

Seminar „Imperfektion und Datenbanken“ im WS 2003/04

Thema 11:

Benutzerfreundliche Beschreibung von unscharfen Zeitintervallen

Cornelius Kuppler

Betreuer: Gábor Nagypál

1. Einführung

In der vorliegenden Arbeit geht es um die Frage, wie unscharfe Zeitintervalle mit Hilfe von Fuzzy-Mengen geeignet beschrieben und (auch für Nicht-Informatiker) anschaulich dargestellt werden können. Zunächst wird der Frage nachgegangen, welche Merkmale in geschichtlichen Zeitangaben auftreten können und auf welche Weise sie für Imperfektion sorgen. Dann wenden wir uns möglichen Darstellungen, so genannten Repräsentationen, aus dem Bereich der Fuzzy-Mengen zu und wenden sie auf unsere Fragestellung an. Diese werden dann auf ihre Eignung untersucht und miteinander verglichen. Zuletzt folgen einige Beispiele sowie ein „mockup“ für ein Tool, welches konkret nach Eingabe der bekannten Daten die gewünschte(n) Darstellung(en) liefert.

2. Merkmale geschichtlicher Angaben

Die Geschichtsforschung arbeitet, im Gegensatz zu den exakten Naturwissenschaften, letztlich immer mit unsicheren und nicht beweisbaren Daten. Wo in den Naturwissenschaften Beweis und Widerlegung bzw. das Experiment zum Nachvollziehen einer Behauptung stehen, können in der Geschichtsforschung nur historische Dokumente und Relikte herangezogen werden, um aus diesen dann eine mehr oder minder schlüssige und mehr oder minder umstrittene Geschichtsschreibung zu gewinnen. Die Probleme, die sich dabei ergeben, lassen sich in drei Bereiche kategorisieren: Unsicherheit, subjektive Angaben und unpräzise Angaben:

2.1 Unsicherheit

Unsicherheit bedeutet, dass aus verschiedenen Gründen keine sicheren Angaben möglich sind: Entweder, verschiedene historische Dokumente widersprechen sich gegenseitig in ihren Aussagen. Oder verschiedene Historiker(-schulen) sind schlicht verschiedener Meinung über die Einordnung oder womöglich die Historizität einzelner Ereignisse.

Besonders schwierig wird die Einordnung, wenn über das untersuchte historische Ereignis keine Zeitangaben vorhanden sind, sondern nur der Bezug zu einem anderen Ereignis bekannt ist. Dann muss anhand dessen das eigentlich untersuchte Ereignis irgendwie ungefähr eingeordnet werden. Dass dabei eine Menge Unsicherheiten und Uneinigkeiten entstehen ist klar.

2.2 Subjektive Angaben

Ein weiterer Punkt, der beachtet werden muss, ist der, dass historische Ereignisse oder Zeiträume oft gar nicht exakt definiert sind. Man hat nur natürlichsprachliche Termina, die sich im Laufe der Zeit, vielleicht sogar zuerst im Volksmund, eingebürgert haben und die nur vage definiert sind oder sein sollen, da es oft auch Übergangszeiten von der einen in die andere Epoche gab. Wenn wir beispielsweise von der „industriellen Revolution“ oder von der „Barock-Zeit“ sprechen, weiß zwar jeder was gemeint ist und was charakteristisch für die einzelnen Zeiten war. Aber wenn man fragt, wann das war, wird man von jedem eine andere Meinung hören.

Historiker sind aus diesen Gründen dazu übergegangen, von Anfangs- bzw. Endphasen, von Prozessen oder Übergangs- und Kernphasen zu sprechen.

2.3 Unpräzise Angaben

Naturgemäß liegen (historische) Zeitangaben in unterschiedlichen Granularitäten vor. Von Ereignissen der Antike wissen wir häufig nur das Jahr, von anderen den Tag oder die Stunde. Wir wählen Formulierungen wie „im Laufe des Tages“ oder „während des Vormittags“. Abhängig von den Anforderungen reicht uns dies aus oder eben nicht. Wir werden immer dann von „unpräzise“ sprechen, wenn uns die vorhandene Präzision für unsere Zwecke nicht ausreicht. Der Grad der Präzision ist also eine Frage der zugrunde gelegten Zeitskala. Insofern können wir jede zeitliche Angabe als unpräzise bezeichnen, wenn wir eben nur unsere Skala entsprechend verfeinern.

3. Darstellung von Fuzzy-Mengen

Die Haupteigenschaft von Fuzzy-Mengen ist die, dass wir nicht mehr von jedem Wert, den wir

untersuchen, angeben können, ob er in der Menge liegt oder nicht. Vielmehr können wir hierfür nur noch Wahrscheinlichkeiten angeben, so genannte „Zugehörigkeitsgrade“. Der Wert „1“ bezeichnet dann die sichere Zugehörigkeit, der Wert „0“ die sichere Unzugehörigkeit.

3.1 Intervalle oder Punkte?

Bevor wir uns der Darstellung von Fuzzy-Mengen zuwenden können, müssen wir noch kurz klären, ob wir als kleinstes Zeit-Primitiv Intervalle oder Punkte wählen. Grundsätzlich ist beides möglich (oder sogar eine Kombination der beiden). Wir entscheiden uns hier kurz und schmerzlos für Intervalle, da diese wohl anschaulicher für die menschliche Wahrnehmung sind und sich auch einfacher „fuzzyfizieren“ lassen.

3.2 Argument-Werte-Paare

Die allereinfachste Möglichkeit zur Darstellung von Fuzzymengen ist die, dass wir jedem Wert, für den wir einen Zugehörigkeitsgrad angeben können, auflisten.

Möchten wir beispielsweise „die späten Neunziger“ definieren, so könnte dies so aussehen:

$$p(1996)=0.3$$

$$p(1997)=0.6$$

$$p(1998)=1.0$$

$$p(1999)=1.0$$

$$p(2000)=0.5$$

Diese Vorgehensweise ist natürlich recht umständlich, wenig anschaulich und auch nur im diskreten Fall anwendbar.

3.3 Vertikale Repräsentationen

Zur grafischen Darstellung von Fuzzy-Mengen gibt es zwei „vordefinierte“ Möglichkeiten: Die vertikale und die horizontale Repräsentation.

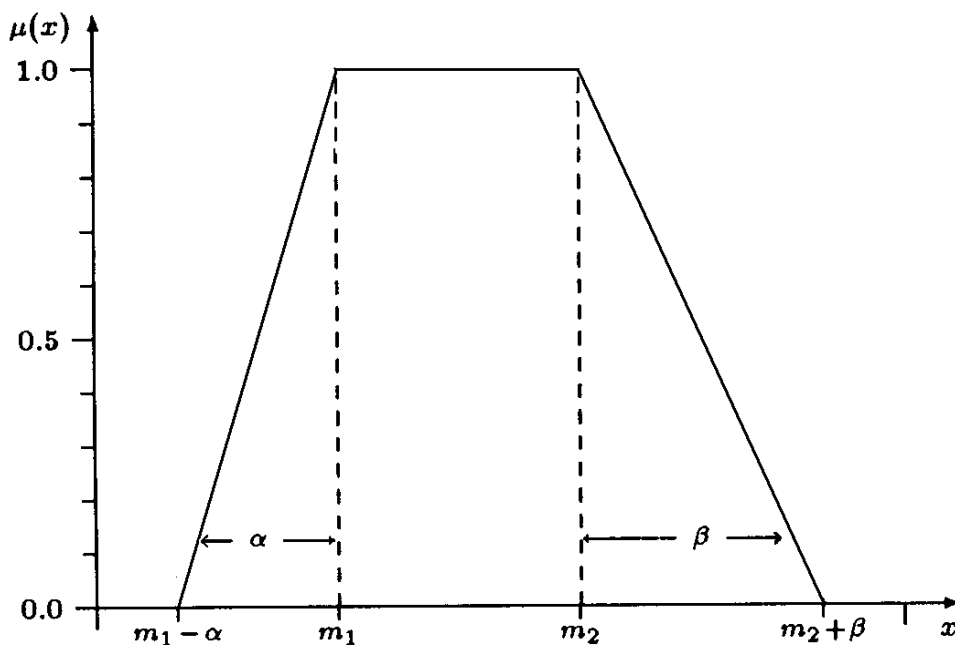
Bei der vertikalen Repräsentation werden einfach die bekannten bzw. gewünschten Werte in ein geeignetes Koordinatensystem eingetragen und dann zu einer abschnittsweise definierten Funktion ergänzt. Je nach Wunsch bzw. Anforderungen können diese Funktionen linear, quadratisch oder sogar noch höherwertiger sein.

3.3.1 Parametrisierte Trapezfunktion

Die parametrisierte Trapezfunktion ist wie folgt definiert:

$$f_{\text{trapez}}(x, m_1, m_2, \alpha, \beta) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{wenn } (x \leq m_1 - \alpha) \vee (x \geq m_2 + \beta) \\ \frac{x - (m_1 - \alpha)}{\alpha} & \text{wenn } (m_1 - \alpha) < x < m_1 \\ 1 & \text{wenn } m_1 \leq x \leq m_2 \\ 1 - \frac{x - m_2}{\beta} & \text{wenn } m_2 < x < (m_2 + \beta). \end{cases}$$

Die Funktion ist also in vier Abschnitte geteilt: links von $m_1 - \alpha$ und rechts von $m_2 + \beta$ ist der Grad der Zugehörigkeit 0. Zwischen m_1 und m_2 ist er 1. Bis hierher haben wir eine normale, scharfe Menge. Die Unschärfe liegt zwischen $m_1 - \alpha$ und m_1 sowie zwischen m_2 und $m_2 + \beta$. Wie steil hier der Kurvenverlauf ist wird also durch die Werte α und β bestimmt:



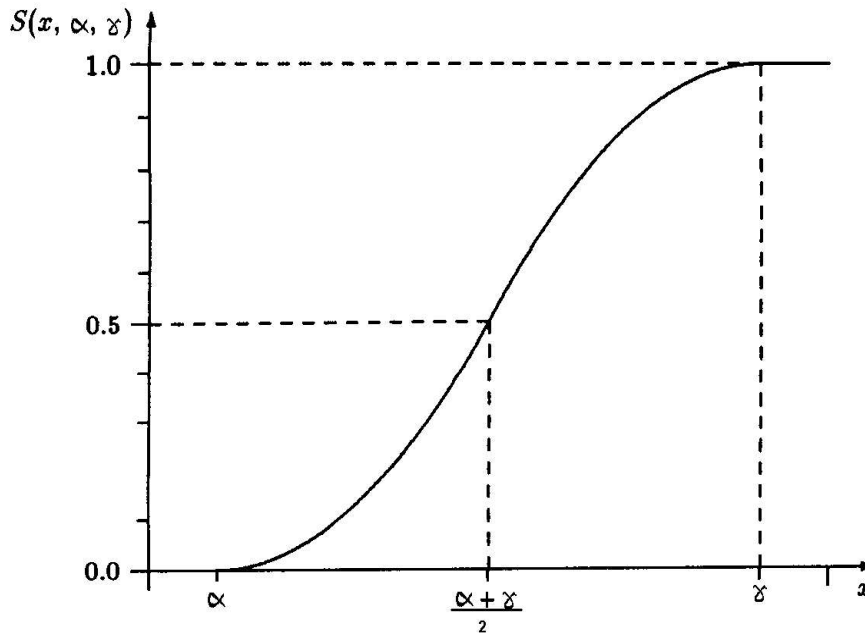
3.3.2 s/z-Funktion

Um die Kanten abzuglätten und einen runden Kurvenverlauf zu erhalten, kann man auch quadratische Funktionen wählen. Eine Möglichkeit ist die so genannte s/z-Funktion, zusammengesetzt aus zwei Halbkurven, der s-Funktion auf der linken und der z-Funktion auf der rechten Hälfte:

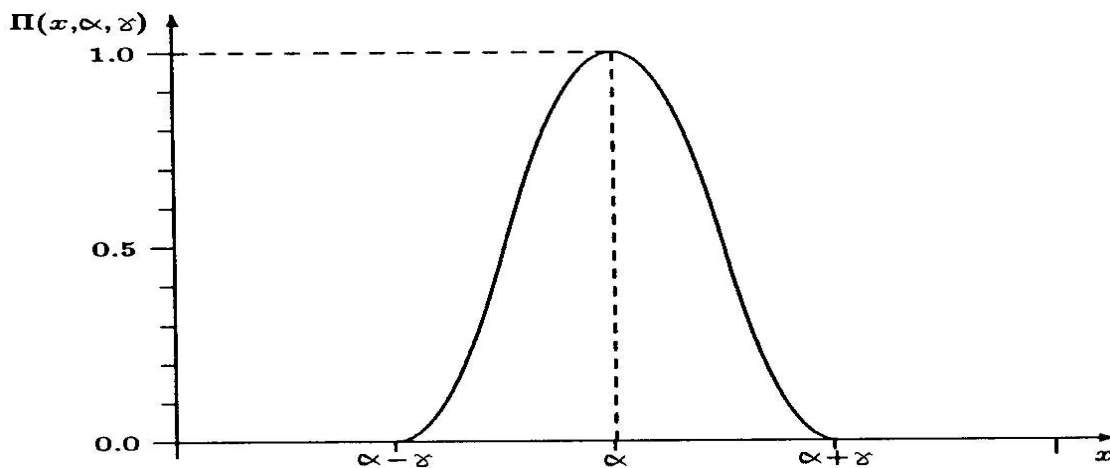
$$s(x, \alpha, \gamma) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{wenn } x \leq \alpha \\ 2 \cdot \left(\frac{x - \alpha}{\gamma - \alpha} \right)^2 & \text{wenn } \alpha < x \leq \frac{\alpha + \gamma}{2} \\ 1 - 2 \cdot \left(\frac{x - \gamma}{\gamma - \alpha} \right)^2 & \text{wenn } \frac{\alpha + \gamma}{2} < x \leq \gamma \\ 1 & \text{sonst} \end{cases}$$

Der Kurvenverlauf ist offensichtlich zwischen α und $(\alpha+\gamma)/2$, dem Wendepunkt, konvex; zwischen $(\alpha+\gamma)/2$ und γ konkav.

Das Schaubild für die s-Funktion sieht dementsprechend so aus:



Wenn wir jetzt als z-Funktion eine entsprechend gespiegelte Funktion rechts im Anschluss an die s-Funktion wählen, so sieht die gesamte s/z-Funktion folgendermaßen aus:



3.4 Horizontale Repräsentationen

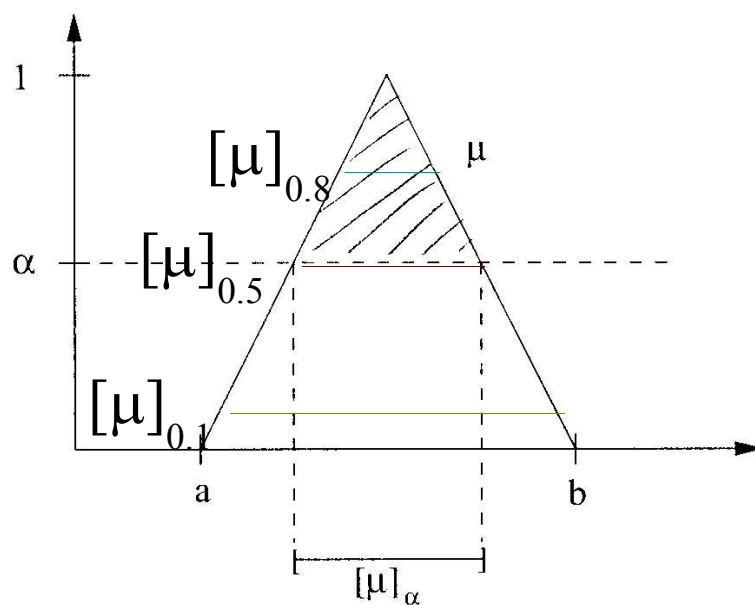
Bei der horizontalen Repräsentation einer Fuzzy-Menge werden eine Anzahl neuer Mengen gebildet, die so genannten α -Schnitte. Diese sind so definiert, dass jeder α -Schnitt genau die Elemente der ursprünglichen Menge enthält, deren Zugehörigkeitsgrad mindestens α ist. Formal ist er so definiert:

$$[\mu]_\alpha \stackrel{\text{def}}{=} \{\omega \in \Omega \mid \mu(\omega) \geq \alpha\}$$

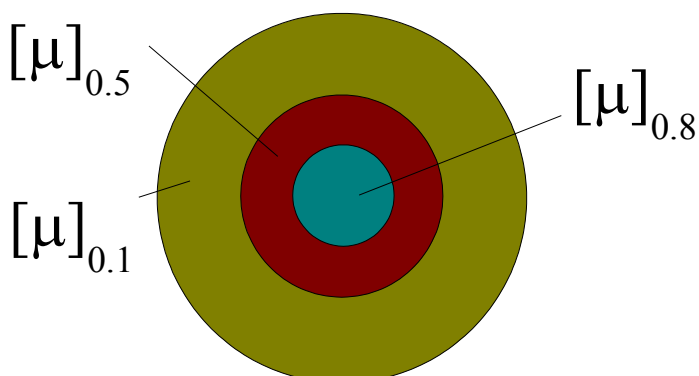
Ω steht hier für die Referenzmenge, ω für deren Elemente.

Die Präzision unserer Repräsentation wird also unmittelbar durch die Anzahl und Dichte der α -Schnitte bestimmt. Je mehr α -Schnitte wir machen, desto genauer wird das Ergebnis; letztlich kann hier (jedenfalls theoretisch) eine beliebig große Genauigkeit erzielt werden wenn wir nur entsprechend viele Schnitte machen.

Veranschaulicht sieht das folgendermaßen aus:



Wir sehen hier drei α -Schnitte: bei 0.1, 0.5 und 0.8. Der Schnitt bei 0.1 ist definitionsgemäß der größte, der bei 0.8 der kleinste. Dies bedeutet auch, dass die Schnitte bei 0.8 bzw. 0.5 vollständig in dem bei 0.1 enthalten sind. Wir können uns dieses Enthaltensein etwa wie folgt vorstellen:



3.5 Vergleich

Ist nun die horizontale oder die vertikale Darstellung für unsere Zwecke geeigneter?

Für die horizontale Darstellung spricht in jedem Fall, dass eine ganz einfache, natürlichsprachliche Eingabe möglich ist. Nämlich mit ganz unmathematischen Ausdrücken wie „voll relevant“, „mäßig relevant“ und „kaum relevant“ (bei drei Schnitten). Somit wird die Eingabe auch für den Nicht-Informatiker/ -Mathematiker zu einem Kinderspiel.

Für die vertikale Repräsentation wiederum spricht die Tatsache, dass die Darstellung einfach intuitiver verständlich, schneller erfassbar ist. Man kann die gesuchten Zugehörigkeitsgrade mit einem Blick ablesen. Nur wird es eben einem Historiker Schwierigkeiten bereiten (wenn er nicht gerade Hobby-Mathematiker ist), die passenden mathematischen Funktionen anzugeben.

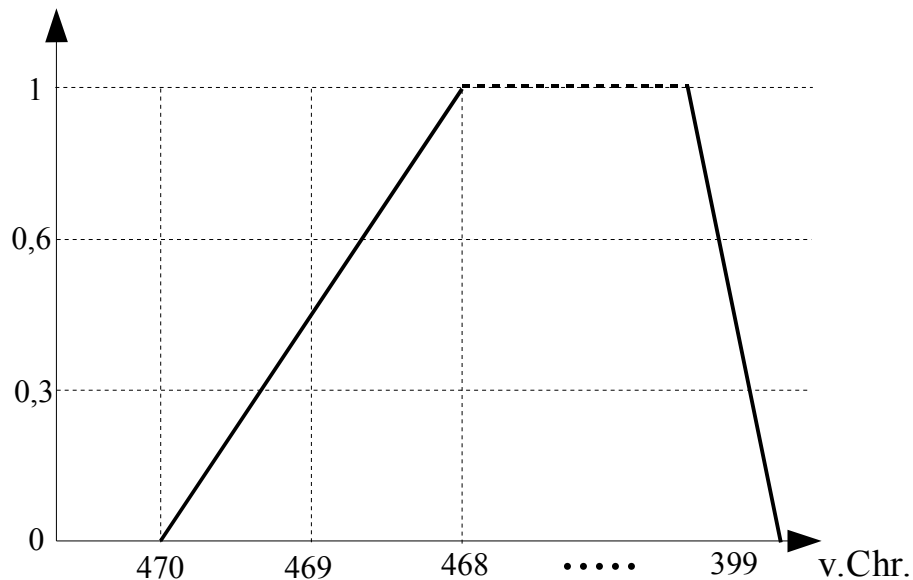
Aus diesen Gründen scheint eine gesunde Mischung optimal: Eingabe in simpler Weise horizontal, dann Überführung (durch den Computer?) in eine vertikale Darstellung.

4. Beispiele

Wir wollen jetzt einige konkrete Beispiele betrachten, die die verschiedenen in Abschnitt 2 erwähnten Probleme umfassen.

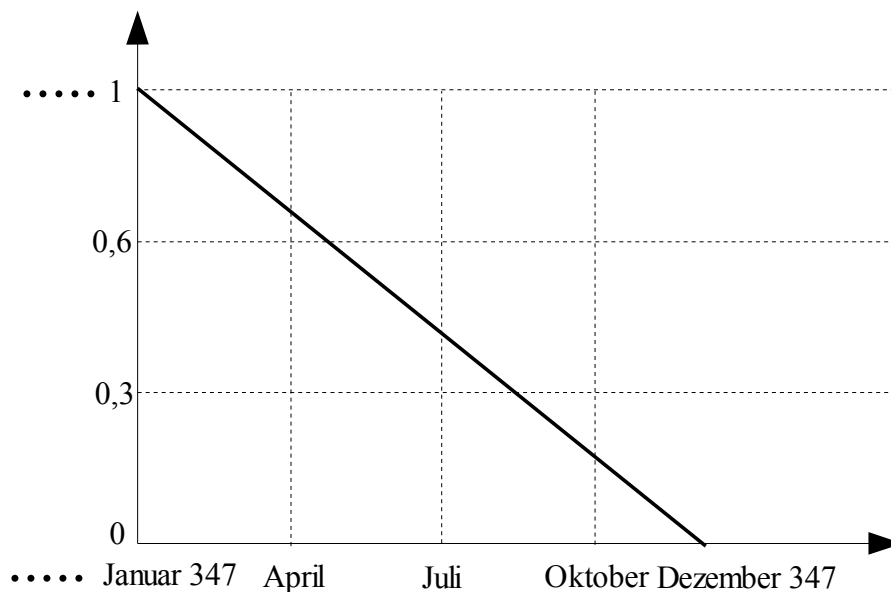
4.1 Beispiel zu Unsicherheit (vertikal)

Von dem Philosophen Sokrates ist überliefert, dass er von 470/469 – 399 v. Chr. lebte. Das genaue Geburtsjahr kann also nicht sicher angegeben werden. Die Wahrscheinlichkeit, dass er „bereits geboren war“, liegt also am 1.1.470 v. Chr. (und davor) bei 0, am 31.12.468 v. Chr. (und danach bis ins Jahr 399) bei 1. Für den dazwischen liegenden Zeitraum nehmen wir mangels genauerer Informationen eine gleichverteilte „Geburtswahrscheinlichkeit“ an und verbinden somit die beiden Punkte linear. Das zugehörige Schaubild sieht dann so aus:



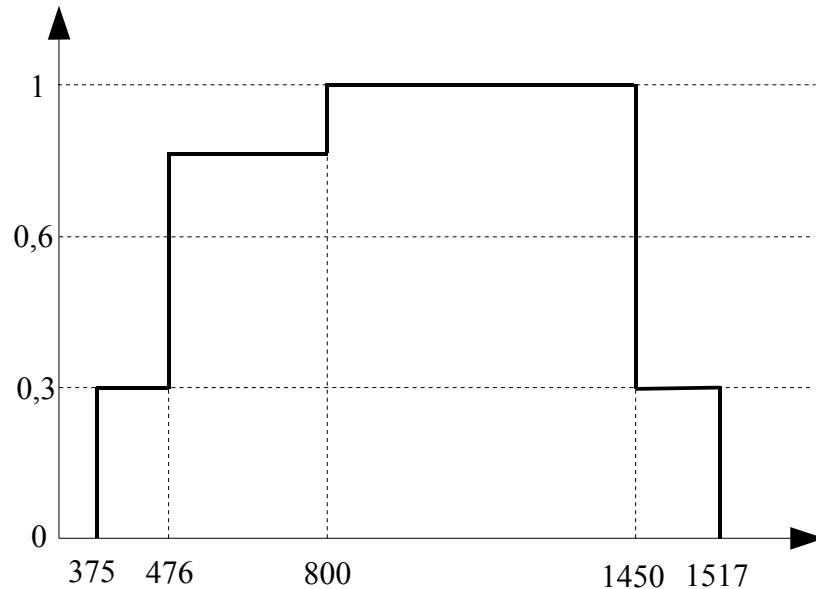
4.2 Unpräzise Angaben (vertikal)

Zur Illustration der sich durch unpräzise Angaben ergebenden Probleme habe ich hier Platon gewählt. Von ihm ist, wie von vielen anderen Berühmtheiten der Antike, nur das Todesjahr, nicht jedoch das genaue Todesdatum bekannt. Wenn nun unser Grund-Primitiv, mit dem wir die Geschichte untersuchen, eben „tagesgenau“ ist, ist die Jahresangabe zu unpräzise. Hier hilft wieder eine Darstellung mit Hilfe der Fuzzy Logic: am 1. Januar 347 v. Chr. lebte er definitiv noch; am 31.12. 347 v. Chr. war er definitiv tot (dass damals ein anderer Kalender verwendet wurde, sei an dieser Stelle einmal vernachlässigt). Den dazwischen liegenden Zeitraum überbrücken wir wieder linear.



4.3 Subjektive Angaben (vertikal)

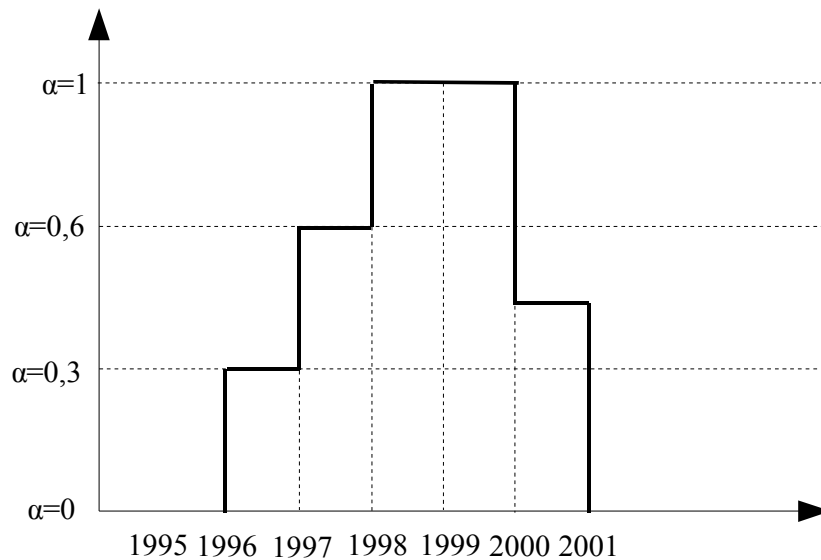
Ein schönes Beispiel für den Fall, dass eine Definition höchst subjektiv ist, ist der Begriff „Mittelalter“. Weder über dessen Beginn noch über dessen Ende herrscht auch nur annähernd Einigkeit.



Die erste Möglichkeit, vom Beginn des Mittelalters zu sprechen, ist um das Jahr 375 mit dem Beginn der Völkerwanderung. Allerdings dürfte dieses Datum kaum konsensfähig sein. Eine große Zahl von Historikern favorisiert etwa das Jahr 476, in dem das weströmische Reich auseinander brach. Spätestens mit dem Jahr 800, in dem Karl der Große gekrönt wurde, sind wir dann mitten im Mittelalter und geben den Zugehörigkeitsgrad 1 an. Auch das Ende des Mittelalters ist wieder eine subjektive Definitionssache. Möglich wäre z. B. ca. das Jahr 1450, wo der Buchdruck von Gutenberg erfunden wurde, oder das Jahr 1517 mit der Reformation durch Martin Luther.

4.4 Subjektive Angaben (horizontal)

Hier verwenden wir nochmals das bereits in Abschnitt 3.2 eingeführte Beispiel „späte Neunziger“ mit den gleichen Werten.

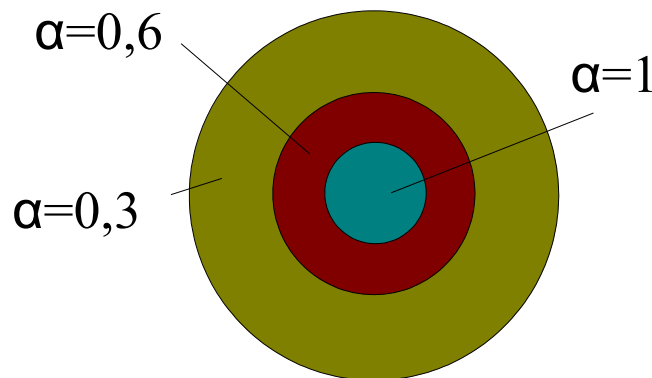


Wenn wir jetzt drei α -Schnitte bei 0,3, 0,6 und 1,0 durchführen, so ergeben sich folgende neue Untermengen:

$\alpha=0,3$: 1996-2001

$\alpha=0,6$: 1997-2000

$\alpha=1,0$: 1998-2000



Man sieht hier sehr schön die Probleme, die eine zu geringe Zahl von Schnitten mit sich bringen können: die Unterscheidung zwischen der Zugehörigkeit 0,5 von 2001 und 0,3 von 1996 geht verloren. Beide liegen noch im 0,3er-Schnitt und nicht mehr im 0,6er-Schnitt, obwohl sie verschieden große Zugehörigkeitsgrade besitzen.

5. Anwendungs-mockup

Nun stellt sich die Frage, wie diese Erkenntnisse konkret nutzbar gemacht werden könnten.

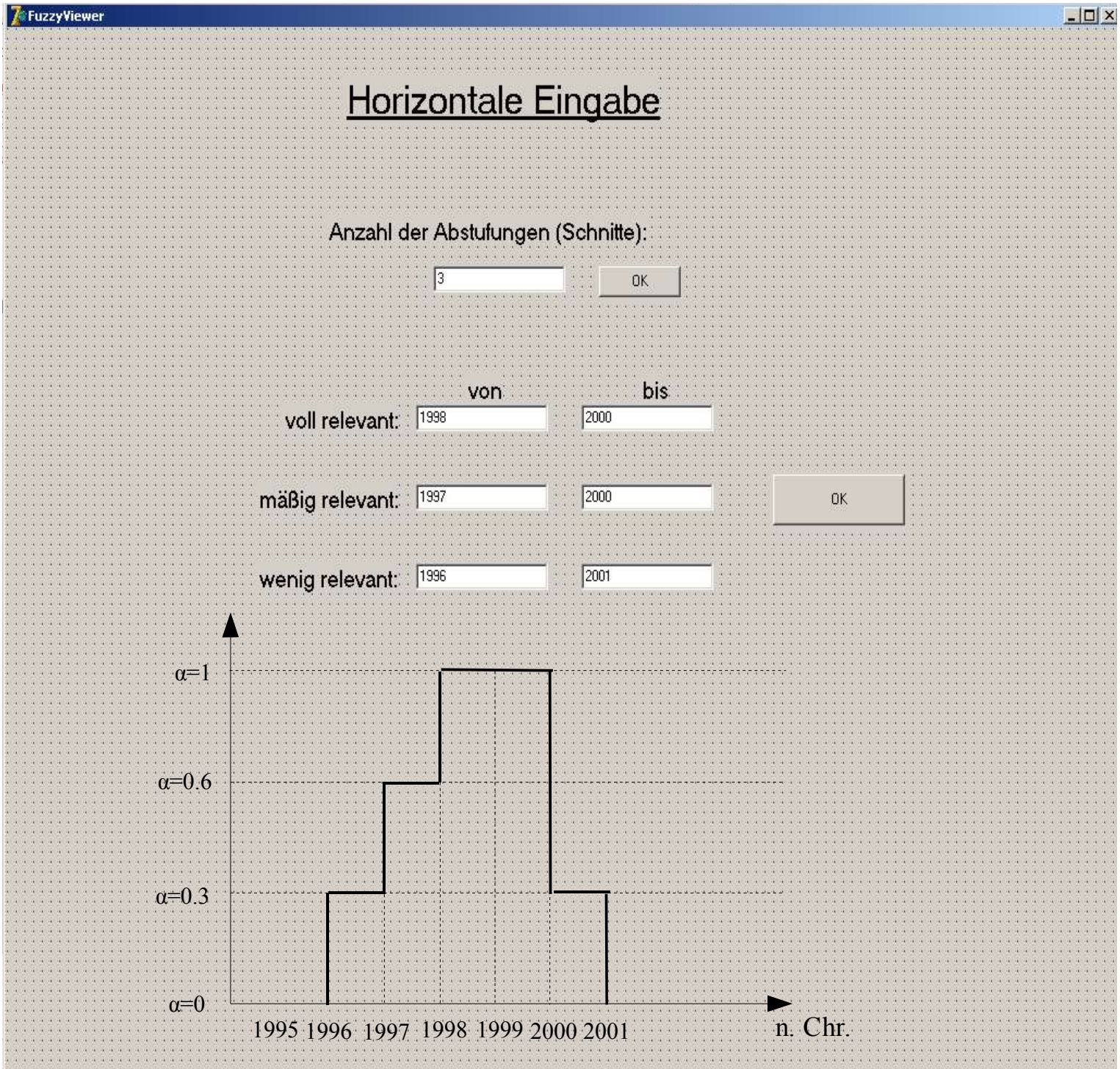
Interessant und sicherlich von großem Nutzen wäre ein Anwendungs-Programm, welches in der Lage wäre, nach entsprechenden horizontalen und/oder vertikalen Eingaben eine oder mehrere verschiedene passende, leicht verständliche (grafische) Ausgaben zu liefern.

Da ein solches Tool natürlich den Rahmen einer Seminararbeit sprengen würde, soll hier nur kurz angedeutet werden, wie eine GUI (grafische Benutzeroberfläche) für eine solche Anwendung

aussehen und welche (Grund-)Funktionalitäten dahinter stecken könnten.

Im folgenden Screenshot sehen wir zunächst die horizontale Variante. Im ersten Eingabefeld lässt sich angeben, wie viele Schnitte bzw. Abstufungen gewünscht werden.

Wenn dann, wie im Beispiel, deren drei gewählt werden, könnte in den Abstufungen „voll relevant“, „mäßig relevant“ und „wenig relevant“ auch ohne Kenntnisse der dahinter stehenden Logik die gewünschte Funktion erzeugt und dargestellt werden.



Vertikale Eingabe

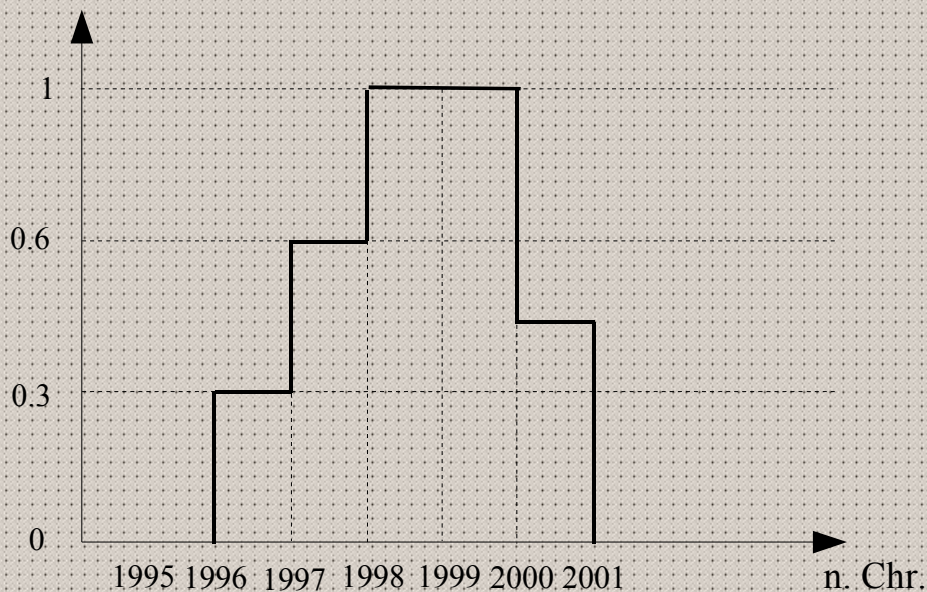
diskrete Eingabe: Zeitpunkt: Wahrscheinlichkeit: Verbindung: linear quadratisch

kontinuierliche Eingabe: von bis Wahrscheinlichkeit:

eingeegebene Werte:

1996: 0.3 - linear
 1997: 0.6 - linear
 1998: 1.0 - linear
 1999: 1.0 - linear
 2000: 0.5 - linear

1996-1997: 0.3
 1997-1998: 0.6
 1998-2000: 1.0
 2000-2001: 1.0



Die vertikale Eingabe unterscheidet zwischen diskreter und kontinuierlicher Eingabe.

Bei der diskreten Eingabe wird schlicht ein Zeitpunkt nach dem anderen mit den zugehörigen Wahrscheinlichkeiten eingegeben und anschließend nur noch gewählt, ob die Verbindung zum nächsten Punkt linear oder quadratisch erfolgen soll. Die bereits eingegebenen Werte erscheinen in der Box ganz rechts. (An dieser Stelle besteht natürlich noch viel Raum für Erweiterungen, man könnte dem ambitionierten Anwender auch die Möglichkeit geben, eigene mathematische

Funktionen anzugeben.)

Bei der kontinuierlichen Eingabe wird dann ein Zeitraum nach dem anderen mit einem Zugehörigkeitswert versehen.

5. Fazit

Aus unterschiedlichen Gründen liegen geschichtliche Daten und zeitliche Angaben meist mehr oder weniger unpräzise vor.

Soll dies im Rahmen eines Fuzzy-Modells berücksichtigt werden, stehen zwei grundlegende Repräsentationen zur Verfügung: Die horizontale und die vertikale Repräsentation.

Erstere ist für die Eingabe durch einen Laien, letztere für eine anschauliche Darstellung geeigneter.

Weiter gearbeitet werden könnte unter anderem an Ideen, Design und Implementierungen einschlägiger Anwendungen.

Literatur:

- Nagypál, Gabor und Motik, Boris: A Fuzzy Model for Representing Uncertain, Subjective and Vague Temporal Knowledge in Ontologies, 2003, ODBASE
- Mayer, Andreas et al.: Fuzzy Logic – Einführung und Leitfaden zur praktischen Anwendung. © 1993 Addison-Wesley GmbH Deutschland
- Renè Witte: Architektur von Fuzzy-Informationssystemen zur Repräsentation und Verarbeitung unscharfer Daten, 2002 (Uni-Karlsruhe Doktorarbeit)