wedge$ interpretiert wird als t-Norm. (Welche ist nicht spezifiziert; es macht auch keinen Unterschied: denn wenn an jedem Punkt ist eine der beiden inversen Funktionen 1, und 1 ist neutral für jede t-Norm).

Hier die Funktionsgraphen von L_w, R_w, M_w für $w = \langle 2, 10, 20 \rangle$:



Hiermit haben wir eine Semantik für atomare Ausdrücke. Als nächstes definieren wir eine Klasse von **Abstrakten Hecken** (komischer Ausdruck – *abstract hedges*). Diese liefern die Bedeutung der Adverbien, also verschärfen/mildern die Bedeutung der Adjektive. Es gibt eine einheitliche Definition dieser Hecken mit 3 Parametern:

$$(70) \quad v_{a,b,c}(y) = \begin{cases} 1 & : c \leq y \\ 1 - \frac{(c-y)^2}{(c-b)(c-a)} & : b \leq y < c \\ \frac{(y-a)^2}{(b-a)(c-a)} & : a \leq y < b \\ 0 & : y < a \end{cases}$$

wobei

- $a < b < c$
- $a, b \in (-\infty, 1)$
- $c \in (0.5, 1]$

Die AH *deformiert* den Horizont. Wir nennen die Menge aller dieser Funktion **AH**, und schreiben $v \in \mathbf{AH}$, was bedeutet $v = v_{a,b,c}$ für bestimmte $a, b, c \in R$ etc.

Wir sagen eine Hecke $v_{a,b,c}$ ist **schärfer** als $v_{a',b',c'}$, falls gilt $a' < a$, $b' < b$, $c' < c$; eine schärfere Hecke deformiert den Horizont stärker.

Die Autoren geben eine Reihe von Hecken an, von denen wir uns nur einen Auszug anschauen. Die Zahlen basieren auf eine "leeren Hecke", die zunächst willkürlich festgelegt ist; einige Hecken sind schärfer, andere weniger scharf.

Hecke	a	b	c
Extremely	0.5	0.75	0.95
...			
Very	0.35	0.58	0.83
leer	0.27	0.5	0.8
Rather	0.4	0.5	0.8
...			
Roughly	0.2	0.4	0.7
Very roughly	0.09	0.2	0.6

Es gibt nun 3 Arten von Intensionen; die entsprechen den oben erwähnten Bedeutungen "klein", "mittel", "groß", jeweils modifiziert durch eine AH.

Intension vom Typ *klein*

$$\mathbf{Sm} = \{Sm_v : Sm_v(w)(x) = v(L_w(x)), v \in \mathbf{AH}\}$$

Intension vom Typ *mittel*

$$\mathbf{Me} = \{Me_v : Me_v(w)(x) = v(M_w(x)), v \in \mathbf{AH}\}$$

Intension vom Typ *groß*

$$\mathbf{Bi} = \{Bi_v : Gr_v(w)(x) = v(R_w(x)), v \in \mathbf{AH}\}$$

Wir können nun für jeden linguistischen Ausdruck nun eine Semantik finden; sei $\mathcal{A}$ die Menge aller Ausdrücke; dann ist

$$Int(\mathcal{A}) = \mathbf{Sm} \cup \mathbf{Mi} \cup \mathbf{Gr}$$

Weiterhin haben wir, gegeben eine Welt $w = \langle v_l, v_s, v_r \rangle$, die Menge der Extensionen eines Ausdrucks an w:

$$Ext_w(\mathcal{A}) = Int(\mathcal{A})(w)$$

Das heißt, für $w = \langle v_l, v_s, v_r \rangle$ ist $Ext_w(\mathcal{A})$ eine Teilmenge der Fuzzy Mengen über $[v_l, v_r]$, oder anders ausgedrückt, eine Menge von Funktionen

$$\mathcal{F} : [v_l, v_r] \to [0, 1]$$

10.5 Das Problem des passenden Ausdrucks

Wir haben nun das inverse Problem: wir können sprachliche Ausdrücke numerisch interpretieren, aber wir wollen jetzt unsere Zahlen in sprachliche Ausdrücke transformieren. Ziel ist natürlich:

- Jede Zahl soll in einen Ausdruck transformiert werden, der für sie wahr ist;
- gleichzeitig soll der Ausdruck die Zahl so genau wie möglich umschreiben.

Hier gibt es verschiedene Algorithmen, und teilweise wird auch manuell gearbeitet. Die folgende Darstellung soll nur eine Idee geben.

Nehmen wir an, unsere Daten sind in Zeilen aufgelistet, wobei eine Zeile einer Messung eines Paars von Variablen entspricht, eine abhängig, die andere soll unabhängig sein, das wissen wir ja noch nicht.

Mächtigkeit der Sedimentschicht 1,	Zeitraum der Ablagerung 1
Mächtigkeit der Sedimentschicht 2,	Zeitraum der Ablagerung 2

Wir haben also

$$\text{Spalte } 1 \cong X, \quad \text{Spalte } 2 \cong Y$$

Als erstes müssen wir den linguistischen Kontext jeder Variable spezifizieren, das sind Intervalle in welcher der Wert jeder Variable fallen kann. Es gibt also einen höchsten und niedrigsten möglichen Wert. Dann suchen wir für den aktuellen Wert der Variable in einer Zeile eine **typische linguistische Beschreibung**. Wir ersetzen die Zahl durch die Beschreibung:

$$X_i \Rightarrow \text{sequence } \mathsf{is} \text{ td}(X_i), \qquad Y_i \Rightarrow \text{sea level change } \mathsf{is} \text{ td}(Y_i)$$

wobei $td(X_i)$ jeweils die **typische Beschreibung** von X_i ist (dazu spter mehr). Ebenso für $td(Y_i)$ wobei ich hier einige Probleme sehe, die nicht näher behandelt werden. Nun haben wir also Zeilen

sequence is td(X_1),	sea level change is td(Y_1)
sequence is td(X_2),	sea level change is td(Y_2)

Aus jeder solchen Zeile machen wir dann eine Fuzzy Implikation:

(71) IF sequence is very thin THEN sea level change is rather small

und so liefert uns eine Zeile eine Fuzzy Implikation.

Wenn wir diese Prozedur für den gesamten Datensatz durchführen, haben wir im Prinzip die numerischen Daten in eine linguistische Beschreibung gepackt. In unserem Fall bestehen die Daten aus aus zwei Teilen:

1. X – ein Zeitrahmen, in dem die Sedimente abgelagert wurden
2. Y – die Mächtigkeit einer Sedimentschicht

Wir ersetzen nun jeden Eintrag X_i durch eine **typische Beschreibung**

Das ist alles was wir brauchen (und haben!), um alte Meeresspiegel zu schätzen. Allerdings wird unsere Methode viele, teilweise widersprüchliche Implikationen generieren, denn genau das eigentliche Problem der unsauberen Daten wird ja noch nicht gelöst. Diese Regeln werden teilweise automatisch reduziert, teilweise von Hand (hier liegt der Vorteil der linguistischen Beschreibung: jeder Geologe kann Hand anlegen!)

Mittels Fuzzy Implikationen kann man aus einer Eingabe eine Ausgabe erzielen: gegeben eine Eingabe (ein Wert für X), findet man zunächst diejenige Regel, die am ehesten darauf appliziert, und dann "feuert" man mittels Modus Ponens: von der Tatsache dass A gilt, $A \Rightarrow B$ eine unserer Fuzzy Implikationen ist, schlussfolgern wir B. Also z.B.:

(72) IF sequence is very thin THEN sea level change is rather small

(73) sequence is very thin

(74) $\because$ sea level change is rather small

10.6 Logische Deduktion mit sprachlichen Beschreibungen

Nehmen wir an, wir haben eine If-Then Regel der Form

(75) $R = \mathsf{If}\ X \mathsf{then}\ Y$

Diese Regel muss nicht unbedingt einen Wahrheitswert von 1 haben; je nach relativer Häufigkeit von R kann R einen Wahrheitswert von $\alpha_R \in [0,1]$ haben. Nimm weiterhin an, wir haben eine linguistische Beschreibung X, wie z.B.

(76) $X =$ layer is more or less thick

Diese Beschreibung des Sachverhaltes muss ebenfalls nicht unbedingt den Wahrheitswert 1 haben, sondern kann einen beliebigen Wahrheitswert α_X haben. Wir haben gesagt dass die Implikation interpretiert wird als Łukasiewicz-Implikation. Was wir jetzt machen ist: wir applizieren Modus Ponens (man nennt das auch *feuern*). Wie funktioniert das? Wir haben hier zwei Wahrheitswerte:

$$v(X) = \alpha_X \tag{77}$$

$$v(\alpha_X \Rightarrow_L \alpha_Y) = \alpha_R \tag{78}$$

Weiterhin wissen wir dass

(79) $$x \Rightarrow_L y = \begin{cases} 1 & \text{falls } x \leq y \\ 1 - x + y & \text{andernfalls} \end{cases}$$

Das erlaubt uns eine Gleichung zu erstellen; wir müssen aber hier zwei Fälle unterscheiden:

Fall 1 $\alpha_R = 1$. In diesem Fall wissen wir, dass $\alpha_X \leq \alpha_Y$. Wir wissen nicht den vollständigen Wahrheitswert, aber wir können eine *untere Schranke* für den Wahrheitswert von Y angeben.

Fall 2 $\alpha_R < 1$. In diesem Fall haben wir

$$\begin{aligned} 1 - \alpha_X + \alpha_Y &= \alpha_R \\ \Leftrightarrow \qquad \alpha_Y &= \alpha_R + \alpha_X - 1 \end{aligned}$$

In diesem Fall ist der Wahrheitswert also genau bestimmt. Wenn also z.B.

$$\alpha_R = 0.8,\ \alpha_X = 0.7,$$

dann haben wir mittels Modus Ponens $\alpha_Y = 0.5$.

Wir können also mit Wahrheitswerten von Regeln und Wahrheitswerten von Prämissen die Wahrheitswerte unserer Konklusionen bestimmen! Also schließen wir, mittels Łukasiewicz Logik, direkt von Gesteinsschichten auf Wahrheitswerte den Meeresspiegel betreffend.

10.7 Defuzzyfizierung

Damit sind wir aber noch nicht am Ziel: wir haben Wahrheitswerte von Fuzzy Propositionen, nämlich von Prädikationen. Wir haben also am Ende, mit ein paar Zwischenschritten die wir ausgelassen haben, eine Verteilung Funktion

(80) $F : \mathbb{R} \to [0,1]$

d.h. eine Fuzzy Menge von Zahlen. Was wir am Ende wollen ist eine **einzelne Zahl** $x \in \mathbb{R}$, die uns sagt wie groß die Meeresspiegelschwankung tatsächlich ist. Das Problem von einer Fuzzy Menge zu einer Zahl zu kommen nennt man **Defuzzyfizierung**. Das ist ein altes und bekanntes Problem; die Autoren der Arbeit nehmen also nur bekannte Methoden auf; denn es gibt eine Vielzahl von Methoden zur Defuzzyfizierung.

Der Grundsatz zur Defuzzyfizerung lautet hier: wir möchten möglichst *konservativ* sein, also von allen Punkten, die in die Extension einer Fuzzy Prädikation fallen, denjenigen nehmen, der die *schwächsten Annahmen* macht. Aber welcher Punkt ist das? Das ist unterschiedlich, je nachdem was für eine Art von Ausdruck wir haben:

- Für Ausdrücke mit einem linken Horizont bedeutet dass, wir nehmen den größten Punkt mit dem maximalen Wahrheitswert
- Für Ausdrücke mit einem rechten Horizont bedeutet dass, wir nehmen den kleinsten Punkt mit dem maximalen Wahrheitswert
- Für Ausdrücke mit einem mittigen Horizont bedeutet dass, wir nehmen den mittelsten Punkt mit dem maximalen Wahrheitswert

Das sollte intuitiv einleuchten: wir möchten jeweils die schwächste Annahme machen, die mit unserem Wissen maximal konsistent ist. Wie wird das technisch umgesetzt? Formal git: eine Defuzzyfizierung ist eine Funktion

$$D : F \to \mathbb{R},$$

wobei $F : \mathbb{R} \to [0,1]$ eine Fuzzy Menge von Zahlen ist.

Wir definieren zunächst einmal drei verschiedene Funktionen zur Defuzzyfizierung, bevor wir sie zusammenfügen. Diese drei Funktionen sind definiert für Extensionen, beziehen sich aber auf die "Welt", die wir als Eingabe bekommen haben.

Sei Ev ein evaluierender sprachlicher Ausdruck (sprich die Bedeutung einer Prädikation), $w = \langle v_l, v_s, v_r \rangle$. Wir definieren zunächst **Least of Maxima**:

(81) $LOM(Ev(w)) = cv_l + (1 - c)v_s$

wobei $c \in (0.5, 1]$ der entsprechende dritte Parameter in der linguistischen Hecke in Ev ist (auch im folgenden). Als nächsten kommt **First of Maxima**:

(82) $FOM(Ev(w)) = cv_r + (1 - c)v_s$

Diese beiden Funktionen sind dual zueinander. Als nächstes kommt die Methode des **Center of Gravity**:

(83) $COG(Ev(w)) = \dfrac{\int_w v_s \cdot Ev_w(v_s)dv_s}{\int_w Ev_w(v_s)dv_s}$

(im Original steht v statt v_s) Es wird also *partiell integriert* über den "Mittelwert" v_s, während die anderen beiden Werte v_l, v_r Variablen der Funktion bleiben.

(Partielles Integral , häh? Das (bestimmte/unbestimmte) Integral liefert ja für die eindimensionale Funktion die Fläche unterhalb des Graphen. Wenn man nun z.B. zwei Dimensionen hat, man integriert nach einer der beiden (sagen wir Nr. 2), dann liefert uns die resultierende Funktion, das **partielle Integral**, an einem Punkt x die Gesamtfläche unterhalb des Graphen, wenn wir Dimension 1 auf x festlegen und den eindimensionalen Graphen in Dimension y betrachten.

Das bedeutet also: partiell nach einer Dimension integrieren heißt messen wieviel Masse jeweils in dieser Dimension liegt, für jeden Punkt der anderen Dimension. Die Fläche eines ganzen eindimensionalen Graphen liegt nun also an einem Punkt.)

Demensprechend liefert uns COG denjeningen Punkt, der im Mittelpunkt der Masseverteilung liegt; daher der Name.

Nun können wir die Defuzzifizierung definieren als Funktion DEE (defuzzyfication of evaluating expressions):

$$(84) \quad DEE(Ev(w)) = \begin{cases} LOM(Ev(v)) & \text{, falls } Ev \in \mathbf{Sm} \\ COG(Ev(v)) & \text{, falls } Ev \in \mathbf{Me} \\ FOM(Ev(v)) & \text{, falls } Ev \in \mathbf{Bi} \end{cases}$$

So kommen wir zu konkreten Zahlen für den Meeresspiegelanstieg zu bestimmten Zeiten. NB: es fehlen hier sehr viele nicht unwichtige Details, was wir hier sehen ist nur eine grobe Skizze der Methode, die die Autoren angewandt haben. Es sollte aber dennoch reichen, um eine Idee zu geben wie man Fuzzy Logik auf gewisse Probleme anwenden kann – und wie sprachliche Beschreibungen dabei helfen können!

11 Allgemeine/abstrakte Logik

11.1 Die Sprache klassischer Logik

Ich gebe zunächst einen Überblick über die Vorraussatzungen, die wir an klassischer Logik brauchen. Die ist weder vollständig noch selbsterklärend, kann aber in jeder Einführung in die Logik nachgelesen werden. Wir haben eine (induktiv definierte) Menge von Formeln, über eine abzählbare Menge von Variablen *Var* und die Konnektoren

$$\{\neg, \wedge, \vee, \rightarrow, \bot, \top\}.$$

Wir nennen die resultierende Menge *Form*(*Var*). Eine **Valuation** ist eine Funktion

$$v : \mathit{Var} \rightarrow \{0, 1\}.$$

Jeder unserer Konnektoren der Stelligkeit n wird interpretiert als Funktion von

$$\{0, 1\}^n \rightarrow \{0, 1\},$$

und auf diese Art erweitern wir v zu einer Funktion

$$\mathit{Form}(\mathit{Var}) \rightarrow \{0, 1\}.$$

Sei Γ eine Menge von Formeln, ϕ eine Formel. Wir schreiben

$$\Gamma \models \phi,$$

falls gilt: f.a. v, falls $v(\gamma) = 1$ f.a. $\gamma \in \Gamma$, dann ist $v(\phi) = 1$. Das ist die **semantische Konsequenz**.

11.2 Hilbert Kalküle

Wir kommen nun zur syntaktischen Konsequenz, Ableitbarkeit und Beweistheorie. Wir präsentieren nur den sog. **Hilbert-Kalkül**. Hilbert Kalküle sind einfach zu präsentieren, aber sperrig zu benutzen und deswegen unbeliebt. Ich benutze sie auch nicht gern, aber die fuzzy Logiken die ich kenne werden allesamt nur im Hilbert Stil präsentiert.

Ein Hilbert Kalkül besteht normalerweise aus einer Menge von Axiomen. Diese Menge ist normalerweise überschaubar und wird endlich präsentiert; die Axiome für klassische Logik sind

$\rightarrow$
(c1) $\alpha \rightarrow (\beta \rightarrow \alpha)$
(c2) $(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$
$\perp$
(c3) $\perp \rightarrow \alpha$
$\neg$
(c4) $\neg\alpha \rightarrow (\alpha \rightarrow \perp)$
(c5) $(\alpha \rightarrow \perp) \rightarrow \neg\alpha$
$\vee$
(c6) $\alpha \rightarrow (\alpha \vee \beta)$
(c7) $\beta \rightarrow (\alpha \vee \beta)$
$\wedge$
(c8) $(\alpha \wedge \beta) \rightarrow \alpha$
(c9) $(\alpha \wedge \beta) \rightarrow \beta$
(c10) $\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))$
GAD
(c11) $\neg\neg\alpha \rightarrow \alpha$

Das sind zwar nur wenige, aber man muss im Kopf behalten dass damit soz. unendlich viele Formeln repräsentiert werden: die griechischen Buchstaben sind sog. Metavariablen, für die wir beliebige Formeln substitutieren können. Die Axiome bleiben gültig, sofern wir die Substitution *einheitlich* machen: gleiche Metavariablen werden gleich substituiert. Zusätzlich zu den Axiomen gibt es **Inferenzregeln**, im propositionalen Hilbert-Kalkül nur eine, nämlich Modus Ponens:

$$\text{(MP)}\ \frac{\alpha \to \beta \quad \alpha}{\beta}$$

Das war es auch schon; was wir noch brauchen ist der Begriff des Beweises: ein Hilberbeweis von einer Formel α ist eine endliche Folge

$$\langle \psi_1, ..., \psi_i \rangle,$$

so dass

1. $\alpha = \psi_i$,
2. f.a. $j \leq i$ gilt: ψ_j ist entweder die Instanz eines Axioms, oder es gibt $j', j'' < j$, so dass ψ_j abgeleitet werden kann aus $\psi_{j'}, \psi_{j''}$ mit Modus Ponens.

Wir schreiben $\vdash_H \phi$, falls es einen Hilbert Beweis von ϕ gibt. Wir schreiben $\Gamma \vdash_H \phi$, falls es einen Hilbert Beweis von ϕ gibt mit zusätzlichen Annahmen $\gamma \in \Gamma$.

Beispiel 1 Wir liefern einen Beweis von $\vdash_H p \to p$, an dem man auch schön sieht, wie man Hilbert Beweise organisiert.

1	$(p \to ((p \to p) \to p)) \to ((p \to (p \to p)) \to (p \to p))$	C2
2	$p \to ((p \to p) \to p)$	C1
3	$((p \to (p \to p)) \to (p \to p))$	MP aus 1,2
4	$p \to (p \to p)$	C1
5	$p \to p$	MP aus 3,4

Man sieht hier: am besten man organisiert einen Hilbert Beweis als Tabelle mit drei Spalten: 1 zählt die Zeilen, 2 enthält die Formel, die im jeweiligen Schritt bewiesen wird, und 3 die Art, auf die es bewiesen wurde: entweder ein Axiom, dessen Instanz es ist, oder MP und die Formeln, auf die MP appliziert wurde.

Beispiel 2: Ein Beweis von Annahmen Wir liefern nun einen Beweis von Annahmen, nämlich die Transitivität logischen Schließens, formaler:

$$p \to q, q \to r \vdash_H p \to r$$

Von Annahmen beweisen bedeutet: wir können die Annahmen beliebig benutzen im Beweis.

1	$(p \to (q \to r)) \to ((p \to q) \to (p \to r))$	C2
2	$(q \to r) \to (p \to (q \to r))$	C1
3	$q \to r$	Annahme
4	$p \to (q \to r)$	MP aus 2,3
5	$(p \to q) \to (p \to r)$	MP aus 1,4
6	$p \to q$	Annahme
7	$p \to r$	MP aus 5,6

Um das ganze übersichtlicher zu gestalten, habe ich also die Annahmen, bevor ich sie benutzt habe, nochmal explizit eingeführt in den Beweis; wenn man viele Annahmen hat, wird es sonst sehr schwierig den Beweis nachzuvollziehen.

Beispiel 3: Prämissen kommutieren Hier wollen wir folgendes beweisen: $p \to (q \to r) \vdash_H q \to (p \to r)$. Das ist aber ziemlich langwierig und schwierig, deswegen schieben wir das nach hinten, bis wir zusätzliche Mittel zur Verfügung haben!

Beispiel 4: Ex falso quodlibet Eine interessante Formel ist folgende: $p \to (\neg p \to q)$. Diese Formel bedeutet: aus einen Widerspruch folgen beliebige Dinge. Auch das ist alles andere als einfach zu zeigen, deswegen verschieben wir das.

Das Deduktionstheorem DT Ein weiterer wichtiger Satz ist das sog. Deduktionstheorem:

Theorem 14 $\Gamma \cup \{\phi\} \vdash_H \psi$ *gdw.* $\Gamma \vdash_H \phi \to \psi$.

Das sagt uns dass die metalogische Relation $\vdash_H$ genau dem logischen Konnektor $\to$ entspricht. Eine Richtung ist trivial und entspricht Modus Ponens. Die andere Richtung ist durchaus kompliziert; man macht das mit einer Induktion über die Länge des Beweises.

11.3 Hilbert Kalküle handhabbar machen

Wie man leicht sieht, ist das Beweisen mit Hilbert Kalkülen eine schwierige und langwierige Arbeit. Hilfreich sind dabei aber folgende Grundsätze: Falls α eine beweisbare Formel ist, dann ist auch jede uniforme Substitution von α eine beweisbare Formel. Also kann α fortan genutzt werden wie die Axiome!

Ebenso: nehmen wir an, wir haben einen Beweis von Annahmen $\Gamma \vdash_H \alpha$. In diesem Fall gilt: nimm an, wir haben eine uniforme Substitution σ von $\Gamma \cup \{\alpha\}$, wobei uniform bedeutet: uniform über die ganze Menge. Dann können wir aus der Beweisbarkeit von $\sigma[\Gamma]$ schließen, dass auch $\sigma(\alpha)$ beweisbar ist. Also gilt:

Grundsatz von Hilbert Kalkülen Jeder Beweis einer Formel liefert uns ein neues Axiom; jeder Beweis einer Formel von Annahmen liefert uns eine neue Inferenzregel.

- Ebenso können wir nun DT nutzen um Dinge zu beweisen!

Damit können wir nun eine etwas kompliziertere Formel beweisen:

Beispiel 3, fortgesetzt Wir können nun den Beweis liefern für $p \to (q \to r) \vdash_H q \to (p \to r)$. Das läuft über DT. Wir beweisen zunächst: $p \to (q \to r), p, q \vdash_H r$. Das ist extrem einfach, mittels 2 Applikationen von MP. Nun nutzen wir DT:

$p \to (q \to r), p, q \vdash_H r$ impliziert
$p \to (q \to r), q \vdash_H p \to r$ impliziert
$p \to (q \to r) \vdash_H q \to (p \to r)$.

Das ist also ein klassischer Fall von einem **metalogischen Beweis**: wir haben nicht mehr den unmittelbaren Hilbert Kalkül, sondern nutzen seine abstrakten Eigenschaften.

Beispiel 4, fortgesetzt Wir beweisen nun $p \to (\neg p \to q)$. Das geht wie folgt; zunächst kommt der Beweis für $p, \neg p \vdash_H q$:

1	$\neg p \to (p \to \perp)$	C4
2	$\neg p$	Annahme
3	p	Annahme
4	$p \to \perp$	MP aus 1,2
5	$\perp$	MP aus 3,4
6	$\perp \to q$	C3
7	q	MP aus 5,6

Nun kommen wir mit dem Deduktionstheorem zum gewünschten Ergebnis:

$p, \neg p \vdash_H q$ impliziert
$p \vdash_H \neg p \to q$ impliziert
$\vdash_H p \to (\neg p \to q)$

Beispiel 5 Wir liefern einen Beweis von $p \to \neg\neg p$.

11.4 Vollständigkeit

Die folgende Eigenschaft nennt man (schwache) **Vollständigkeit**:

Lemma 15 $\vdash_H \phi$ *gdw.* $v(\phi) = 1$ *f.a. Valuationen* v.

Wir können den Begriff erweitern: wir sagen $\langle \psi_1, ..., \psi_i \rangle$ ist ein Beweis von $\Gamma \vdash_H \phi$, falls gilt: jedes ϕ_j ist ein Axiom, ableitbar aus Vorgängern, oder in Γ. Wir nehmen uns also zusätzliche "Axiome". Das sind aber keine abstrakten Axiome über Metavariablen, sondern konkrete Instanzen im Normalfall. Das nächste Theorem, die starke Vollständigkeit, ist folgendes:

Theorem 16 $\Gamma \models \phi$ *gdw.* $\Gamma \vdash_H \phi$.

Das nennt man die Vollständigkeit des Beweiskalküls.

12 Fuzzy Logik - Beweistheorie

Wie wir gesehen haben, haben wir vier Fuzzy Logiken, die wir rein semantisch konstruieren können: pro t-Norm eine Logik, sowie eine Logik für die Tautologien aller t-Normen. Das Problem ist nun: die semantische Verifikation von Tautologien, Implikationen etc. ist recht mühsam. Hier sieht man, wie wichtig eine vollständige Beweistheorie ist. In klassischer Logik ist das nicht unbedingt offensichtlich, da die einfache wahrheitstheoretische Semantik eine einfache Verifizierung erlaubt.

Die Semantik ist also gegeben, nun betrachten wir die Beweiskalküle, versuchen deren Vollständigkeit zu zeigen, und und untersuchen, in inwieweit klassische metalogische Resultate noch Gültigkeit haben.

Fuzzy Logik ist – wie klassische Logik – wahrheitsfunktional. Damit ist klar, wie sich Valuationen und Konnektoren verhalten: Valuationen bilden Var nach $[0, 1]$ ab, und n-äre Konnektoren werden interpretiert also Funktionen

$$[0, 1]^n \to [0, 1].$$

Die starke Konjunktion wird interpretiert als t-Norm, die Implikation als deren Residuum, und alle anderen Konnektoren werden durch diese Konnektoren definiert. Wir haben also wie besprochen 4 Logiken:

1. Gödels Logik, mit $* := min$;
2. Produkt Logik, mit $* := \cdot$;
3. Łukasiewicz's Logik, mit $* := max(0, a + b - 1)$
4. Hajeks Logik (*basic logic*, *BL*), mit beliebigen t-Normen (gültig ist was für jede stetige t-Norm gültig ist)

Jede dieser Logiken ist vollständig für die oben beschriebene Semantik und der entsprechenden Definition von $*$; Hajeks Logik ist vollständig für *jede* Interpretation. D.h. jede der drei erstgenannten Logiken ist eine Erweiterung von Hajeks Logik, deswegen nennt man die auch **Basic Logic** oder *BL*. Wir werden daher mit *BL* anfangen.

13 Hajeks Logik BL

Eine Konvention in der abstrakten Logik ist folgende: wir identifizieren eine Logik mit der Menge ihrer Theoreme (oder Tautologien, falls wir semantisch denken). Logiken sind also schlichtweg Mengen. In diesem Sinne können wir sagen:

Definition 17 *Logik $\mathcal{L}_1$ ist größer als Logik $\mathcal{L}_2$ gdw. das bedeutet: jedes Theorem von $\mathcal{L}_2$ ist ein Theorem von $\mathcal{L}_1$.*

Das setzt natürlich voraus, das wir entweder dieselben Konnektoren haben in $\mathcal{L}_1, \mathcal{L}_2$, oder das wir zusätzlich eine **Einbettung** angeben, mit der jeder Konnektor von $\mathcal{L}_2$ in eine $\mathcal{L}_1$-Formel übersetzt wird (diese Einbettung muss natürlich injektiv sein).

Beispiel 1: Klassische Logik Z.B. ist klassische Logik *maximal*, d.h. es gibt keine größere Logik mit denselben Konnektoren, die nicht trivial ist.

Beispiel 2: Hajeks Logik Von allen fuzzy Logiken (basierend auf stetiger t-Norm) ist Hajeks Logik *minimal*, denn es gilt:

Definition 18 *α ist eine Tautologie in Hajeks Logik BL, falls gilt: für jede Interpretation $v : prop \to [0,1]$, und für jede Erweiterung $\overline{v}$ für eine beliebige, stetige t-Norm, gilt $\overline{v}(\alpha) = 1$.*

Weiterhin gilt: $\Gamma \models \alpha$ in BL, falls gilt: für jede Interpretation $v : prop \to [0,1]$, und für jede Erweiterung $\overline{v}$ für eine beliebige, stetige t-Norm, falls $\overline{v}(\gamma) = 1$ f.a. $\gamma \in \Gamma$, dann $\overline{v}(\alpha) = 1$:

Also, für die Tautologien/Theoreme von BL gilt: egal in welcher t-Norm wir sie interpretieren, unsere Theorem sind immer Tautologien; und umgekehrt ist jede Formel, die unter jeder Evaluation in einer t-Norm wahr ist, ein Theorem von Hajeks Logik. Deswegen heißt sie auch **basic logic**, oder kurz BL.

13.1 Syntax und Semantik von BL

BL hat drei "primitive" Konnektoren; das bedeutet, aus diesen dreien werden alle anderen definiert, und nur für diese drei brauchen wir eine Semantik. Die Konnektoren sind

$$\&, \rightarrow, \bot;$$

sie werden interpretiert als

- $*$ (eine beliebige stetige t-Norm),
- $\Rightarrow$ (das dazugehörige Residuum),
- und 0, die arithetmetische 0.

Wir haben

- $Var = \{p_0, p_1, ...\}$ sind wohlgeformte Formeln,
- und falls ϕ, χ wohlgeformt sind, sind es auch $\phi\&\chi, \phi \rightarrow \chi, \bot$,
- sonst (erstmal) nichts.

Es gibt aber eine Reihe anderer Konnektoren, die durch diese definiert werden können:

- $\neg\phi \equiv \phi \rightarrow \bot$
- $\phi \wedge \chi \equiv \phi\&(\phi \rightarrow \chi)$
- $\phi \vee \chi \equiv ((\phi \rightarrow \chi) \rightarrow \chi) \wedge ((\chi \rightarrow \phi) \rightarrow \phi)$
- $\phi \leftrightarrow \chi \equiv (\phi \rightarrow \chi)\&(\chi \rightarrow \phi)$

Was wir haben auf der semantischen Seite ist:

$$v : Var \rightarrow [0, 1];$$

wir erweitern das zu $\overline{v}$ wie folgt:

- $\overline{v}(\phi\&\chi) = \overline{v}(\phi) * \overline{v}(\chi)$
- $\overline{v}(\phi \rightarrow \chi) = \overline{v}(\phi) \Rightarrow \overline{v}(\chi)$.

- $\overline{v}(\bot) = 0$.

Natürlich gilt, wie wir schon gesehen haben, folgendes

Lemma 19 *1.* $\overline{v}(\phi \wedge \chi) = min(\overline{v}(\phi), \overline{v}(\chi))$;

2. $\overline{v}(\phi \vee \chi) = max(\overline{v}(\phi), \overline{v}(\chi))$.

Wir beweisen nur die erste Behauptung.

Beweis 1. Wir zeigen dass $x*(x \Rightarrow y) = min(x, y)$. Fall i: $x \leq y$; dann ist $x \Leftarrow y = 1$; also ist $x*(x \Rightarrow y) = x = min(x, y)$. Fall ii: falls $y < x$. Dann gibt es ein $z \in [0, 1]$, s.d. $z*x = y$, denn für $z = 0$ haben wir $z*x = 0$, und für $z = 1$ haben wir $z*x = x$, und $y \in [0, x)$ und $*$ ist stetig. Für das maximale $z : z*x = y$ haben wir $z = x \to y$, also ist $x*(x \to y) = y = min(x, y)$. ⊣

BL ist also die Logik aller stetigen t-Normen. Man könnte meinen, dass es damit recht "künstlich" oder "umstndlich" ist. Tatsächlich gibt es aber auch eine recht einfache Beweistheorie für BL. Wir definieren nun die Beweistheorie für BL; wir geben eine Hilbert Kalkül:

(BL1) $(\alpha \to \gamma) \to ((\gamma \to \beta) \to (\alpha \to \beta))$

(BL2) $(\alpha \& \beta) \to \alpha$

(BL3) $(\alpha \& \beta) \to (\beta \& \alpha)$

(BL4) $(\alpha \& (\alpha \to \beta)) \to (\beta \& (\beta \to \alpha))$

(BL5) $(\alpha \to (\gamma \to \beta)) \leftrightarrow ((\alpha \& \gamma) \to \beta)$

(BL6) $((\alpha \to \gamma) \to \beta) \to (((\gamma \to \alpha) \to \beta) \to \beta)$

(BL7) $\bot \to \alpha$

Natürlich stehen diese 7 wieder für unendlich viele Formeln, da wir den Abschluss unter Substitution haben. NB: die Axiome referieren nur auf die primitiven Konnektoren, denn alle anderen sind ja über sie definiert.

Die einzige Ausnahme ist (A5); allerdings ist hier der Doppelpfeil nur eine Abkürzung für zwei Implikationen. Einige Worte zur Erklärung:

BL1 ist eine Form des klassischen Dreierschlusses: falls γ aus α folgt, dann folgt alles, was γ folgt, auch aus α.

BL2 ist klar

BL3 garantiert die Kommutativität von &.

BL4,BL5 definieren die Interaktion von $\&, \rightarrow$.

BL6 ist eine Form des Beweises über Fälle: falls β aus $\gamma \rightarrow \alpha$ folgt und aus $\alpha \rightarrow \gamma$, dann gilt β – denn da Formeln in $[0,1]$ interpretiert werden, muss eines der beiden gelten. In gewissen Sinne implementiert dass die lineare Ordnung der Wahrheitswerte – eine Sache die in nichtklassischen Logiken nicht mehr garantiert wird.

BL7 ist klar.

Die einzige Deduktionsregel ist **modus ponens** (MP); Beweise sind wie für klassische Logik definiert.

Folgendes Ergebnis ist eine fundamentale Voraussetzung dafür, dass unser Kalkül korrekt ist:

Lemma 20 *Alle BL-Axiome sind fuzzy Tautologien, d.h. für jede Interpretation $v : var \rightarrow [0,1]$ und jede stetige t-Norm $*$ haben wir $\overline{v}(\alpha) = 1$*

Das ist offensichtlich für (A2),(A3),(A7); für die anderen Axiome muss man etwas arbeiten. Wir lassen den Beweis aus. Eine weitere Eigenschaft ist:

Lemma 21 *Falls $\alpha \rightarrow \gamma$ eine Tautologie ist, α eine Tautologie ist, dann ist γ eine Tautologie.*

Beweis. Wir haben oben gesehen dass f.a. t-Normen, $x \Rightarrow y = 1$ gdw. $x \leq y$. Da α eine Tautologie ist, ist $\overline{v}(\alpha) = 1$; es muss aber $\overline{v}(\alpha) \leq \overline{v}(\gamma)$ sein, also $\overline{v}(\gamma) = 1$. $\dashv$

Was das zeigt ist das unser Kalkül **korrekt** ist: alles, was wir damit beweisen, ist eine Tautologie. Man sagt auch: die Menge der BL-Tautologien sind abgeschlossen unter *modus ponens*.

Das Gegenstück hierzu ist die **Vollständigkeit**: jede Tautologie soll in unserem Kalkül beweisbar sein. Man zeigt das wie folgt: zunächst gibt man BL eine sog. **algebraische Semantik**, die BL-Algebren. Hierfür lässt sich leicht Vollständigkeit beweisen. Dann zeigt man, dass sich jede BL-Algebra als ein Produkt von t-Normen darstellen lässt. Der Beweis ist also eher algebraisch; ich werde ihn hier nicht darstellen.

13.2 BL-Tautologien

Wir zeigen zunächst mal, was man nicht beweisen kann in BL:

Lemma 22 *Die folgenden sind keine BL-Tautologien:*

Gödel $\alpha \to \alpha \& \alpha$

Łuk $\neg\neg\alpha \to \alpha$

Prod $\neg\alpha \vee ((\alpha \to (\alpha \& \gamma) \to \gamma)$

GAD $\alpha \vee \neg\alpha$

Was haben die Namen zu bedeuten? Das sind jeweils die Logiken, in denen sie gelten. Wir sehen also das BL nicht die Logik einer bestimmten t-Norm ist! Das nächste Ergebnis ist ebenfalls ernüchternd:

Lemma 23 *Falls* $\Gamma, \alpha \vdash_{BL} \beta$*, dann gilt nicht notwendig* $\Gamma \vdash \alpha \to \beta$*.*

D.h. wir verlieren das Deduktionstheorem. Das hat einen sehr einfachen Grund: in einem Beweis von $\Gamma, \alpha \vdash_{BL} \beta$ dürfen wir das α beliebig oft nutzen. In BL macht es aber einen Unterschied, ob wir α oder $\alpha \& \alpha$ beweisen! Es gibt aber ein etwas schwächeres Ergebnis, nämlich das sog. lokale Deduktionstheorem:

Theorem 24 $\Gamma, \alpha \vdash_{BL} \beta$ *genau dann wenn es ein* n *gibt, so dass* $\Gamma \vdash \underbrace{(\alpha \& .... \& \alpha)}_{n \text{ mal}} \to \beta$*.*

Das DT ist hier lokal in dem Sinne: wir müssen in der Implikation lokal festlegen, wie oft wir die Prämisse brauchen. Wir werden das nicht beweisen, aber wir erklären kurz wie es hierzu kommt: wir haben allgemein, in allen Fuzzy Logiken,

$$v(\alpha \& ... \& \alpha) \leq v(\alpha)$$

Gleichzeitig gilt, ebenso in allen Fuzzy Logiken,

$$v(\alpha \to \beta) = 1 \text{ gdw. } v(\alpha) \leq v(\beta)$$

Das bedeutet: wir können den Fall haben dass $v(\underbrace{(\alpha \& .... \& \alpha)}_{n+1 \text{ mal}} \to \beta) = 1$, aber $\underbrace{(\alpha \& .... \& \alpha)}_{n \text{ mal}} \to \beta < 1$. Das ist der Grund warum die Eigenschaften der Idempotenz und Archimedeität so wichtig waren. Nun können wir einige Tautologien von BL beweisen:

Übung

1. Zeigen Sie, dass folgende Formel eine Tautologie ist in Produkt Logik: $\neg p \vee ((p \to (p \& q)) \to q)$

2. Beweisen Sie die Formel $p \to \neg\neg p$ im Hilbert-Kalkül für klassische Logik. Sie dürfen dabei DT benutzen!

Lösung 1 Fallunterscheidung: $v(p) = 0$, $v(p) > 0$

Lösung 2 Wir zeigen erstmal $p \vdash \neg\neg p$

1	$\neg p \to (p \to \bot)$	C4
2	$(\neg p \to (p \to \bot)) \to (p \to (\neg p \to \bot)$	Beispiel 3, DT
3	$p \to (\neg p \to \bot)$	MP aus 1,2
4	p	Annahme
5	$\neg p \to \bot$	MP aus 3,4
6	$(\neg p \to \bot) \to \neg\neg p$	C5
7	$\neg\neg p$	MP aus 5,6

Das Beweist $p \vdash \neg\neg p$ (ein Beweis von $\neg\neg p$, der Annahme p nutzt). Qua DT bekommen wir $\vdash p \to \neg\neg p$ QED.

13.3 Theorien und ihre Anwendung – anhand BL

BL liefert uns zunächst nur Sätze, die allgemein gültig sind (in jeder stetigen t-Norm). Die sind natürlich für die Anwendungen nicht sehr interessant; was aber interessant ist ist die Axiomatik im Rahmen dieser Logik. Wir führen nun den Begriff der **Theorie** ein: eine Theorie T ist eine (endliche) Menge von Formeln. Wir können auch den Begriff des Beweises erweitern: wenn wir

$$\vdash_{BL} \alpha$$

schreiben für: α ist beweisbar in BL, dann meinen wir mit

$$T \vdash_{BL} \alpha$$

dass α in BL *und* den zusätzlichen Annahmen in T beweisbar ist. Z.B.: wenn wir das $p\&q$ in unserer Theorie haben, dann ist in BL p beweisbar, denn

$$\{p\&q\} \vdash_{BL} p.$$

Bevor wir Theorien anwenden können, müssen wir uns noch kurz über propositionale Semantik Gedanken machen. Nehmen wir an, p bedeutet so etwas wie: "es ist kalt". Klassische gesprochen ist das wahr oder falsch; für uns sind die Dinge anders: die Bedeutung von p ist eine Funktion

$$p : \mathbb{R} \to [0, 1].$$

Das ist so zu verstehen: wir messen die Temperatur, und abhängig davon ändert sich der Wahrheitswert von p. Jetzt können wir ein Axiom hinzufügen:

$$p \to q,$$

wobei q die Bedeutung hat: "Heizung läuft". Auch das ist ein numerischer Parameter, dessen Wahrheit in $[0, 1]$ liegt, wobei die Skala natürlich zu definieren ist. Nun bedeutet

$$v \models p \to q$$

nicht, dass sobald $v(p) = 1$, dann $v(q) = 1$, sondern es bedeutet:

$$v(p) \leq v(q)$$

Die Formel ist also erfüllt, wenn wir "mehr heizen, als es kalt ist".

Nun füttern wir unser Kontrollsystem mit Daten zu unseren Variablen, und wann immer $T \vdash_{BL} \alpha$ gilt, soll unser System sicherstellen dass α gilt. Wir müssen natürlich sicherstellen, dass es in seiner Macht liegt.

Was wir dabei natürlich eigentlich benutzen (theoretisch) ist die Relation $\models_{BL}$: wir können unsere Meßdaten als eine Belegung v auffassen; die erweitern wir auf eine Art und Weise so dass f.a. nichtlogischen Axiome ϕ gilt dass

$$\overline{v}(\phi) = 1.$$

Wir möchten also, falls gilt:

$$\overline{v} \models_{BL} \alpha \text{ impliziert } \overline{v} \models_{BL} \beta$$

dann soll unser System Sorge tragen dass β gilt. Der Punkt ist: mit unserem Vollständigkeitsergebnis fallen $\vdash_{BL}$ und $\models_{BL}$ zusammen! Hier gibt es einige Dinge zu beachten (Konsistenz, Kontrolle).

Wenn man noch weiter geht, könnte man verlangen: β soll so manipuliert werden, dass $\overline{v}(\beta)$ den (mindesten) Wert annimmt, den es logisch Annehmen muss. Leider ist das nicht so ohne weiteres möglich, unser Kalkül ist nicht so stark, dass es das leisten könnte: nimm an, wir haben v so dass

(85) $min(\overline{v}(\phi), \overline{v}(\phi \to \beta)) = x$

Es kann dennoch sein dass

(86) $\overline{v}(\beta) < x$

Dazu müssen wir annehmen, dass

(87) $\overline{v}(\phi) < 1$,

außerdem

(88) $\overline{v}(\beta) < \overline{v}(\phi)$

Dann ist

(89) $\overline{v}(\phi) \Rightarrow \overline{v}(\beta) = max\{z : \overline{v}(\phi) * z \leq \overline{v}(\beta)\}$

Falls nun $* = \cdot$, dann ist

(90) $\overline{v}(\beta) < min(\overline{v}(\phi) \Rightarrow \overline{v}(\beta), \overline{v}(\phi))$

Der einzige Fall, wo das nicht gilt, ist tatsächlich, falls $* = min$, also Gödel Logik.

Dann stellt sich die Frage: warum brauchen wir BL, wenn wir ohnehin nur mit den diskreten Werten richtig arbeiten können? Wir können auch mit BL fuzzy Sachverhalte erfassen: insbesondere $\rightarrow$ erlaubt es uns, beliebige Größenrelationen zu beschreiben, denn es gilt

$$\phi \rightarrow \beta \text{ genau dann wenn } \phi \leq \beta.$$

Wir können also folgendes machen: wir legen für $v(p)$ einen bestimmten Wert fest, und nehmen das Axiom $\phi \rightarrow p$. Wann immer

(91) $\overline{v}(\phi) > \overline{v}(p)$,

dann haben wir

(92) $\overline{v}(\phi \rightarrow p) = 0$

und mit diesem Sachverhalt können wir weiter räsonnieren (über Negation etc.).

14 Łukasiewicz Logik, Wetten und Spiele

Es gibt einige interessante Verbindungen von Fuzzy Logik, Spieltheorie und Wahrscheinlichkeit. Als einen kleinen Ausblick würde ich das gerne hier vorstellen. Es gibt zwei Kalküle für Ł, der erste ist der Kalkül für BL mit folgendem Axiom:

$(\neg\neg\alpha) \to \alpha$

Es gibt aber noch einen kompakteren Kalkül für Łukasiewicz Logik, der hat folgende Axiome:

L1 $\alpha \to (\beta \to \alpha)$

L2 $(\alpha \to \beta) \to ((\beta \to \gamma) \to \alpha \to \gamma))$

L3 $((\alpha \to \beta) \to \beta) \to ((\beta \to \alpha) \to \alpha)$

L4 $(\neg\beta \to \neg\alpha) \to (\alpha \to \beta)$

(L5 $(\alpha \to \beta) \vee (\beta \to \alpha)$)

L5 ist tatsächlich ableitbar aus den anderen, das wurde aber erst relativ spät gezeigt (Hilbert Kalküle!), dewegen steht es in Klammern. Die einzige Inferenzregel ist wie immer MP. Es fällt auf dass es keine Axiome für & gibt; der Grund ist folgender: in Ł kann & definiert werden aus $\to, 0$:

$$\neg\alpha \equiv \alpha \to\perp \tag{93}$$

$$\alpha\&\beta \equiv \neg(\alpha \to \neg\beta) \tag{94}$$

14.1 Übung

Beweisen Sie, dass die Axiome von BL ableitbar sind aus (L1-L4), und umgekehrt!

Ł ist besonders interessant aus einem bestimmten Grund; es gibt nämlich Axiome, mit denen wir die Wahrheitswerte auf eine beliebige endliche Zahl beschränken können.

AT1 $\nabla\alpha \leftrightarrow \overbrace{\neg\nabla\alpha\&...\&\neg\nabla\alpha}^{n-2 \text{ mal}}$

AT2 $\bigvee_{i=1}^{n-1}(\alpha \leftrightarrow \overbrace{\neg\nabla\alpha\&...\&\neg\nabla\alpha}^{i \text{ mal}}$

Damit bekommen wir Łukasiewicz Logik mit endlich vielen Wahrheitswerten, nämlich

$$L_n = \{0, \tfrac{1}{n-1}, \tfrac{2}{n-1}, ..., \tfrac{n-2}{n-1}, 1\}$$

Wir haben dann die Valuation

(95) $v(\nabla\alpha) = \dfrac{n-2}{n-1}$

Man kann das scheinbar auch ohne ∇ axiomatisieren, geht aber schwieriger.

15 Gödel Logik (Fragment)

Wir kommen nun zur Logik der Gödel Norm. Die Menge der wohlgeformten Formeln ist ebenso definiert wie für BL. $\overline{v}$ wie folgt:

- $\overline{v}(\phi \& \beta) = min(\overline{v}(\phi), \overline{v}(\beta))$
- $\overline{v}(\phi \to \beta) = \overline{v}(\phi) \Rightarrow \overline{v}(\beta)$.
- $\overline{v}(\bot) = 0$.

Die anderen Konnektoren sind wie üblich definiert:

- $\neg\phi \equiv \phi \to \bot$
- $\phi \vee \beta \equiv ((\phi \to \beta) \to \beta) \wedge ((\beta \to \phi) \to \phi)$
- $\phi \leftrightarrow \beta \equiv (\phi \to \beta) \& (\beta \to \phi)$

Das $\wedge$ ist nicht nötig, da es äquivalent ist zu &.

16 Spieltheoretische Semantik für Fuzzy Logiken (Auftakt)

16.1 Spieltheoretische Semantik

Für klassische Logik gibt es eine einfache spieltheoretische Interpretation. Es gibt Spieler 1 und 2. Spieler 1 nimmt die Rolle W, Spieler 2 die Rolle F. W will die Formeln wahr machen, F will sie falsch machen; wir sagen dann, sie haben jeweils gewonnen. Am Anfang haben wir die Formel α. Weiterhin haben wir eine atomare Belegung

$$\sigma : prop \to \{0, 1\}$$

die jeder atomaren Proposition einen Wahrheitswert zuweist. Nun folgende Definitionen:

- $\alpha = \beta \wedge \gamma$ F darf sich eine Teilformel aussuchen, auf der das Spiel weiterläuft
- $\alpha = \beta \vee \gamma$ W darf sich eine Teilformel aussuchen, auf der das Spiel weiterläuft
- $\alpha = \neg\beta$ Spieler 1 und 2 tauschen die Rollen W und F: W muss falsifizieren, F muss verifizieren. Auch die obigen Regeln gelten nun für die umgekehrten Rollen.

Das Spiel ist beendet, falls die Formel, auf der gespielt wird, die Form $p \in prop$ hat.

- Falls $\sigma(p) = 0$, dann hat derjenige Spieler, der Rolle F hat, gewonnen.
- Falls $\sigma(p) = 1$, dann hat derjenige Spieler, der Rolle W hat, gewonnen.

Am Ende sagen wir:

- α ist wahr unter σ, wenn es für Spieler 1 eine Strategie gibt, wie er **immer** gewinnt (unabhängig von den Entscheidungen von Spieler 2).
- α ist falsch unter σ, wenn es für Spieler 2 eine Strategie gibt, wie er **immer** gewinnt (unabhängig von den Entscheidungen von Spieler 1).

- α ist eine Tautologie, wenn es für jedes σ eine Strategie gibt, so dass Spieler 1 immer gewinnt (unabhängig von den Entscheidungen von Spieler 2).

- α ist ein Widerspruch, wenn es für jedes σ eine Strategie gibt, so dass Spieler 2 immer gewinnt (unabhängig von den Entscheidungen von Spieler 1).

Soweit, so simpel. Man kann dieses Spiel sogar mit Quantoren spielen (bei $\forall$ darf F aussuchen), $\rightarrow$ ist ebenfalls kein Problem. Aber wie soll das funktionieren mit Fuzzy Logik?

16.2 Ulam Reny Spiele

Stanislav Ulam. Adventure of a Mathematician, 1976:

> Someone thinks of a number between one a one million (which is just less than 2^{20}). Another person is allowed to ask up to twenty questions, to each of which the first person is supposed to answer only yes or no. Obviusly the number can be guessed by asking first: Is the number in the first half-million? and then again reduce the reservoir of numbers in the next question by one-half, and so on. Finally, the number is obtained in less than $log_2(1,000,000)$. Now suppose one were allowed to **lie** once or twice, then how many questions would one need to get the right answer? One clearly needs more than n questions for guessing one of the 2^n objects because one does not know when the lie was told. This problem is not solved in general.

Das führt uns direkt zu Fuzzy Logik und deren spieltheoretischer Semantik. Man sieht also: es gibt hier so etwas wie "Lügen"!

Etwas formaler sieht das wie folgt aus:

- Zwei Spieler: Frager F, Antworter A;
- Der Suchraum $S = \{1, ..., n\}$, $n \in \mathbb{N}_+$;
- A wählt in $z \in S$;
- Um z zu identifizieren stellt F Ja-Nein Fragen an A;
- Man nennt das auch das 20-Fragen Spiel!

Natürlich ist klar: jede Ja/Nein Frage hat die Form:

> Ist $x \in M$, für $M \subseteq S$?

Das bedeutet: eine Frage ist eigentlich nur eine Menge M, und die Frage hat die Form: Ist $z \in M$?

Es gibt jetzt wieder zwei einfache Strategien von F:

1. Wir gehen einfach durch die Liste $1, ..., n$ – also n Fragen.
2. Wir halbieren jedes Mal die Menge S. So brauchen wir höchstens $\lceil log_2(n) \rceil$ Fragen.

Wir formalisieren das wie folgt:

- Eine Frage Q ist eine Menge $Q \subseteq S$
- Die Antwort ist ja, falls $x \in Q$
- und nein andernfalls.

Ein wichtiger Begriff für unser Spiel ist der des **Zustands**. Ein Zustand s ist für uns die **charakteristische Funktion** einer Teilmenge $\Omega \subseteq S$, d.h.

(96) $$s(x) = \begin{cases} 1, \text{ falls } x \in \Omega \\ 0 \text{ andernfalls} \end{cases}$$

s ist also unser Wissen: je stärker wir die Menge, die z enthält, einschränken können, desto größer ist unser Wissen.

- Wir starten also mit $\overline{1}$, wobei $s_0(x) = 1$ f.a. $x \in S$
- Unser Ziel ist ein Zustand $s(x) = 1$ gdw. $x = z$
- Es gibt noch einen weiteren speziellen Zustand $\overline{0}$ so dass $\overline{0}(x) = 0$. Das entspricht der *inkonsistenten Information.*

Das interessante ist jetzt: Zustände sind partiell geordnet nach **Informativität**. Wir schreiben

$$s_1 \leq s_2 \text{ gdw. für alle } x,\, s_1(x) \leq s_2(x)$$

Das bedeutet: s_1 ist informativer als s_2 (wichtig: nicht vertauschen!). Damit kann man auch Operationen definieren:

- $s_1 \wedge s_2(x) = min(s_1(x), s_2(x))$ ist der am wenigsten informative Zustand, der informativer ist als s_1, s_2 (also die Information der beiden vereint)
- $s_1 \vee s_2(x) = max(s_1(x), s_2(x))$ ist die Information, die s_1, s_2 gemeinsam ist
- $\neg s_1(x) = \overline{1} - s_1(x)$...

Es ist also offensichtlich, dass die Zustände eine Boolesche Algebra bilden. Das ist auch offensichtlich, denn jeder Zustand repräsentiert einfach eine Menge. Damit ist das Kapitel natürlich auch schnell abgehakt.

16.2.1 Spiele mit Lügen

Jetzt wird es interessant, denn nun betrachten wir die Möglichkeit, dass unser Gegenüber lügt.

- Zwei Spieler: Frager F, Antworter A;
- Der Suchraum $S = \{1, ..., n\}$, $n \in \mathbb{N}_+$;
- A wählt in $z \in S$;
- Um z zu identifizieren stellt F Ja-Nein Fragen an A;
- Dem Antworter A stehen nun $e \geq 0$ **Lügen** zur Verfügung.

Also: Ein Kandidat x kann nicht z sein, falls x $e+1$ Fragen falsifiziert.

Aber: falls ein Kandidat $e' \leq e$ Fragen falsifiziert, dann kann er noch nicht ausgeschlossen werden!

Das bedeutet: wir können Zustände nicht einfach als Mengen auffassen, denn es macht einen großen Unterschied, ob ein Objekt zwei Fragen nicht erfüllt, oder nur eine! Wir müssen also in jedem Schritt uns merken, welche Objekte welche Frage nicht erfüllen.

Beispiel 1 $F1$: ist x gerade? $A1$: Ja.
$F2$: ist x gerade? $A2$: Nein.

Hier sind die beiden Antworten inkompatibel, also ist die Anzahl der Lügen um 1 gesunken (eine Antwort war eine Lüge; wir wissen natürlich nicht welche!).

Beispiel 2 Nimm an, $e = 1$. Weiterhin:
$F1$: ist x gerade? $A1$: Ja.
$F2$: ist x gerade? $A2$: Ja.

In diesem Fall folgt, dass z gerade ist, denn wir können nicht zweimal lügen! Das bedeutet also: $a \wedge a$ ist eine stärkere Aussage als a!

Wie formaliseren wir unsere Zustandsfunktion? Wir nehmen eine Zustandsfunktion

(97) $s(x) = \begin{cases} \frac{e+1-i}{e+1} \text{ falls } x\ i \leq e \text{ Fragen falsifiziert} \\ 0 \text{ falls } x \text{ mehr als } e \text{ Fragen falsifiziert} \end{cases}$

Das ist eine Funktion

$$s : S \to \{0, \tfrac{1}{e+1}, ..., \tfrac{e}{e+1}, 1\}$$

Man kann sich das also vorstellen als eine Funktion mit e+2 **Wahrheitswerten**.

Als nächstes brauchen wir Funktionen, die die Information einer Frage Fi *und* Antwort Ai bemessen. Das geht wie folgt: wir repräsentieren das mit einer Funktion

$$i_{Fi,Ai} : S \to \{\tfrac{e}{e+1}, 1\}$$

definiert durch

(98)

$$i_{Fi,Ai}(x) = \begin{cases} 1 & \text{, falls } (x \in Fi \text{ und } Ai = \text{Ja}) \text{ oder } (x \notin Fi \text{ und } Ai = \text{Nein}) \\ \frac{e}{e+1} & \text{, falls } (x \in Fi \text{ und } Ai = \text{Nein}) \text{ oder } (x \notin Fi \text{ und } Ai = \text{Ja}) \end{cases}$$

Sei s der Zustand vor Fi. Dann berechnen wir den Nachfolgezustand, den wir durch Fi und Ai bekommen durch

(99) $s' = s \odot i_{Fi,Ai}$

Das ist definiert durch

(100) $s \odot i_{Fi,Ai}(x) = s(x) \odot i_{Fi,Ai}(x)$

wobei also $\odot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ definiert ist durch

(101) $x \odot y = max(0, x + y - 1)$

So können wir also unsere Zustände beschreiben. Und was hat das mit Fuzzy Logik zu tun? Kommen wir zu.

16.3 n-wertige Łukasiewicz Logik

Wir haben bislang Łukasiewicz Logik aufgefasst als Logik mit unendlich vielen Wahrheitswerten. Historisch hat Łukasiewcz glaube ich angefangen mit 3. Man kann das aber beliebig im Endlichen erweitern, es gibt als für jedes $n > 2$ eine n-wertige Łukasiewcz Logik (wahrheitswerte sind dabei immer linear geordnet, ist ja eine Fuzzy Logik).

Hier kommt nun die Korrespondenz:

Ulam-Reny Spiel mit n Lügen $\cong$ Łukasiewicz Logik mit n Wahrheitswerten

Das kann man wie folgt aufbauen: sei

(102) $I_n = \{0, \frac{1}{n}, ..., \frac{n-1}{n}, 1\}$

Wichtig ist: I_n ist geschlossen unter Łukasiewicz t-Norm und Łukasiewicz Residuum. Sprich: wir müssen gar nichts weiter ändern, wir erhalten n-wertige Łukasiewicz Logik einfach, indem wir die möglichen Wahrheitswerte von atomaren Propositionen beschränken. Wir kommen also niemals I_n raus!

Wenn wir Łukasiewicz Logik mit n Wahrheitswerten nehmen, dann ist jede Formel ϕ der Logik, bestehend aus n verschiedenen propositionalen Variablen, eine (fuzzy) Wahrheitsfunktion

(103) $\phi : (I_n)^n \to I_n$

so wie man das von klassischer Logik kennt.

Als nächstes definieren wir

(104) $C(n,e) = \{\frac{1}{e+1}, \frac{e}{e+1}\}^n$

(also n-faches kartesisches Produkt der Menge mit sich selber).

Folgendes Theorem (Mundici, 1992):

Theorem 25 *Für jede Funktion $f : C(n,e) \to I_{e+2}$ gibt es eine Łukasiewicz Formel ϕ so dass f_ϕ (die Funktion der Formel), restringiert auf $C(n,e)$, gleich f ist.*

Weiterhin: für jede Łukasiewicz Formel ϕ ist die Restriktion von f_ϕ auf $C(n,e)$ eine Funktion $s : C(n,e) \to \{0, \frac{1}{e+1}, ..., \frac{e}{e+1}, 1\}$

Wir bekommen dann folgende Korrespondenz:

Ulam-Reny Spiel	Łukasiewicz Logik
max e Lügen	Wahrheitswerte $\{0, \frac{1}{e+1}, ..., \frac{e}{e+1}, 1\}$
Suchraum $S = \{0,1\}^n$	$C(n,e) = \{\frac{1}{e+1}, \frac{e}{e+1}\}^n$
$z \in S$	ein Punkt in $C(n,e)$
Anfangszustand (kein Wissen)	Łukasiewicz Tautologie
Finaler Zustand (z bekannt)	ϕ hat genau eine Belegung in $C(n,e)$
Aktueller Zustand	ϕ so dass $f_\phi = s \in C(n,e)$
Frage Ai	Teilmenge von $C(n,e)$
Positive Antwort ψ auf Ai	Formel ϕ so dass $f_\psi = f_\phi$ auf $C(n,e)$
Zustand s nach $A1, ..., Ai$	Konjunktion der entsprechenden Formeln
Menge der ausgeschlossenen Nummern	Punkte in $C(n,e)$ die ϕ falsifizieren

17 Der Akinator und Łukasiewicz Moisil

17.1 Das Problem, der einfache Fall

Der Akinator hat eine Datenbank mit (fiktiven) Charakteren.

- Don Vito Corleone
- Hannibal Lecter
- Prinzessing Leia
- Jean-Luc Picard
- Cruella de Vil

Wir nennen die Menge der Charaktere C.

Zudemn hat der Akinator eine (binäre) Merkmalszuschreibung für jeden Charakter:

	Corleone	Lecter	Leia	Picard	Cruella
villain	1	1	0	0	1
excentric hair	0	0	1	1	1
bossy	1	1	1	1	1
european	1	1	0	1	1

Eine Frage des Genies lautet dann:

`Is your character bossy?`

Eine **Frage** ist ein Element $\delta \in L_2^C$ (hier ist $L_2 = \{0, 1\}$)

Die **Beschreibung** der Charaktere ist eine Menge von Fragen, also

$$\Delta \subseteq L_2^C$$

Eine **Runde** ist ein paar (Frage, Antwort). Man schreibt das so:

$$\text{Runde } Q_\delta^a, \text{ wobei } a \in L_2, \text{ und } \delta \in \Delta.$$

Z.B.:

`Is your character a villain?  No.`

wird: $Q^0_{villain}$

Ein **Spiel** ist eine Folge von Runden $\gamma = Q_{\delta 1}^{a1} ... Q_{\delta n}^{an}$

Beispiel Unser Wissen nach einer Runde, ist, z.B. $Q^0_{villain}$, wäre $K^0_{villain}$ (wie knowledge):

	Corleone	Lecter	Leia	Picard	Cruella
villain	1	1	0	0	1
$K^0_{villain}$	0	0	1	1	0

Also besteht das **Wissen** nach einer Runde Q^a_δ aus

(105) $K^a_\delta = 1_{\delta^{-1}(a)} \in L^C_2$

wobei 1_X die charakteristische Funktion ist von $X \subseteq C$.

Beispiel, fortgesetzt Nimm an, wir haben das Spiel mit $(Q^0_{villain}, Q^1_{european})$. Dann bekommen wir:

	Corleone	Lecter	Leia	Picard	Cruella
villain	1	1	0	0	1
$K^0_{villain}$	0	0	1	1	0
$K^1_{european}$	1	1	0	1	1

Das characterisiert eindeutig **Jean-Luc Picard**. Also: nach jeder Runde wählen wir eine Menge

(106) $1_{Ck} = K^{a1}_{\delta 1} \wedge ... \wedge K^{ak}_{\delta k}$

(wahrheitstheoretisches $\wedge$ auf charakteristischen Funktionen entspricht $\cap$ auf Mengen).

17.2 Der komplexere Fall

Soweit, so einfach. Was ist daran *fuzzy*? Ach so, da war ja was: die Antworten beim Akinator sind:

$$\{\text{Yes, Probably, Don't know, Probably not, No}\}$$

Diese Antworten werden partiell geordnet, nicht nach Wahrheitsgrad, sondern nach **Informationsgehalt**:

Don't know $\prec$ Probably $\prec$ Yes
Don't know $\prec$ Probably not $\prec$ No

Man bekommt grob gesagt das folgende Szenario:

- Charaktere: eine Menge C
- Antworten: eine partielle geordnete Menge $(L_p, \preceq)$ mit $p \geq 2$
- Beschreibungen der Charaktere: $\Delta \subseteq L_p^C$
- Eine Runde (Frage,Antwort): Q_a^δ, wobei $\delta \in \Delta$ und $a \in L_p$
- Das Wissen nach Runde Q_a^δ:

  (107) $K_a^\delta = \bigvee_{a \preceq x} 1_{\delta^{-1}(x)} \in L_2^C$

  NB: obwohl es mehr als zwei Antworten gibt, ist die Auswahl eine (knackige) Menge an Charakteren)

- Ein Spiel: eine Folge $\gamma = Q_{\delta 1}^{a1}...Q_{\delta n}^{an}$ (wie vorher)

Hier fehlen eine Menge Details....

18 Algebraische Semantik: BL-Algebren

Auch die Semantik ist eigentlich geradeaus, wir müssen aber, um bei BL zu bleiben, den Begriff der BL-Algebra einführen. Wir haben ja gesagt dass BL die Logik aller stetigen t-Normen ist; es gibt also erstmal keine einheitlich definierte Klasse von Modellen. Um diese Lücke zu füllen wurden BL-Algebren definiert, die eine korrekte und vollständige Semantik für den Beweiskalkül BL darstellen.

Definition 26 *Ein **Verband** $(M, \wedge, \vee, \leq)$ ist eine partielle Ordnung, in der es f.a. $a, b \in M$ eine kleinste obere Schranke und eine kleinste obere Schranke gibt, d.h.:*

$$\begin{aligned} x \leq y \wedge z &\iff x \leq y \& x \leq z \\ y \vee z \leq x &\iff y \leq x \& z \leq x \end{aligned}$$

Man kann das ganze auch ohne $\leq$ axiomatisieren mit einigen Gleichungen, und $\leq$ ist dann definierbar mittels

(108) $x \leq y \iff x \wedge y = x \iff x \vee y = y$

Wir lassen in Zukunft also das Symbol $\leq$ weg. Ein Verband ist **beschränkt**, falls es noch zusätzlich noch 0,1 gibt mit

(109) $0 \leq x$

(110) $x \leq 1$

(freie Variablen sind immer implizit universell quantifiziert).

Definition 27 *Ein **verbands-geordneter Monoid** ist eine Struktur $(M, \wedge, \vee, *, 0, 1)$ wobei*

1. *$(M, \wedge, \vee, 0, 1)$ ein beschränkter Verband ist*
2. *$(M, *, 1)$ ein Monoid (assoziativ, neutrales Element).*

Man beachte dass das Maximum 1 neutral ist für $*$! Außerdem nehmen wir immer an, dass $*$ **kommutativ** ist:

(111) $x * y = y * x$

Und weiter gehts:

Definition 28 *Ein **residuierter Verband** ist ein **verbands-geordneter Monoid** $(M, \wedge, \vee, *, \Rightarrow, 0, 1)$ mit Konnektor $\Rightarrow$, der das Gesetz des Residuums erfüllt:*

(112) $x \leq y \Rightarrow z \Longleftrightarrow x * y \leq z$

Zu guter letzt:

Definition 29 *Eine **BL-Algebra** ist ein residuierter Verband $(M, \wedge, \vee, *, \Rightarrow, 0, 1)$, welcher zusätzlich folgende Bedingungen erfüllt:*

$$x * (x \Rightarrow y) = x \wedge y \tag{113}$$

$$(x \Rightarrow y) \vee (y \Rightarrow x) = 1 \tag{114}$$

Das erste Axiom ist wichtig für die Interaktion von $*, \wedge$ (schwaches und!), das zweite Axiom ist die sog. **Prälinearität**. Es gibt also auch nichtlineare BL-Algebren!

BL-Algebren sind eine korrekte und vollständige Semantik für BL – trotz Nichtlinearität! Wir sind also hier mal den anderen Weg gegangen: von der Beweistheorie zur Semantik. Um das zu zeigen, brauchen wir eine Reihe von Ergebnissen:

Lemma 30 *In allen BL-Algebren gelten folgende Gleichungen:*

(115)	$x * (x \Rightarrow y)$	$\leq$	x
(116)	y	$\leq$	$x \Rightarrow (x * y)$
(117)	$x \leq y$	*gdw.*	$x \Rightarrow y = 1$
(118)	$x \leq y$	*impliziert*	$x * z \leq y * z$
(119)	$x * (y \vee z)$	$=$	$(x * y) \vee (x * z)$
(120)	$x \vee y$	$=$	$((x \Rightarrow y) \Rightarrow y) \wedge ((y \Rightarrow x) \Rightarrow x)$

Die Beweise sind nicht allzu kompliziert und zeigen, wie BL-Algebren mit Fuzzy Logiken zusammenhängen. Das folgende Ergebnis ist interessant, um zu verstehen welche Rolle Linearität spielt:

Lemma 31 *Eine BL Algebra ist **linear geordnet**, falls gilt: f.a. x, y, entweder $x \wedge y = x$ oder $x \wedge y = y$.*

Umgekehrt ist jeder lineare residuierte Verband eine BL Algebra, sofern f.a. x, y gilt: $x \wedge y = x * (x \Rightarrow y)$.

Wie man die Sprache unserer (propositionalen) Fuzzy Logiken in BL-Algebren interpretiert, ist geradeaus: sei $\mathbf{L} = (M, \wedge, \vee, *, \Rightarrow, 0, 1)$ eine BL-Algebra; eine AL-Algebra Interpretation ist eine Abbildung

$$\tau : Var \to M$$

welche wie folgt auf beliebige Formeln erweitert wird:

$$\tau(\alpha\&\beta) = \tau(\alpha) * \tau(\beta) \tag{121}$$

$$\tau(\alpha \to \beta) = \tau(\alpha) \Rightarrow \tau(\beta) \tag{122}$$

$$\tau(0) = 0 \tag{123}$$

Die weiteren Konnektoren sind nicht atomar und müssen daher nicht interpretiert werden. Unsere Ergebnisse stellen aber sicher:

$$\tau(\alpha \vee \beta) = \tau(\alpha) \vee \tau(\beta) \tag{124}$$

$$\tau(\alpha \wedge \beta) = \tau(\alpha) \wedge \tau(\beta) \tag{125}$$

Damit nun folgendes, zentrales Ergebnis:

Theorem 32 *Es gilt, für beliebige Mengen von Formaln Γ, und beliebiges α, $\Gamma \vdash_{BL} \alpha$ genau dann wenn gilt:*

$$\tau(\gamma) = 1 \textit{ f.a. } \gamma \in \Gamma \textit{ impliziert } \tau(\alpha) = 1.$$

Beachte den Spezialfall $\Gamma = \emptyset$: $\vdash_{BL} \alpha$ gdw. $\tau(alpha) = 1$ f.a. BL-Algebren und alle τ.

Wir haben also eine einheitliche, vollständige Semantik für BL! Das interessante ist dass man f.a. axiomatischen Erweiterungen des Hilbert-Kalküls von BL ebenfalls eine axiomatische Erweiterung von BL-Algebren machen kann, und damit wiederum eine vollständige Semantik bekommt. Man nennt das auch **Algebraisierung**.

19 Fuzzy Prädikatenlogik

19.1 Syntax I: Formeln

Die Syntax ist im Prinzip geradeaus (d.h. wie immer). Es gibt:

- $Var = \{x_1, x_2, x_3, ...\}$, eine abzählbare Menge von Variablen
- $a_1, ..., a_m$ eine endliche (möglicherweise leere) Menge von Konstanten für Individuen
- $P_1, ..., P_n$ eine endliche nichtleere Menge an Prädikaten, zusammen mit der Stelligkeitsfunktion Ω, wobei $\Omega(P_i) \mapsto \mathbb{N}$
- die propositionalen Konnektoren $\&, \rightarrow$
- die Quantoren $\forall, \exists$
- sonst nix!

Funktionssymbole lassen wir weg, ebenso die Gleichheit: fuzzy Gleichheit und fuzzy Funktionen sind nicht ganz einfach.

Um die Formelsprache zu definieren, müssen wir wie immer erstmal wohlgeformte Terme definieren:

1. Falls $t \in Var$, dann ist t ein Term
2. Falls t eine Individuenkonstante ist, ist t ein Term
3. Sonst nix!

Jetzt zu Formeln. Wir nennen die Menge der wohlgeformten Formeln `WFF`

1. Falls P ein Prädikat, $\Omega(P) = n$, $t_1, ..., t_n$ Terme, dann ist $P(t_1, ..., t_n) \in$ `WFF`
2. Falls $\alpha, \beta \in$ `WFF`, dann sind $(\alpha \& \beta), (\alpha \rightarrow \beta) \in$ `WFF`
3. Falls $\alpha \in$ `WFF`, $x \in Var$, dann sind $(\forall x.\alpha), (\exists x.\alpha) \in$ `WFF`
4. Sonst nix!

Soviel zur Syntax. Konventionen sind:

- Quantoren nehmen, falls Klammern weggelassen werden, weiten Skopus

19.2 Semantik

Wie wir gesehen haben, brauchen wir (wie immer) ein Vokabular $V = (P_1, ..., P_n, a_1, ..., a_m)$. Dementsprechend ist das Modell eine Struktur

$$\mathcal{M} = (U, R_1, ..., R_n, c_1, ..., c_m)$$

wobei

1. U das Universum ist
2. $R_i : U^{\Omega(P_i)} \to \mathbf{L}$ eine Fuzzy Relation der passenden Stelligkeit, wobei

   $$\mathbf{L} = (M, \wedge, \vee, *, \Rightarrow, 0, 1)$$

   eine **lineare BL-Algebra** ist.
3. $c_i \in U$ eine Konstante

Weiterhin brauchen wir die Belegungsfunktion

$$v : \mathit{Var} \cup (\{a_1, ..., a_m\}) \to U$$

wobei wir annehmen dass

(126) $v(a_i) = c_i$

Eine weitere wichtige (und übliche) Konvention ist die folgende: wir schreiben

(127) $v \sim_x v'$

falls gilt: f.a. $y \in \mathit{Var} - \{x\}$, $v(y) = v'(y)$

Damit können wir beliebigen Formeln Wahrheitswerte zuweisen, wie folgt:

$$\|P_i(t_1, ..., t_{\Omega(P_i)})\|_{\mathcal{M},v} = R_i(v(t_1), ..., v(t_{\Omega(P_i)}))$$

$$\|\alpha \& \beta\|_{\mathcal{M},v} = \|\alpha\|_{\mathcal{M},v} * \|\beta\|_{\mathcal{M},v}$$

$$\|\alpha \to \beta\|_{\mathcal{M},v} = \|\alpha\|_{\mathcal{M},v} \Rightarrow \|\beta\|_{\mathcal{M},v}$$

$\|\forall x.\alpha\|_{\mathcal{M},v} = \mathbf{inf}\{\|\alpha\|_{\mathcal{M},v'} : v{\sim_x}v'\}$

$\|\exists x.\alpha\|_{\mathcal{M},v} = \mathbf{sup}\{\|\alpha\|_{\mathcal{M},v'} : v{\sim_x}v'\}$

Man beachte: für die Quantoren nutzen wir das Supremum und Infimum, den Minimum und Maximum sind nicht immer definiert, während sup und inf in jeder linearen Ordnung definiert sind. Folgendes ist leicht zu sehen:

- Jede Formel α bekommt so einen Wahrheitswert $\|\alpha\|_{\mathcal{M},v}$, abhängig von Modell und Belegung
- Wenn aber α eine **geschlossene Formel** ist, ist der Wahrheitswert unabhängig von der Belegung:

  (128) $\|\alpha\|_{\mathcal{M},v} = \|\alpha\|_{\mathcal{M},v'}$

  f.a. v, v'
- Insbesondere gilt: falls $\mathbf{L} = \mathbf{B}_1$ (die BA über $\{0,1\}$), dann bekommen wir genau die Semantik für klassische Logik!

Insbesondere können wir hiermit schreiben:

$$\|\alpha\|_{\mathcal{M}}, \text{ vorausgesetzt } \alpha \text{ ist geschlossen}$$

Weiterhin bekommen wir dadurch die semantische Konsequenz:

$$\alpha \models \beta \text{ gdw. f.a. } \mathcal{M}, \|\alpha\|_{\mathcal{M}} = 1 \text{ impliziert } \|\beta\|_{\mathcal{M}} = 1$$

Das reicht fürs Erste.

19.3 Syntax II: Beweistheorie

Wir haben die Formelsprache und die Semantik (basierend auf BL-Algebren) der Prädikatenlogik von BL besprochen. Wir präsentieren nun seine Beweistheorie. Zunächst haben wir natürlich immer noch (BL1-7), die wir hier wiederholen:

(BL1) $(\alpha \to \gamma) \to ((\gamma \to \beta) \to (\alpha \to \beta))$

(BL2) $(\alpha \& \beta) \to \alpha$

(BL3) $(\alpha \& \beta) \to (\beta \& \alpha)$

(BL4) $(\alpha \& (\alpha \to \beta)) \to (\beta \& (\beta \to \alpha))$

(BL5) $(\alpha \to (\gamma \to \beta)) \leftrightarrow ((\alpha \& \gamma) \to \beta)$

(BL6) $((\alpha \to \gamma) \to \beta) \to (((\gamma \to \alpha) \to \beta) \to \beta)$

(BL7) $\bot \to \alpha$

Dazu kommen noch die folgenden **Quantorenaxiome**:

($\forall$1) $(\forall x.\alpha) \to \alpha[t/x]$

($\forall$2) $(\forall x.\alpha \to \beta) \to (\alpha \to \forall x.\beta)$ (vorausgesetzt $x \notin FV(\alpha)$)

($\forall$3) $(\forall x.\alpha \& \beta) \to (\forall x.\alpha) \& \beta$ (vorausgesetzt $x \notin FV(\beta)$)

($\exists$1) $\alpha[t/x] \to \exists x.\alpha$

($\exists$2) $(\forall x.\alpha \to \beta) \to ((\exists x.\alpha) \to \beta)$ (vorausgesetzt $x \notin FV(\beta)$)

Diese Quantorenaxiome sind relativ universell, ebenso wie die Semantik der Quantoren. In Prädikatenlogiken gibt es meist **zwei** Inferenzregeln anstatt einer, nämlich (MP) und **Generalisierung**:

$$\text{(MP)}\ \frac{\alpha \to \beta, \alpha}{\beta} \qquad\qquad \text{(gen)}\ \frac{\alpha}{\forall x.\alpha}$$

Mittels der Regel (gen) können wir beispielsweise beweisen $\forall x.P(x) \to P(x)$. Das ist die einzige Regel, welche Quantoren einführt; der Existenzquantor kann eingeführt werden mit ($\exists$1).