

VSB – TECHNICAL UNIVERSITY OF OSTRAVA



DEPARTMENT OF SYSTEMS ENGINEERING

Assessing the Risk Profile of a Retail Investor with a Focus on Complementing
Retirement Investments – a Fuzzy Logic Approach

Posuzování Rizikového Profilu Retailového Investora se Zaměřením na
Doplnění Investic na Penzi – Přístup s Využitím Fuzzy Logiky

Student: Bc. Vit Chrubasik, BSc (Hons)
Supervisor: doc. Dr. Ing. Miroslav Hudec

Ostrava 2022

Abstract

In order to provide the best services to their customers for undertaking investments for retirement as part of the third pillar of the retirement reform from the World Bank, financial institutions assess private investors' risk profiles. There are many approaches to assessing a risk profile. Since risk profile definitions are vague and assessment deals with many imprecisions, uncertainty, and missing information, we proposed an approach based on fuzzy logic which proved to have capabilities for dealing with those phenomena. We described a framework where we can translate risk profile components into fuzzy sets and assess risk profiles using fuzzy inference. We argued for flexibility, understandability (to a non-technical audience), and transparency. We proposed that future work focuses on parts of our proposed framework, as we have given those very broadly to provide proof of concept.

Keywords: Fuzzy Logic, Fuzzy Inference, Risk Assessment, Investing, Retirement

Acknowledgements

I want to thank my supervisor, doc. Dr. Ing. Miroslav Hudec, for being of great support in providing feedback and pointing me in the right direction. I also want to thank doc. Dr. Ing. Miroslav Hudec for introducing me to the captivating world of fuzzy logic.

I want to thank all of the conscientious people at VSB – Technical University of Ostrava for not hesitating to go above and beyond their standard duty when providing my peers and me with academic and non-academic support.

I have to give great thanks to my supportive family and friends, who always wanted the best for me. I am forever grateful for having them in my life and will always check myself not to take them for granted.

Last but not least I also want to go beyond just listing the bibliography and give my thanks explicitly to all scholars referenced in this thesis. Namely Professor Lotfi A. Zadeh and his remarkable seminal work. Through this work, Professor Zadeh added the kindling to ignite the fire of passion for fuzzy logic in me after doc. Hudec gave the initial spark.

Contents

Abstract	i
1 Introduction	1
2 Literature Review	2
2.1 Fuzzy Set Theory	2
2.2 Operations on Fuzzy Sets	5
2.2.5 Fuzzy Relations	8
2.3 Fuzzy Logic	9
2.3.1 Fuzzy Propositions	9
2.3.2 Fuzzy Logical Operators	9
2.3.6 Fuzzy Predicates	10
2.3.7 Fuzzy Quantifiers	10
2.3.10 Fuzzy Predicate Modifiers	11
2.3.12 Qualification of Propositions	11
2.4 Linguistic Variables	12
2.5 Linguistic Summaries	13
2.6 Fuzzy Inference	16
2.6.1 Generalized Modus Ponens	16
2.6.2 Mamdani Inference System	17
2.7 Defuzzification	19
2.7.2 Defuzzification based on extreme value	19
3 Analysis of the state of the art of risk profile assessment	20
4 Methodology	22
4.1 Risk Profile Assessment	22
4.2 Building the Knowledge Base and the Inference System	22

5 Application of the Proposed Method	25
5.1 Constructing fuzzy sets and computing memberships for elements	27
5.2 Model Example	28
6 Discussion	31
7 Conclusion	35
Bibliography	39
List of Abbreviations	40
List of Annexes	41
Annexes	42
A Ageing Societies – Alternatives to employment for older citizens: Ex-	
ploring challenges and opportunities from the point of view of the cur-	
rent young workforce as future retirees	43
B Querying and Data Summarization with Fuzzy Logic: Helping Pensions	
Advisors to Search for Clients based on Risk Profile and other Fuzzy	
Data – Research Project Proposal	49
C Application Implementation	57
D Application Notebook	63

List of Figures

2.1	An example of a convex fuzzy set.	4
2.2	An example of a non-convex fuzzy set.	5
2.3	Fuzzy complement.	6
2.4	Fuzzy max s -norm.	7
2.5	Fuzzy min t -norm.	8
2.6	Linguistic Variable $L = \text{"Water Teperature"}$ (Chrubasik, 2022b)	12
2.7	"Is young" predicate membership function.	14
2.8	"Most" validity function.	15
2.9	Mamdani inference example with one crisp antecedent (Hudec, 2016; Ross, 2005).	18
2.10	Mamdani inference example with one fuzzy antecedent (Hudec, 2016; Ross, 2005).	18
2.11	Mamdani inference system diagram (Hudec, 2016; Ross, 2005).	18
4.1	General decomposition mind map diagram.	23
4.2	Diagram of Equation 4.1	24
5.1	Decomposition created from the literature.	26
5.2	Types of used fuzzy numbers.	28
6.1	The magic triangle of investing (Musílek, 1999).	34

Chapter 1

Introduction

Retirement is a ubiquitous topic in the developed world. Even though having enough funds is not as necessary as alternatives to employment arrive for ageing societies (Chrubasik, 2022a), it is highly unlikely it would ever be recommended to overlook the third pillar of retirement, i.e., undertaking private investment projects (Holzmann et al., 2008). However, investing can be an overburdening endeavor for an inexperienced individual. Therefore, individuals usually invest through an intermediary. Intermediaries assess the investor based on a plethora of aspects in order to recommend and broker relevant products. A collection of those aspects is commonly known as a *risk profile*.

Many factors play a role in risk profile assessment and are mostly not easily quantifiable. The data that come into the assessment process are of a large variety. Those data relate to aspects such as credit score, accreditation, or financial literacy. The models of assessment are not set in stone; they usually have to be altered, if not replaced altogether, when new information comes to light.

This thesis focused on an approach incorporating selected techniques from fuzzy logic, which builds on fuzzy set theory (Lotfi A Zadeh, 1996). We outlined the theoretical foundation and some selected related applications in Chapter 2. We analyzed the state-of-the-art of risk profile assessment in Chapter 3. We explained our general approach in Chapter 4 and, based on that, created and described an example model in Chapter 5. We focused on the approach itself; we assumed our example model in Chapter 5 is incomplete. We discussed our findings and gave pointers for possible future work in Chapter 6. We concluded the thesis in Chapter 7.

Chapter 2

Literature Review

Our literature review focused on the approach we wanted to investigate – fuzzy logic. In the following sections, we review the necessary theory from the contemporary literature about fuzzy logic, which build on top of fuzzy set theory. If not stated otherwise, definitions are paraphrased from Hudec (2016).

2.1 Fuzzy Set Theory

Fuzzy set theory (Lotfi A Zadeh, 1996, originally 1965) is an extension of the classical set theory (Zermelo, 1908; Fraenkel, 1925).

Definition 2.1.1 (Fuzzy Set). Given a universe of discourse X , we say a fuzzy set \tilde{A} is set of ordered pairs: (i) an element $x \in X$, and (ii) a membership function $\mu(x)$, such that $\mu : X \rightarrow [0, 1]$. Formally:

$$\tilde{A} := \{(x \in X, \mu(x)) \mid \mu : X \rightarrow [0, 1]\}. \quad (2.1)$$

When talking about membership, we say that for fuzzy set \tilde{A} an element a belongs to \tilde{A} with a degree of membership $\mu(a)$.

Remark 2.1.2. It holds that

$$\begin{aligned} a \in A &\leftrightarrow \mu(a) = 1 \\ a \notin A &\leftrightarrow \mu(a) = 0. \end{aligned}$$

A classical (crisp) set would therefore be a special case of a fuzzy set with $\mu : X \rightarrow \{0, 1\}$.

Definition 2.1.3 (Notation). For finite and discrete X , we denote a fuzzy set \tilde{A} :

$$\tilde{A} = \left\{ \frac{\mu_{\tilde{A}}(x_1)}{x_1} + \frac{\mu_{\tilde{A}}(x_2)}{x_2} + \dots + \frac{\mu_{\tilde{A}}(x_n)}{x_n} \right\} = \left\{ \sum_{i=1}^n \frac{\mu_{\tilde{A}}(x_i)}{x_i} \right\}. \quad (2.2)$$

The horizontal fraction line serves as a delimiter. $+$ is a union operator. For infinite and continuous X :

$$\tilde{A} = \left\{ \int_a^b \frac{\mu_{\tilde{A}}(x)}{x} \right\}. \quad (2.3)$$

Definition 2.1.4 (Support of a fuzzy set). Support of a fuzzy set \tilde{A} is defined as a set of all elements of \tilde{A} with a non-zero degree of membership. Formally:

$$\text{supp}(\tilde{A}) := \{x \in X \mid \mu(x) > 0\}. \quad (2.4)$$

An equivalent definition can be made using strong α -cut (see Definition [2.1.6](#))

$$\text{supp}(\tilde{A}) := \tilde{A}^{0+}. \quad (2.5)$$

Definition 2.1.5 (α -cut). α -cut of a fuzzy set \tilde{A} is defined as:

$$\tilde{A}^\alpha := \{x \in X \mid \mu(x) \geq \alpha\}, \quad (2.6)$$

Example 2.1.5.1. $\tilde{A}^0 = \tilde{A}$

Definition 2.1.6 (Strong α -cut). Strong α -cut of a fuzzy set \tilde{A} is defined as:

$$\tilde{A}^{\alpha+} := \{x \in X \mid \mu(x) > \alpha\}, \quad (2.7)$$

Definition 2.1.7 (Height of a fuzzy set). The height of a fuzzy set \tilde{A} is the highest membership degree in fuzzy set \tilde{A} (supremum of the membership degree set). Formally:

$$h(\tilde{A}) := \sup_{x \in X} \mu_{\tilde{A}}(x) \quad (2.8)$$

Definition 2.1.8 (Cross-over points). Crossover points of a fuzzy set \tilde{A} are a set of elements of \tilde{A} such that their membership degrees are half that of $h(\tilde{A})$.

$$x_c := \{x \in X \mid \mu_{\tilde{A}}(x) = \frac{h(\tilde{A})}{2}\} \quad (2.9)$$

Definition 2.1.9 (Core/kernel of a fuzzy set). Core of a fuzzy set \tilde{A} is a (crisp) set of elements where $\mu_X(x) = 1$, formally:

$$\text{Core}(\tilde{A}) = \{x \in X \mid \mu_X(x) = 1\} \quad (2.10)$$

Definition 2.1.10 (Normalized fuzzy set). We call a fuzzy set \tilde{A} a normalized fuzzy set if and only if there exists an element $x \in X$ such that $\mu_{\tilde{A}}(x) = 1$, therefore $h(A) = 1$. Formally:

$$\begin{aligned}\text{Normalized}(\tilde{A}) &\leftrightarrow \exists x \in X \mid \mu_{\tilde{A}}(x) = 1 \\ \text{Normalized}(\tilde{A}) &\rightarrow h(A) = 1\end{aligned}$$

Definition 2.1.11 (Convexity of a fuzzy set). We say fuzzy set \tilde{A} is convex if and only if the following holds:

$$\mu_{\tilde{A}}(\lambda x + (1 - \lambda)y) \geq \min(\mu_{\tilde{A}}(x), \mu_{\tilde{A}}(y)) \quad \forall x, y \in X, \forall \lambda \in [0, 1] \quad (2.11)$$

Examples of a convex and a non-convex fuzzy set given in Figures [2.1](#), [2.2](#) respectively.

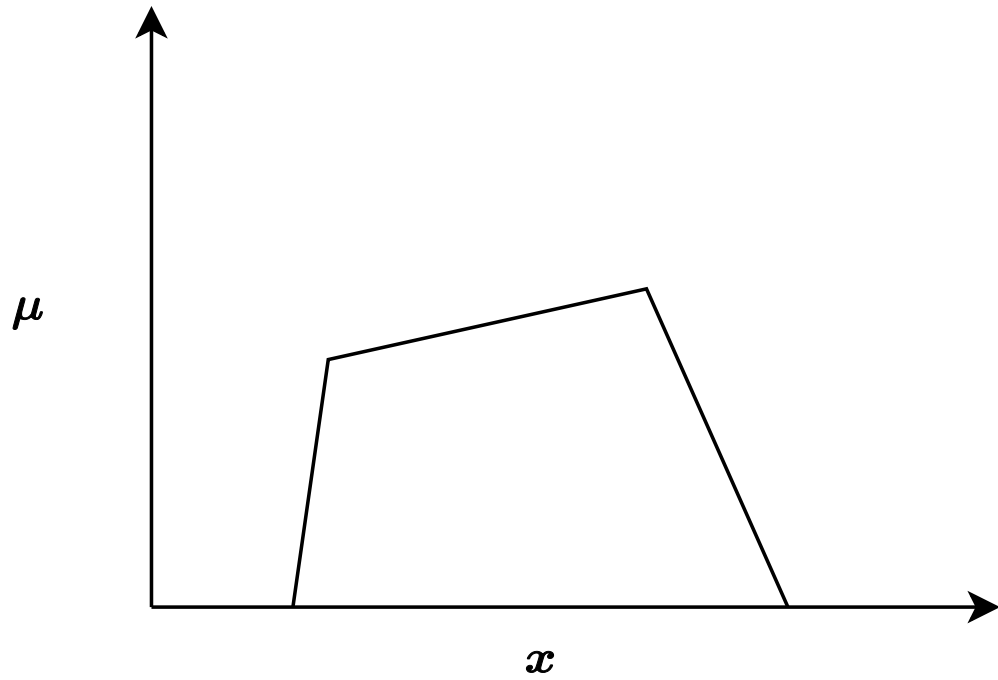


Figure 2.1: An example of a convex fuzzy set.

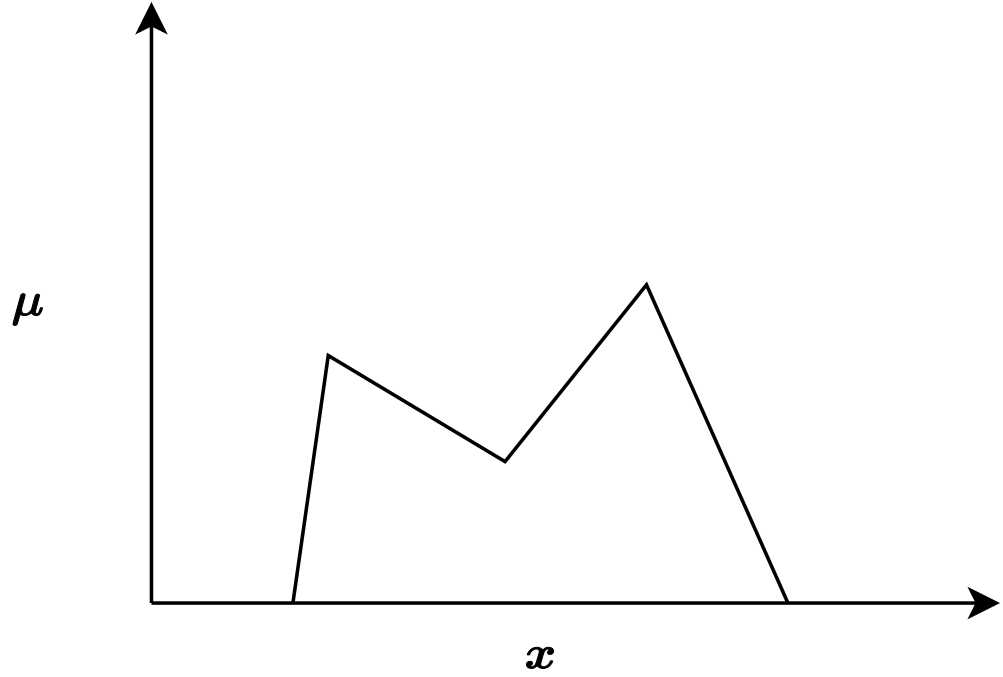


Figure 2.2: An example of a non-convex fuzzy set.

Definition 2.1.12 (Fuzzy Number). We say fuzzy set \tilde{M} is a fuzzy number if and only if it is normalized (2.1.10), convex (2.1.11), and $X = \mathbb{R}$. Fuzzy numbers can, therefore, be understood as an extension of real numbers (Dijkman et al., 1983).

Examples of fuzzy numbers can be viewed in Chapter 5, Figure 5.2.

2.2 Operations on Fuzzy Sets

Operations on fuzzy sets affect the membership function.

Definition 2.2.1 (Equality). We say that fuzzy set \tilde{A} is equal to fuzzy set \tilde{B} if and only if

$$\mu_{\tilde{A}}(x) = \mu_{\tilde{B}}(x), \forall x \in X. \quad (2.12)$$

Definition 2.2.2 (Fuzzy Complement). Fuzzy complement of a fuzzy set \tilde{A} is a unary operation defined as (see Figure 2.3 for a diagram):

$$\mu_{\tilde{A}}(x) := 1 - \mu_{\tilde{A}}(x), \forall x \in X \quad (2.13)$$

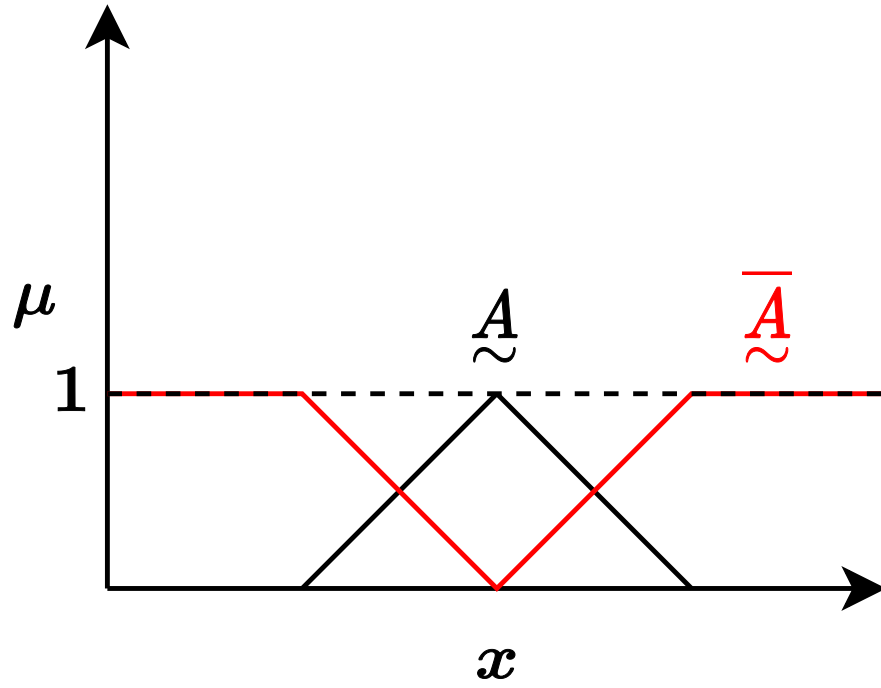


Figure 2.3: Fuzzy complement .

Definition 2.2.3 (Fuzzy Union).

$$\mu_{\tilde{A} \cup \tilde{B}}(x) = s(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)), \forall x \in X, \quad (2.14)$$

where s is an interpretation of the union operation. The most common s for the union operation is the \max function (Figure 2.4), i.e.:

$$\mu_{\tilde{A} \cup \tilde{B}}(x) = \max(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)), \forall x \in X, \quad (2.15)$$

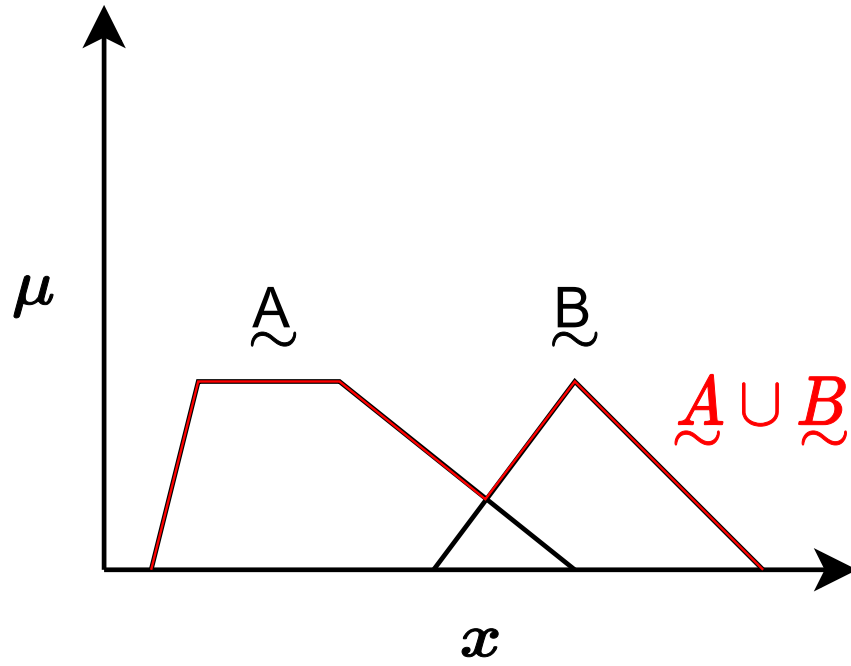


Figure 2.4: Fuzzy max s -norm .

Definition 2.2.4 (Fuzzy intersection).

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = t(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)), \forall x \in X, \quad (2.16)$$

where t is an interpretation of the intersection operation. The most common t for the intersection operation is the min function (Figure 2.5), i.e.:

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = \min(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)), \forall x \in X, \quad (2.17)$$

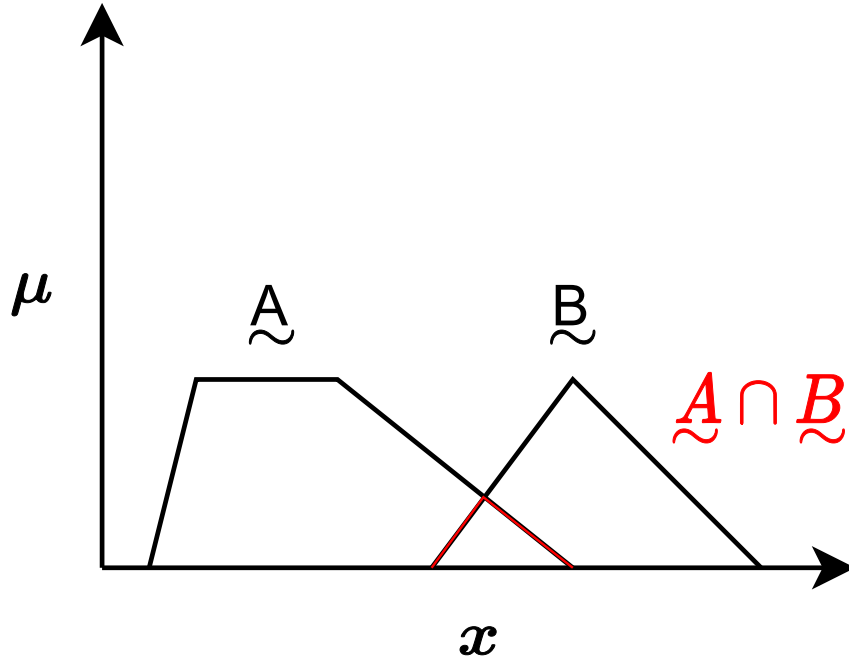


Figure 2.5: Fuzzy min t -norm .

2.2.5 Fuzzy Relations

Given a set A and a set B (both could be either classical or crisp), we define a relation $R(A, B)$ as a subset of a Cartesian product $A \times B$. Formally:

$$R := \{(x, y) \mid \forall (x \in A, y \in B) \in A \times B\} \quad (2.18)$$

Definition 2.2.6 (Fuzzy Relation). We define a membership function $\mu : R \rightarrow [0, 1]$. Fuzzy relation \tilde{R} is then defined as:

$$\tilde{R} := \{((x, y), \mu_{\tilde{R}}(x, y)) \mid \text{supp}(A \times B)\} \quad (2.19)$$

Even though it is possible to define a (fuzzy) relation between more than two sets, in this thesis, we only work with relations defined on two sets, i.e., *binary relations*.

Given a crisp relation $R_1(A_1, A_2)$, and a crisp relation $R_2(A_2, A_3)$, composition is defined as:

$$R_1 \circ R_2 = \{(x_1, x_3) \in R_1 \times R_2 \mid \exists x_2 \in R_2, (x_1, x_2) \in R_1 \wedge (x_2, x_3) \in R_2\} \quad (2.20)$$

Definition 2.2.7 (Fuzzy Composition). Given a fuzzy relation $\tilde{R}(A_1, A_2)$, and a fuzzy relation $\tilde{R}(A_2, A_3)$ we define fuzzy composition as a membership function $\mu_{\tilde{R}_1 \circ \tilde{R}_2} : [0, 1]^2 \rightarrow [0, 1]$

We can define $\mu_{\tilde{R}_1 \circ \tilde{R}_2}$ arbitrarily. However, the most commonly used definition is the *max-min composition rule*.

Definition 2.2.8 (Max-min Fuzzy Composition). Given sets A_1, A_2, A_3 , and fuzzy relations

$\tilde{R}_1(A_1, A_2), \tilde{R}_2(A_2, A_3)$, we define Max-min fuzzy composition membership function as:

$$\mu_{\tilde{R}_1 \circ \tilde{R}_2} := \sup_{x_2 \in A_2} \min(\mu_{\tilde{R}_1}(x_1, x_2), \mu_{\tilde{R}_2}(x_2, x_3)), \forall (x_1, x_3) \in A_1 \times A_3 \quad (2.21)$$

2.3 Fuzzy Logic

Fuzzy logic build on fuzzy set theory to extend classical and many-valued logic.

2.3.1 Fuzzy Propositions

Unlike in two-valued logic, where propositions are either *true*, or *false*, a proposition p in fuzzy logic can have a truth value in the real interval $[0, 1]$ (Lotfi A Zadeh, 1988).

2.3.2 Fuzzy Logical Operators

In fuzzy logic, logical operators on propositions are isomorphic to fuzzy set operations.

Given a fuzzy proposition \tilde{p} , we define:

Definition 2.3.3 (Logical NOT).

$$\neg \tilde{p} := 1 - \tilde{p} \quad (2.22)$$

Given a fuzzy proposition \tilde{p} and a fuzzy proposition \tilde{q} , we define:

Definition 2.3.4 (Logical OR).

$$\tilde{p} \vee \tilde{q} := s(\tilde{p}, \tilde{q}), \quad (2.23)$$

where s is an interpreting s -norm. For maximum s -norm:

$$\underline{p} \vee \underline{q} := \max(\underline{p}, \underline{q}), \quad (2.24)$$

Definition 2.3.5 (Logical AND).

$$\underline{p} \wedge \underline{q} := t(\underline{p}, \underline{q}), \quad (2.25)$$

where t is an interpreting t -norm. For minimum t -norm:

$$\underline{p} \wedge \underline{q} := \min(\underline{p}, \underline{q}), \quad (2.26)$$

2.3.6 Fuzzy Predicates

Predicates in fuzzy logic can therefore be created about vague phenomena, such as:

$$\text{Tall}(x \in X)$$

where X is the universe of people.

2.3.7 Fuzzy Quantifiers

We can construct fuzzy quantifiers about elements from the universe of discourse in addition to traditional universal and existential quantifiers from first-order logic.

We can create extended existential quantifiers – *absolute quantifiers*.

Example 2.3.7.1. We can relax the universal quantifier for all to for *almost* all. There exist n occurrences can be relaxed (see Linguistic Hedges in [2.3.11](#)) to there exist *about* n occurrences.

These quantifiers are then defined as fuzzy sets

Definition 2.3.8 (Proportional Quantifiers).

$$\forall_{x \in X} := \{(p, \mu_{\forall}(p)) \mid \mu_{\forall} : [0, 1] \rightarrow [0, 1]\} \quad (2.27)$$

Definition 2.3.9 (Absolute Quantifiers).

$$\exists_{x \in X} := \begin{cases} (n, \mu_{\exists}(n)) \mid \mu_{\exists} : \mathbb{N} \rightarrow [0, 1] \text{ for discrete quantities } n \\ (x, \mu_{\exists}(x)) \mid \mu_{\exists} : \mathbb{R} \rightarrow [0, 1] \text{ for continuous quantities } x \end{cases} \quad (2.28)$$

Examples and how to come up with proportion p and absolute quantities n, x are discussed in [2.5](#).

We can then make statements like:

$$(\forall \mid \exists)x \in X, \text{ Tall}(x)$$

where X is the universe of people.

\forall can be, e.g., almost all, about half, quarter, thirty percent... \exists can then be, e.g., about three billion, not more than three billion, several...

2.3.10 Fuzzy Predicate Modifiers

We can make modifiers more strict or relaxed in fuzzy logic. We can either use (linguistic) hedges (Definition [2.3.11](#)) or construct linguistic variables (Section [2.4](#)).

Definition 2.3.11 (Linguistic Hedges). Linguistic hedge is a function $H : [0, 1] \rightarrow [0, 1]$, which takes a membership function of a fuzzy set A as an input and transforms it. We say the hedge H transforms meaning of A .

Example 2.3.11.1. Given the universe of discourse of people X and a fuzzy predicate $Y(x \in X)$ meaning x is *young* defined as:

$$Y(x) = \int_0^{\infty} \frac{1}{1 + e^{x-35}}, \quad (2.29)$$

we can define linguistic hedge *very* of a fuzzy predicate Y as:

$$\text{very}(Y) := (\mu_Y)^2. \quad (2.30)$$

2.3.12 Qualification of Propositions

Unlike classical logic (Hughes et al., [1996](#)), fuzzy logic can qualify truth, probability, and possibility (Lotfi A Zadeh, [1988](#)).

Example 2.3.12.1. • **Truth:** it is *not quite true* that violets are blue.

- **Probability:** it is *likely* that it will rain.
- **Possibility:** there is *a slight possibility* that it will rain.

2.4 Linguistic Variables

Linguistic variables take their values from words or sentences from the natural language.

Example 2.4.0.1. *Temperature* is an instance of a linguistic variable. Instances of its values could be *hot*, *very hot*, *lukewarm*, *freezing*, *slightly cold*...

Definition 2.4.1 (Linguistic Variable). More rigorously, we define a linguistic variable LV as a set $LV = \{L, T(L), X, G, H\}$, where

- L is the name of the variable
- $T(L)$ is the set of all linguistic labels related to L (from Example 2.4.0.1, *hot* would correspond to label "hot" etc.)
- X is the universe of discourse
- G is the syntactic rule to generate $T(L)$ values
- H is the semantic rule relating each label $T(L)$ to its meaning $H(L)$ (Lotfi Asker Zadeh, 1975; Hudec, 2016; Ross, 2005).

Example 2.4.1.1. Consider linguistic variable $L = \text{"Water Temperature"}$ with

$T = \{\text{"Freezing", "Very Cold", "Cold", "Slightly Cold", "Lukewarm", "Warm", "Hot", "Very Hot"}\}$

This example is illustrated in a diagram in figure 2.6.

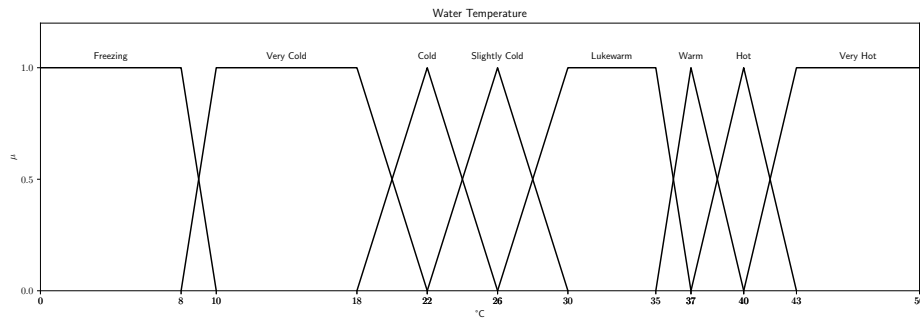


Figure 2.6: Linguistic Variable $L = \text{"Water Temperature"}$ (Chrubasik, 2022b)

Linguistic variables are argued to be the tool to bridge real-world and computer data representation (Lotfi A Zadeh, 1983) – they are the building block of a bridge connecting the world of human and computer reasoning.

2.5 Linguistic Summaries

Linguistic summaries summarise data in sentences from the natural language (Yager, 1982).

Definition 2.5.1 (Linguistic Summary). We define linguistic summary as a fuzzy quantifier \tilde{Q}_x of a fuzzy predicate $\tilde{P}(x)$ ($x \in X$), both constructed from natural language. Formally:

$$\tilde{Q}_x(\tilde{P}(x \in X)) \quad (2.31)$$

validity v of a linguistic summary is then a function

$$v(\tilde{Q}_x) \quad (2.32)$$

Summaries serve as an alternative to numeric quantities such as mean or variance (Yager, 1982).

$$\tilde{v}(v) := \int_0^{0.4} 0 + \int_{0.4}^{0.9} 2x + \int_{0.9}^1 1 \quad (2.33)$$

Given a data-set $D = \{d_1, d_2, \dots, d_n\}$, $n \in \mathbb{N}$, we can create linguistic summaries in the form:

$$Q \text{ elements of } D \text{ have property } P. \quad (2.34)$$

This is a fuzzy statement and has a truth value, also called validity $v \in [0, 1]$.

$$\begin{aligned} P : D &\rightarrow \tilde{D} \\ Q(D) &= \begin{cases} \frac{1}{n} \sum_{i=1}^n \mu_{\tilde{D}}(d_i) & \text{if } Q \text{ proportional} \\ \sum_{i=1}^n \mu_{\tilde{D}}(d_i) & \text{if } Q \text{ absolute} \end{cases} \\ v : Q &\rightarrow [0, 1] \end{aligned}$$

Example 2.5.1.1 (Taken from Chrubasik, 2022b). Let D be the dataset of people's ages (duplicates allowed, people are distinct).

$$D = \{15, 13, 14, 33, 35, 20, 19, 8, 5\} \quad (2.35)$$

Most people are young. (2.36)

$P : d \in D$ is young

$Q : \text{most}$

Given fuzzy predicate "is young" \tilde{P} (similar to the one in Yager, 1982):

$$\tilde{P}(d \in D) = \int_0^{40} \frac{1}{1 + e^{x-35}} + \int_{40}^{\infty} 0, \quad (2.37)$$

yielding membership function of shape in Figure 2.7,

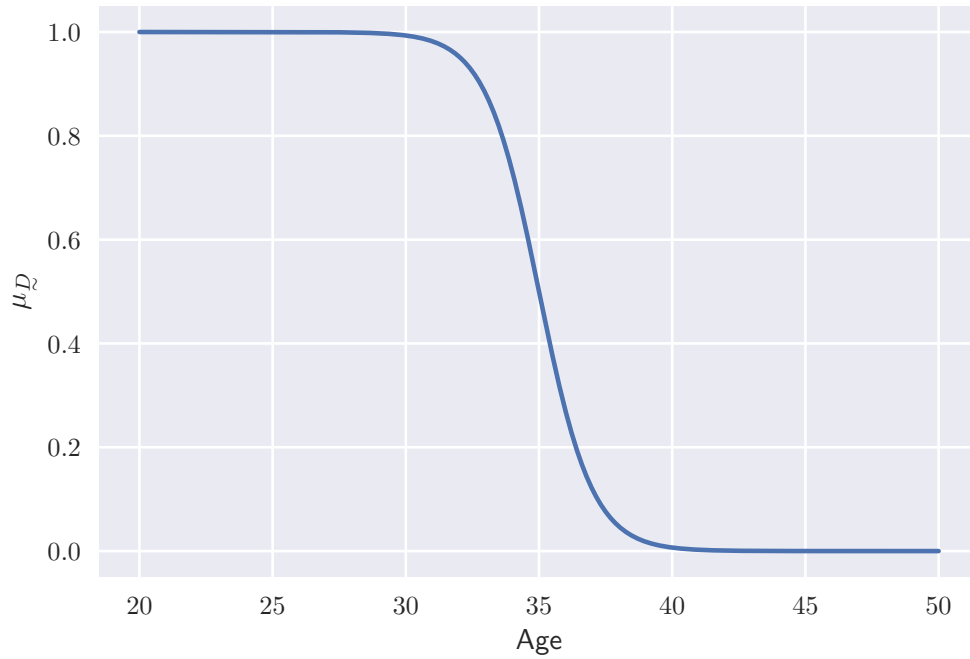


Figure 2.7: "Is young" predicate membership function .

computation of the fuzzy predicate for D yields fuzzy set \tilde{D}

$$P(d \in D) = \tilde{D} \approx \left\{ \frac{1}{15} + \frac{1}{13} + \frac{1}{14} + \frac{0.88}{33} + \frac{0.5}{35} + \frac{1}{20} + \frac{1}{19} + \frac{1}{8} + \frac{1}{5} \right\}$$

We compute the value of fuzzy quantifier \tilde{Q} :

$$\begin{aligned} \text{most of (proportional)} &= \tilde{Q}(D) \\ &= \frac{1}{n} \sum_{i=1}^n \mu_{(d_i)} \approx 0.93. \end{aligned}$$

Given the validity function (for shape see Figure 2.8)

$$v(\tilde{Q}) = \int_0^1 \frac{1}{1 + e^{-30x + 30 \cdot 0.7}}, \quad (2.38)$$

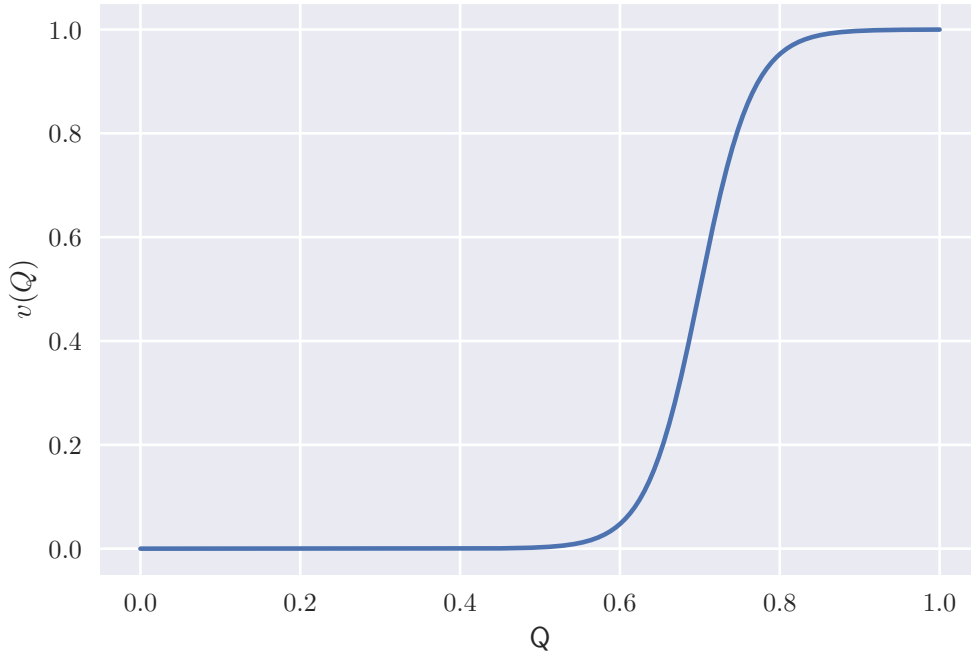


Figure 2.8: "Most" validity function .

We get $v(0.71) \approx 0.999 \doteq 1$. Hence we can say that most of people in D are young. Using the classical methods, we could compute the mean (18), the median

(20), and the standard deviation (≈ 9.65). Linguistic summary immediately yields meaning understandable in natural language. Using classical methods, we would require an additional expert interpretation (Yager, 1982; Hudec, 2016; Kacprzyk et al., 2001).

2.6 Fuzzy Inference

One of the main usages of logic always was to arrive from premises to conclusion using a rigorous mechanism, also called inferring. Fuzzy logic accommodates logical reasoning and generalizes rules of inference when the degrees of truth are introduced .

2.6.1 Generalized Modus Ponens

In classical propoositional logic, modus ponens (also known as affirming the antecedent) is the fundamental rule of inference. Given a proposition P and a proposition Q , a premise $P \rightarrow Q$ is true, and a premise P is true, we conclude that Q is true. Formally (Lotfi A Zadeh, 1996; Lotfi A Zadeh, 1988):

$$\begin{array}{c} P \rightarrow Q \\ P \\ \hline \therefore Q. \end{array} \quad (2.39)$$

This rule can be extended to predicate logic as

$$\begin{array}{c} P(x) \rightarrow Q(y) \\ P(x) \\ \hline \therefore Q(y), x \in X, y \in Y \end{array} \quad (2.40)$$

Example 2.6.1.1. If the measured wavelength is 550 nm, then the color is green. The wavelength was measured to be 550 nm. We, therefore, conclude that the color is green.

More formally, given the universe of discourse of measurements X , the universe of discourse of possible colours Y , the predicate "measured wavelength" $W(x \in X)$, and the predicate "is color" $C(y \in Y)$, we say

$$\begin{array}{c} W(550 \text{ nm}) \rightarrow C(\text{green}) \\ W(550 \text{ nm}) \\ \hline \therefore C(\text{green}). \end{array} \quad (2.41)$$

In basic inference, we expect absolute truth in our premises and absolute precision and certainty in our predicates. In the real world, this is rarely the case. To accommodate this reality, we extend modus ponens with fuzzy values. In example [2.6.1.1](#), we might get fuzzy values in (i) the rule, e.g., if the wavelength is *about* 550 nm, then the color is *somewhat* green. (ii) We also might get fuzzy values in the input, e.g., the color is *somewhat* green *but a little* yellow.

In fuzzy logic, we represent fuzzy implication as a binary relation between the antecedent fuzzy set \tilde{A} and the consequent fuzzy set \tilde{B}

$$\tilde{R}_{\rightarrow} = \tilde{A} \rightarrow \tilde{B} \quad (2.42)$$

Fuzzy inference is then given by a fuzzy composition (Lotfi A Zadeh, [1973](#); Hudec, [2016](#)).

$$\tilde{B}' = \tilde{A}' \circ \tilde{R}_{\rightarrow} \quad (2.43)$$

Generalized modus ponens can be represented from classical modus ponens as:

$$\begin{array}{c} \tilde{A}(x) \rightarrow \tilde{B}(y) \\ \tilde{A}'(x) \\ \hline \therefore \tilde{B}'(y), x \in X, y \in Y \end{array} \quad (2.44)$$

Expressed in terms of membership functions for:

$$\forall y \in Y, \mu_{\tilde{B}'}(y) = \sup_{x \in X} t(\mu_{\tilde{A}'}, \mu_{\tilde{R}_{\rightarrow}}(x, y)), \quad (2.45)$$

where $\sup-t$ is a generalized composition rule .

2.6.2 Mamdani Inference System

A special case of fuzzy inference (general in [2.45](#)) using the minimum t -norm (Mamdani et al., [1975](#)).

$$\forall y \in Y, \mu_{\tilde{B}'}(y) = \sup_{x \in X} \min(\mu_{\tilde{A}'}, \min(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(y))), \quad (2.46)$$

This rule is depicted with a crisp and a fuzzy antecedent in Figures [2.9](#), [2.10](#) respectively.

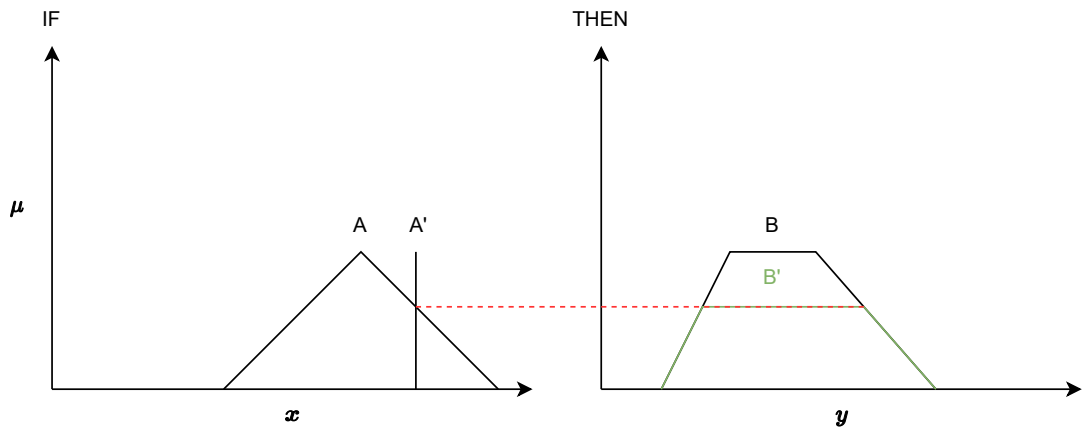


Figure 2.9: Mamdani inference example with one crisp antecedent (Hudec, 2016; Ross, 2005).

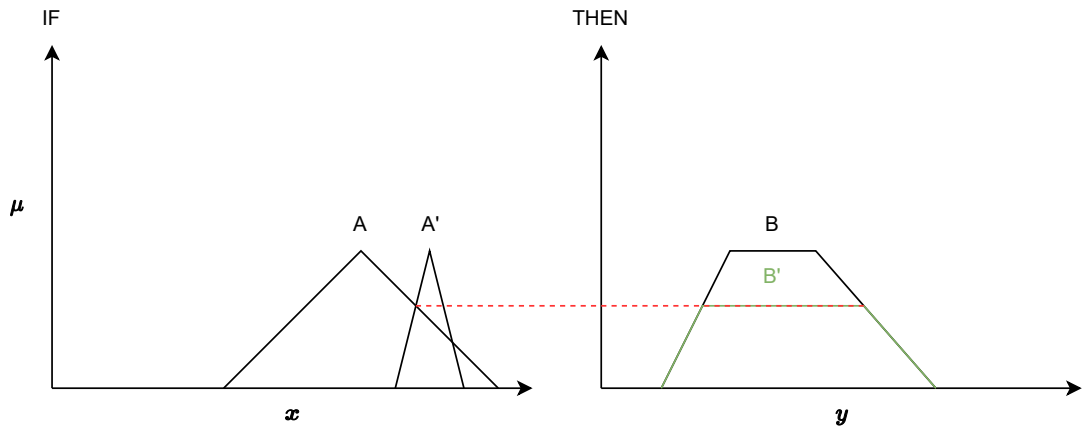


Figure 2.10: Mamdani inference example with one fuzzy antecedent (Hudec, 2016; Ross, 2005).

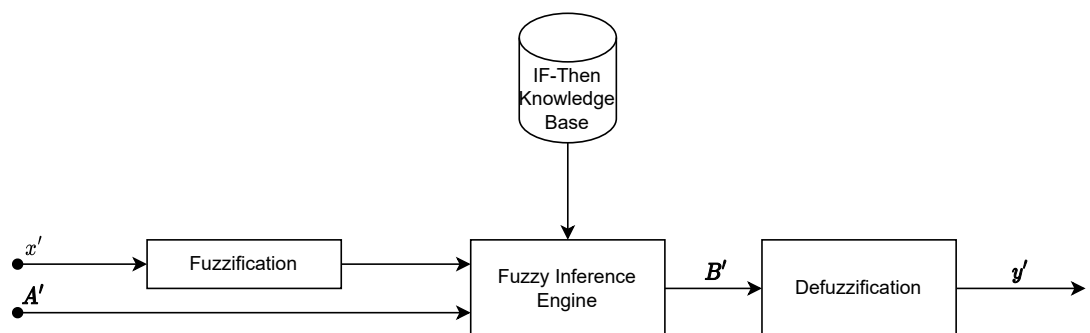


Figure 2.11: Mamdani inference system diagram (Hudec, 2016; Ross, 2005).

2.7 Defuzzification

The fuzzy consequent is a fuzzy set whose shape is harder to interpret by the consumer (human or machine). Defuzzification is a method of producing a crisp value from a fuzzy set (Hudec, 2016; Ross, 2005).

Definition 2.7.1 (Defuzzification). Defuzzification is a function taking a fuzzy set on X and mapping it to an element from X $D : \tilde{A} \rightarrow x \in X$.

2.7.2 Defuzzification based on extreme value

If one unique maximum does not exist in \tilde{A} (we say \tilde{A} is *non-unimodal*). We express a set of such values as :

$$E_s(x) = \{x \mid y \in Y \wedge \neg(\exists y \in Y)(\mu_{\tilde{B}'}(y) > \mu_{\tilde{B}'}(x))\} \quad (2.47)$$

Definition 2.7.3 (Center of Maxima (COM) defuzzification).

$$x = \frac{\min\{x \mid E_s\} + \max\{x \mid E_s\}}{2} \quad (2.48)$$

Chapter 3

Analysis of the state of the art of risk profile assessment

An argument could be made that a company that can correctly assess an investor's risk profile has the potential to have satisfied clients. For example, the correct assessment of risk tolerance could be a long-term investment, as evidence shows that risk tolerance is a quite stable personality trait and is unlikely to change for at least five years (Van de Venter et al., 2012). Customer satisfaction could then be the harbinger of further inflow of customer interest. However, the inverse also holds. The incorrect (by negligence or malice) assessment played role in the financial crisis of 2008. Investors were assessed to have higher risk tolerance and were dissatisfied with their portfolios (Pan et al., 2012).

The standard technique for investor risk profile assessment is the questionnaire (also called risk assessment questionnaire, investor questionnaire, etc.) The landscape of risk profile assessment questionnaires is very heterogeneous and the techniques of processing the results are opaque. There is a large variability in predictive performance among risk assessment questionnaires (Yook et al., 2003).

The literature usually agrees on three important components of the risk profile (*Personal financial planning – Requirements for personal financial planners*, 2005).

- (i) The risk requirement is the target wealth investor wants to achieve in a target time.
- (ii) The risk capacity is what risk the investor really can afford to take.
- (iii) The risk tolerance is the risk that the investor is willing to take.

Questionnaires are typically used to assess the risk tolerance and risk capacity component of the risk profile (Pan et al., 2012; *Personal financial planning – Requirements for personal financial planners*, 2005).

The classical risk assessment methods performed through questionnaires can be classified into questionnaires derived from economic and psychometric theory

(Grable et al., 2018). Grable et al. (2018) found that only psychometric data was a good predictor of risk-taking behavior.

When talking about risk assessment from the point of view of retirement investing, the most common representation is the equity:debt ratio. Markowitz (2009) describes this generally on a model where portfolios are created from $n \in \mathbb{N}$ uncorrelated assets of varying risk and return with proportions represented by weights (w_1, w_2, \dots, w_i) $\sum_{i=1}^n w_i = 1, w_i \in \mathbb{R}$. The model only talks about risk, i.e., volatility in the form of variance from the expected returns (assumed normal). Representing a risk profile as a combination of risky assets is a valuable tool; however, the assumptions about what risk is are too strict to be applied in the real world.

Chapter 4

Methodology

4.1 Risk Profile Assessment

In the real world, we say about an arbitrary investor x that their risk profile is R , or that they have a risk profile R . R can manifest itself as any terminus technicus used in the finance profession and the literature. E.g., institutions offering long-term investments for retirement (mutual funds, Exchange-traded Funds, etc.) would use terms such as *conservative*, *moderate*, or *aggressive* as characterizations of the investor's risk profile (Cipra, 2012; Carducci et al., 1998).

4.2 Building the Knowledge Base and the Inference System

We started with defining the risk profile \mathbb{R} in terms of our theoretical framework – fuzzy logic and context – retirement investments. We chose a top-down approach. We mind-mapped the essential components and determined how they might be measured (questionnaire, financial data...). The mind map is a tree-like structure, where the leaves resemble the most-atomic and the least-vague terms, for which conventional measurements are established (see Figure 4.1 for a diagrammatic representation).

¹We need to address that in finance, the risk usually pertains to known probabilities. In fuzzy logic, we tend to work more with uncertainty, i.e., the probability distribution is unknown. However, we believe it is unnecessary to introduce new nomenclature.

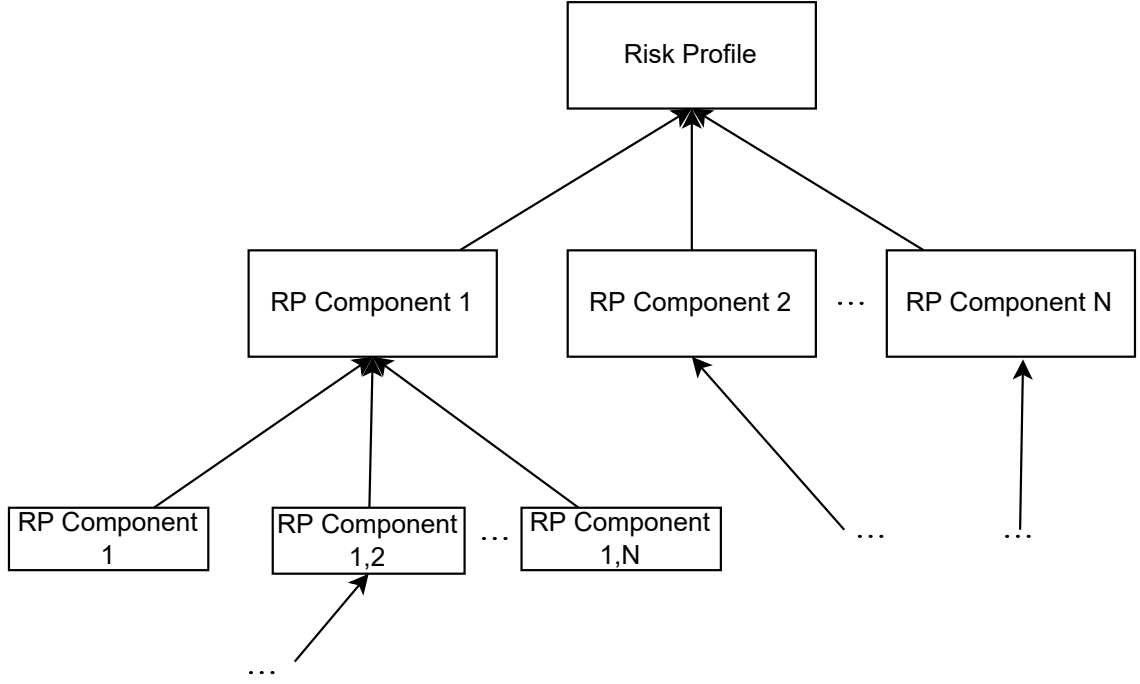


Figure 4.1: General decomposition mind map diagram.

We selected representative components to demonstrate our approach; our demonstration was compared to the classical methods. The advantages and limitations were discussed in Chapter 6. The selected components were then translated into linguistic variables.

The knowledge base was built from the relationships between the components.

We used the general inference system architecture. For the sake of simpler calculations, we used the Mamdani inference system (Mamdani et al., 1975). We selected maximum and minimum functions (Hudec, 2016) for s -norms and t -norms, respectively. General inference mechanism:

$$\begin{array}{l}
 \frac{A_1^{m+1}(x_1) \frown A_2^{m+1}(x_2) \frown \cdots \frown A_n^{m+1}(x_n) \rightsquigarrow C_j^m(y)}{A_1^{m+1}(x_1) \frown A_2^{m+1}(x_2) \frown \cdots \frown A_n^{m+1}(x_n)} \\
 \therefore C_j'(y)
 \end{array} \tag{4.1}$$

$x \in X, y \in Y, n, m, j \in \mathbb{N},$

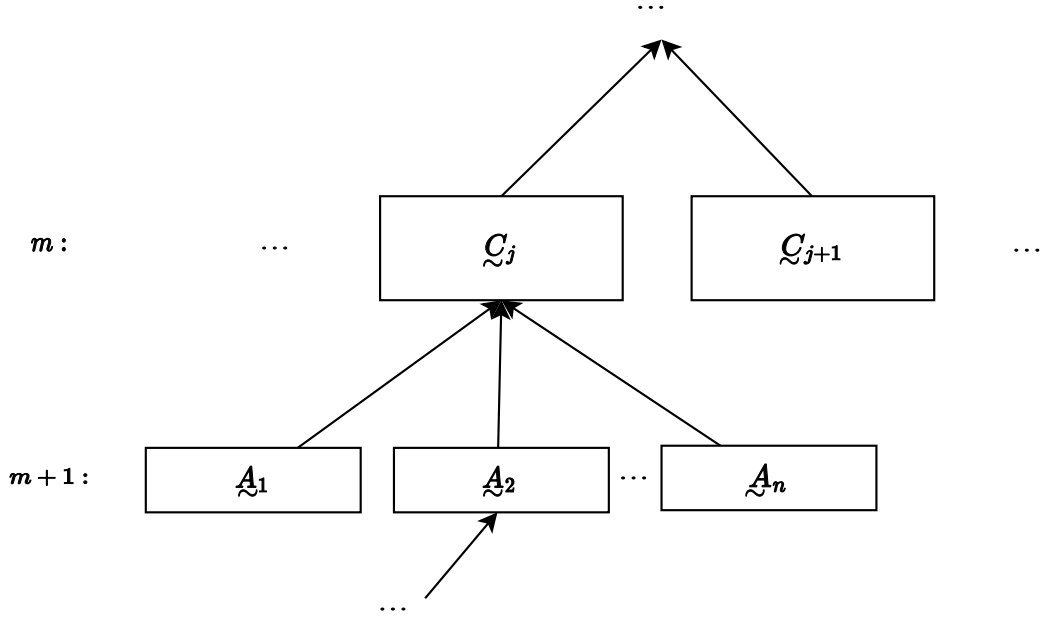


Figure 4.2: Diagram of Equation 4.1

m is the order of the tree mind-map. See Figure 4.2 for a diagram. Equation 4.1 and Figure 4.2 show that when not on the top of the tree (where the risk profile sits), any consequent goes into other inference sub-system's antecedent.

The mentioned context here was complementing retirement investments. We worked with the third pillar of the World Bank's framework (Holzmann et al., 2008). Because we work from the point of standard investor, we worked with classical conventional capital instruments such as equity shares, debt bonds, and public funds (we abstracted away from currency hedging) investing in such instruments (Cipra, 2012).

We opted for the Mamdani Inference Method and used piecewise linear fuzzy numbers (2.1.12) for easy computations.

Chapter 5

Application of the Proposed Method

The literature typically mentions three main components of risk profile (Cipra, 2012; Carducci et al., 1998; Grable et al., 2018; Barberis et al., 2003). (i) *Risk tolerance* is defined (linguistically) as the amount of risk the investor is willing to take. Components of risk tolerance could be education-based but also personality/behavior-based. We selected neuroticism as a psychometric measure of dealing with negative emotions. Conscientiousness is a psychometric measure of discipline. Financial literacy measures qualifications. Planned investments are scheduled payments. We classified planned investments as a component of risk tolerance (although we could reason for other classifications) because the discipline of sticking to the plan is related to Conscientiousness. (ii) *Risk capacity* measures how much risk the investor can withstand. In our model, risk capacity consists of the level of current wealth, the current plans undertaken by the investor, and current income. (iii) *Risk requirement* entails the requirement for future value. Because we are dealing with retirement investments, age is also an important component.

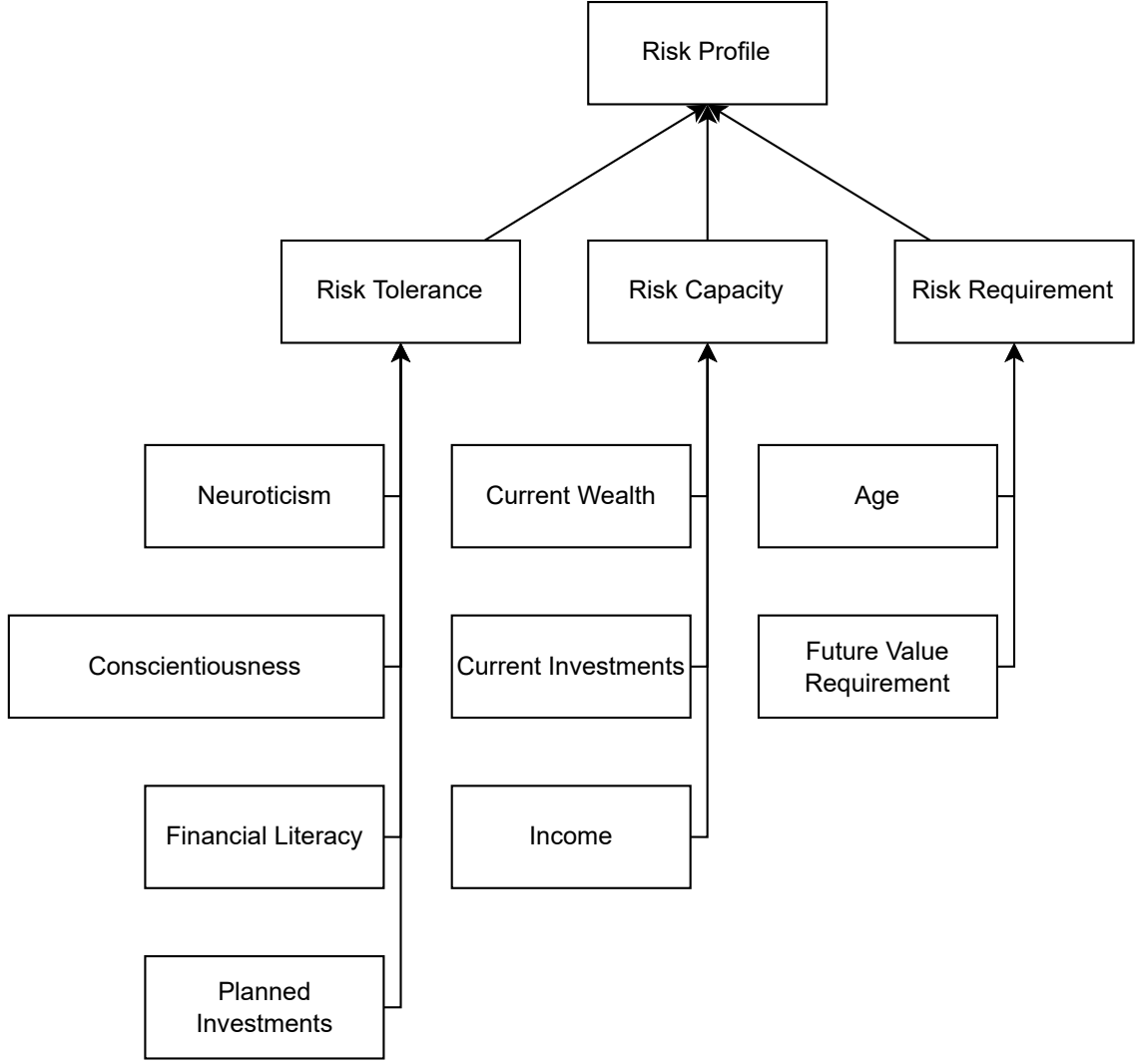


Figure 5.1: Decomposition created from the literature.

This risk profile in our context (see Chapter 3) then to the percentage of riskier instruments. Considering retail investors, a special case of two asset classes (equity and debt instruments) suffices (Markowitz, 2009). Because this is a common trend in the field, we selected these terms as a starting point. Given the universe X of retail investors, we can construct fuzzy predicates about each individual in the universe and assign them validity.

Let T be the set of term labels:

$$T = \{\text{Very Conservative, Conservative, Moderate, Aggressive, Very Aggressive}\} \quad (5.1)$$

let the universe of discourse be the proportions of equity in the portfolio $P = [0, 1]$.

5.1 Constructing fuzzy sets and computing memberships for elements

For the sake of simpler computations, we only opted for using piecewise linear functions given to the computer as an ordered n -tuple of significant points. Those points could then be obtained from an arbitrary function $F : \mathbb{R} \rightarrow [0, 1]$.

Any arbitrary degree of membership $\mu_p(x \in X)$ could be obtained by linear interpolation like so:

$$\mu_p(x) = \mu_p(x_0) + (x - x_0) \frac{\mu_p(x_1) - \mu_p(x_0)}{x_1 - x_0} = \frac{\mu_p(x_0)(x_1 - x) + \mu_p(x_1)(x - x_0)}{x_1 - x_0} \quad (5.2)$$

where $x_0, x_1 \in X$ are the nearest points from x with pre-computed μ . $x_0 < x < x_1$ holds. The points could then be passed to the computer raw or generated through a template function. Although libraries exist (Rada-Vilela, 2018; MATLAB, 2010), we opted for our implementation in Python 3. Rada-Vilela (2018) does not support fuzzy output, and MATLAB (2010) is a proprietary implementation.

Although the function in Equation 5.2 could theoretically approximate any template function, we only dealt with the following fuzzy numbers: triangular, trapezoidal, and left and right ramps. Diagrams of those numbers are shown in Figure 5.2, and membership degree functions are shown in Equations 5.3, 5.4, 5.5, and 5.6 (Hudec, 2016).

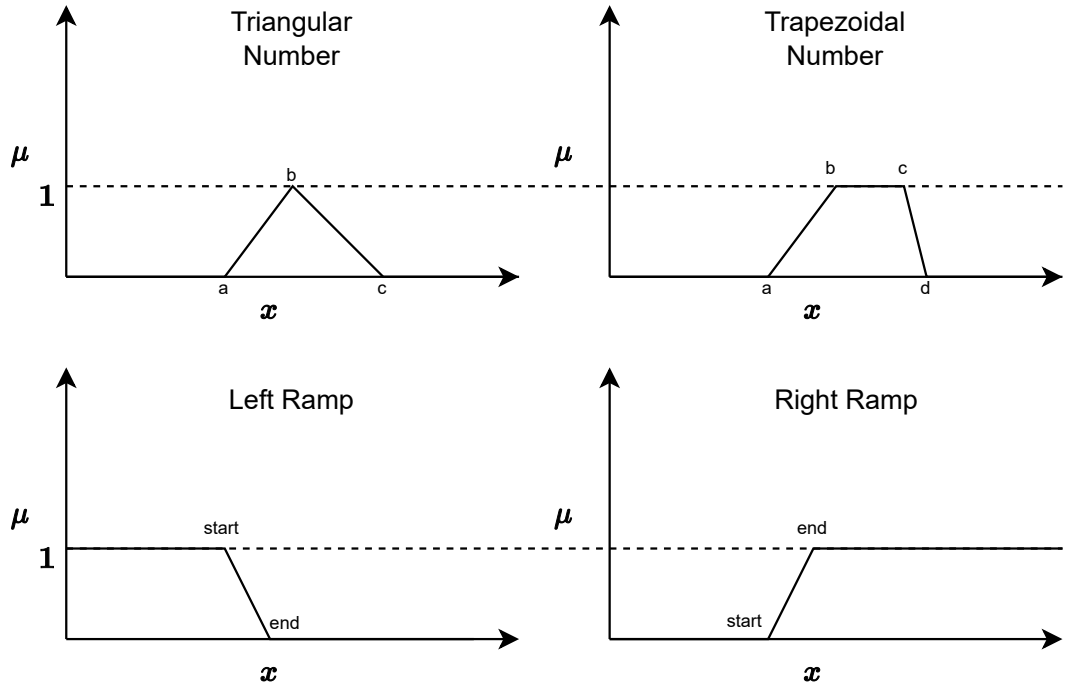


Figure 5.2: Types of used fuzzy numbers.

$$\mu_{\text{Triangular}}(x) = \int_{-\infty}^a 0 + \int_a^b \frac{x-a}{b-a} + \int_b^c \frac{c-x}{c-b} + \int_c^{\infty} 0 \quad (5.3)$$

$$\mu_{\text{Trapezoidal}}(x) = \int_{-\infty}^a 0 + \int_a^b \frac{x-a}{b-a} + \int_b^c 1 + \int_c^d \frac{d-x}{d-c} + \int_d^{\infty} 0 \quad (5.4)$$

$$\mu_{\text{Left Ramp}}(x) = \int_{-\infty}^{\text{start}} 0 + \int_{\text{start}}^{\text{end}} \frac{\text{end}-x}{\text{end}-\text{start}} + \int_{\text{end}}^{\infty} 1 \quad (5.5)$$

$$\mu_{\text{Right Ramp}}(x) = \int_{-\infty}^{\text{start}} 1 + \int_{\text{start}}^{\text{end}} \frac{x-\text{start}}{\text{end}-\text{start}} + \int_{\text{end}}^{\infty} 0 \quad (5.6)$$

5.2 Model Example

Please see the Annexes for the Application Example. Linguistic Variables used in the rule base are visualized in the Jupyter Notebook (Kluyver et al., [2016](#)).

Let us have an arbitrary, put particular investor. We collected information about their obligation adherence in the form of late obligation payments (e.g., mortgage, rent payments). We did not get the full list. However, the investor

told us there might be just a couple of additional late payments. We summarized the known missed payments and ended up with a triangular fuzzy number $A'_{\text{No. late payments}} = (a, b, c) = (7, 7, 8)$. We obtained the measure of conscientiousness and neuroticism from a questionnaire. Because it was not a conventional psychometric questionnaire, we obtained an imprecise value given by a triangular fuzzy number $A'_{\text{conscientiousness}} = (a, b, c) = (0.8, 0.85, 0.9)$ and $A'_{\text{neuroticism}} = (a, b, c) = (0.6, 0.8, 0.9)$.

The inference system took the questioned conscientiousness (taken from McCrae et al., 2008) along with evidence of discipline (no. late payments) and output adjusted (ideally real) measure of Conscientiousness.

Let rule base assessing risk conscientiousness be:

$$\begin{aligned}
& A_{\text{conscientiousness}}^{\text{high}}(x_1) \wedge A_{\text{no. late payments}}^{\text{at most few}}(x_2) \rightarrow C_{\text{conscientiousness}}^{\text{high}}(y) \\
& A_{\text{conscientiousness}}^{\text{high}}(x_1) \wedge A_{\text{no. late payments}}^{\text{more than few}}(x_2) \rightarrow C_{\text{conscientiousness}}^{\text{low}}(y) \\
& A_{\text{conscientiousness}}^{\text{moderate}}(x_1) \wedge A_{\text{no. late payments}}^{\text{more than few}}(x_2) \rightarrow C_{\text{conscientiousness}}^{\text{low}}(y) \\
& A_{\text{conscientiousness}}^{\text{low}}(x_1) \wedge A_{\text{no. late payments}}^{\text{more than few}}(x_2) \rightarrow C_{\text{conscientiousness}}^{\text{low}}(y)
\end{aligned} \tag{5.7}$$

The adjusted conscientiousness was then paired with measured neuroticism and run through the following rule base to obtain risk tolerance:

$$\begin{aligned}
& A_{\text{conscientiousness}}^{\text{low}}(x_1) \wedge A_{\text{neuroticism}}^{\text{high}}(x_2) \rightarrow C_{\text{risk tolerance}}^{\text{low}}(y) \\
& A_{\text{conscientiousness}}^{\text{high}}(x_1) \wedge A_{\text{neuroticism}}^{\text{low}}(x_2) \rightarrow C_{\text{risk tolerance}}^{\text{low}}(y) \\
& A_{\text{conscientiousness}}^{\text{moderate}}(x_1) \wedge A_{\text{neuroticism}}^{\text{moderate}}(x_2) \rightarrow C_{\text{risk tolerance}}^{\text{low}}(y)
\end{aligned} \tag{5.8}$$

From other (the origin does not have to be specified for the demonstration) sources, we got the measures of the remaining two components:

$$\begin{aligned}
A'_{\text{risk capacity}} &= \text{very low capacity} \\
&= \text{Hedged Left Ramp} = (a, b, c)^2 = (0, 0.1, 0.2)^2 \\
A'_{\text{risk requirement}} &= \text{somewhat moderate capacity} = \\
&= \text{Hedged Triangular} = (a, b, c)^2 = (0.3, 0.5, 0.4)^{0.5}
\end{aligned}$$

For the definition of linguistic hedges, see [2.3.11](#)

Those were then inputs in the final rule base:

$$\begin{aligned}
& \tilde{A}_{\text{risk tolerance}}^{\text{low}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{low}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{low}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{very conservative}}(y) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{low}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{high}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{moderate}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{conservative}}(y) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{low}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{moderate}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{low}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{conservative}}(y) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{moderate}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{high}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{low}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{conservative}}(y) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{low}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{high}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{low}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{Moderate}}(y) \quad (5.9) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{moderate}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{high}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{moderate}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{aggressive}}(y) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{moderate}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{high}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{moderate}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{aggressive}}(y) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{moderate}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{moderate}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{high}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{aggressive}}(y) \\
& \tilde{A}_{\text{risk tolerance}}^{\text{high}}(x_1) \wedge \tilde{A}_{\text{risk capacity}}^{\text{high}}(x_2) \wedge \tilde{A}_{\text{risk requirement}}^{\text{high}}(x_3) \rightarrow \tilde{C}_{\text{risk profile}}^{\text{very aggressive}}(y)
\end{aligned}$$

All inference calculation were carried according to Mamdani inference (Mamdani et al., 1975, see 2.46).

The system was programmed into Python3 (Van Rossum et al., 2009) (see Annexes for code).

The final measurement was defuzzified by 2.7.3. The defuzzified value for the risk profile was 0.5, which signifies a moderate risk profile.

Chapter 6

Discussion

We are able to pull out the set of rules from the knowledge base. We are able to do so in the natural language. This means that our approach accommodates transparency for non-technical users. Basic human reasoning is enough to get the gist of what goes into one's risk profile. This transparency can then be consumed by human specialists/advisors to use their technical expertise to polish the client's portfolio in order to account for the residual vagueness and uncertainty. Given that the client can also read the mechanism, they can get pointers to fine-tune their behavior.

We included personality traits in our model. Behaviour is a vital component of risk profile (Barberis et al., 2003), but it is also tough to quantify in terms of risk. We wanted to demonstrate the possibility of providing a vague value as an input. We drew from the big five personality model (McCrae et al., 2008) which describes personality dimensions in terms of (crisp) percentiles. Given the nature of psychometry, the measurements carry vagueness and imprecision. There are questionnaires to assess the traits as accurately as possible. Individuals can, however, also assess themselves, albeit with greater inaccuracies. We can, therefore, easily accommodate both approaches. If we get the input from an "accurate" questionnaire, we can use the crisp percentiles and put them through fuzzifications. Otherwise, we can use just a fuzzy estimate as an input.

One substantial advantage of linguistic variables is the ability to update them. It is possible to update any part of the variable when the knowledge becomes less uncertain.

There are also limitations. The example in this thesis is small. The real-world model can get arbitrarily large, and the transparency could get overshadowed by an information overload. The advantage of transparency could be preserved by deliberately omitting some details (on the leaves).

Another question is about the technical complexity of various membership functions. Terms such as triangle or trapezoid are straight-forward, as they can be understood as piece-wise linear functions. Therefore, given fuzzy antecedent \underline{A}' , we can select the necessary value for $x \in X$ (input for $\mu_{\underline{A}'}(x)$) by taking $h(\underline{A}' \cap \underline{A})$ (Definition 2.1.7). When dealing with arbitrary membership functions, the solution becomes non-trivial. There is not one general analytical solution to get an intersection. One way would be to approximate the functions. Non-linear terms must be approximated sufficiently, but the resolution cannot be too fine, lest the task becomes too computationally demanding.

Monetary measures such as monthly/yearly income (pay grade) suffer from two shortcomings. First, orders of magnitudes affect the fuzziness substantially. The timeliness of the solution must be considered. One should, for example, consider the difference between *about one thousand euro* and *about one million euro*. One solution would be to assign the interval around the core as a percentage rather than an absolute value.

Phenomena such as inflation change the semantic value of some variables over time. A possible solution could be periodic updating. Frequent updating would be simple, as adjusting the underlying fuzzy sets is a trivial task. Given an example variable $L = \text{"Income bracket,"}$ the semantic rule could be shifted (e.g., yearly) by the value of the inflation measure, e.g., the **Harmonized Consumer Price Index (HCPI)**. One possible shortcoming would be ending up with a semantic disparity. For example, if one income bracket B is represented by a triangular fuzzy number $\underline{B}(50,000; 75,000; 100,000)$ then given inflation $I = 3\%$, we would get the adjusted number $\underline{B}'(51,500; 77,250; 103,000)$. These numbers make sense in theory but may become disconnected from reality. Additional adjustments could then further shroud the reasoning behind the number. Therefore, we argue that adopting a similar approach to other parts of the knowledge base would result in opaqueness, which would go against the reason inference/expert systems are used in contrast to methods such as artificial neural networks.

In Chapter 5, a specific model example was given. The problems of how to construct fuzzy sets and which t/s -norms to use were put out of the scope of this thesis, as they would deserve a scholarly paper on their own.

When extracting results from the mechanism, defuzzification is used. Usually, this means that a crisp value is given to the user. Crisp values make sense in systems that still work with somewhat precise values, such as a cooling system adjusting the rotation per second on a cooling fan. There is much more uncertainty in risk assessment. Given output being a risk profile in portfolio allocation,

we argue it is acceptable and even somewhat preferable to give the output as a fuzzy value. We propose defuzzification happens *to a degree*. Some of the fuzziness would still be removed through defuzzification, but fuzziness would intentionally be added to the final value. The final result would then be presented as a linguistic term. In Chapter 5, our example risk profile came out as 0.5, which means 50% recommendation of equity asset in investor's portfolio and 50% of debt assets. The final result would probably be consumed by a specialist, who might appreciate fuzziness as room for adjustment. The question then is, how much defuzzification to apply in the final result.

The third pillar of retirement reform is a narrow context. We argue that our model can be extended to assessing risk profiles in other cases. Because retirement investing is just investing, we can relax the condition of retirement. The relaxation would mainly reflect in the risk requirement component. Risk profiles in the context of retirement investing are built from the equity:debt ratio (Cipra, 2012). In the context of general investments, other outputs could be used. For example, another (more general) model could be built assuming the Efficient Market Hypothesis (Malkiel, 1989). In investing, this hypothesis is usually explained by the "magic triangle of investing, (Figure 6.1' representing a growth-liquidity-risk tradeoff (Musílek, 1999).

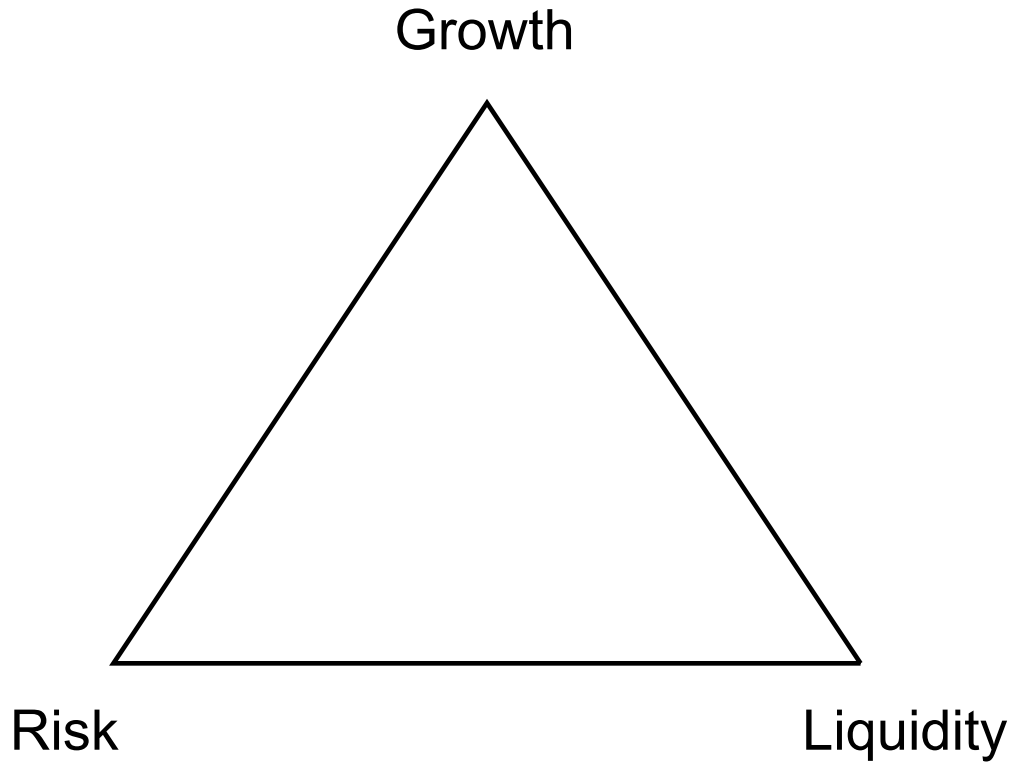


Figure 6.1: The magic triangle of investing (Musílek, 1999).

This general model would then map the components to the triangular space.

Another possible expansion is to the field of debt and credit. There are also many systems and models to assess credit risk. A generally known example is the credit score in the United States (Kiviat, 2019). This system is known to suffer from limitations such as opaqueness and hard understandability, i.e. the limitations we intended to tackle in this thesis. On top of the outlined limitations, it is alleged to suffer from unfairness, a problem that can be addressed by transparency and understandability (Campisi, 2021).

The future evolution of the inference system must take into account the quality of inference. This entails consistency in the meaning of variables across usages and the criteria of explainability and interpretability of the system (Moral et al., 2021; Mináriková, 2021).

Chapter 7

Conclusion

Fuzzy logic is a suitable tool for modeling uncertainty, vagueness, and imprecision, among many other things. Our thesis aimed to lay the foundation for using the fuzzy logic approach in private investors' risk assessment. We built our model framework as multiple fuzzy inference systems. We showed that this model could be understood by a non-technical audience thanks to the representation written in natural language. The representation is only subsequently translated through fuzzy logic to be understood by the computer. We see the main contribution in this transparency. We demonstrated the flexibility in updatability of the produced models, which could compensate for the non-timely nature of the models if done carefully. We implemented an example model in Python3 as a proof of concept and demonstrated its usage on a theoretical investor. Last but not least, we discussed the possibility of generalizing the model framework from the scope of retirement investing to investing in general.

Bibliography

- BARBERIS, Nicholas; THALER, Richard, 2003. A survey of behavioral finance. *Handbook of the Economics of Finance*. Vol. 1, pp. 1053–1128.
- CAMPISI, Natalie, 2021. *From Inherent Racial Bias to Incorrect Data—The Problems With Current Credit Scoring Models* [online] [visited on 2022-06-30]. Available from: <https://www.forbes.com/advisor/credit-cards/from-inherent-racial-bias-to-incorrect-data-the-problems-with-current-credit-scoring-models/>.
- CARDUCCI, Bernardo J; WONG, Alan S, 1998. Type A and risk taking in everyday money matters. *Journal of business and psychology*. Vol. 12, no. 3, pp. 355–359.
- CHRUBASIK, Vit, 2022a. *University of Huddersfield, MSc Fintech Course: Ageing Societies – Alternatives to employment for older citizens: exploring challenges and opportunities from the point of view of the current young workforce as future retirees*.
- CHRUBASIK, Vit, 2022b. *University of Huddersfield, MSc Fintech Course: Research Methods – Project Proposal*.
- CIPRA, Tomáš, 2012. *Penze: kvantitativní přístup*. Ekopress. ISBN 978-80-86929-87-3.
- DIJKMAN, JG; VAN HAERINGEN, H; DE LANGE, SJ, 1983. Fuzzy numbers. *Journal of Mathematical Analysis and Applications*. Vol. 92, no. 2, pp. 301–341.
- FRAENKEL, Adolf, 1925. Untersuchungen über die Grundlagen der Mengenlehre. *Mathematische Zeitschrift*. Vol. 22, no. 1, pp. 250–273.
- GAUTAM MITRA, Yu Xiang, 2016. *The Handbook of Sentiment Analysis in Finance*. Albury Books. ISBN 1910571571; 9781910571576. Available also from: libgen.li/file.php?md5=c14d4860ade44be22e4d91e8884907e9.
- GRABLE, John; HUBBLE, Amy; KRUGER, Michelle; VISBAL, Melissa, 2018. Predicting Financial Risk Taking Behavior: A Comparison of Questionnaires. In: *2019 Academic Research Colloquium for Financial Planning and Related Disciplines*.

- HARRIS, Charles R.; MILLMAN, K. Jarrod; WALT, Stéfan J. van der; GOMMERS, Ralf; VIRTANEN, Pauli; COURNAPEAU, David; WIESER, Eric; TAYLOR, Julian; BERG, Sebastian; SMITH, Nathaniel J.; KERN, Robert; PICUS, Matti; HOYER, Stephan; KERKWIJK, Marten H. van; BRETT, Matthew; HALDANE, Allan; RÍO, Jaime Fernández del; WIEBE, Mark; PETERSON, Pearu; GÉRARD-MARCHANT, Pierre; SHEPPARD, Kevin; REDDY, Tyler; WECKESSER, Warren; ABBASI, Hameer; GOHLKE, Christoph; OLIPHANT, Travis E., 2020. Array programming with NumPy. *Nature*. Vol. 585, no. 7825, pp. 357–362. Available from DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- HOLZMANN, Robert; HINZ, Richard Paul; DORFMAN, Mark, 2008. Pension systems and reform conceptual framework. *World Bank Discussion Paper*. Vol. 824.
- HUDEEC, Miroslav, 2016. Fuzziness in information systems. *Switzerland (CHE): Springer Nature*. ISBN 9783319425184.
- HUGHES, George Edward; CRESSWELL, Max J; CRESSWELL, Mary Meyerhoff, 1996. *A new introduction to modal logic*. Psychology Press.
- HUNTER, J. D., 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*. Vol. 9, no. 3, pp. 90–95. Available from DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- Personal financial planning – Requirements for personal financial planners*, 2005. Geneva, CH. Standard. International Organization for Standardization.
- KACPRZYK, Janusz; YAGER, Ronald R, 2001. Linguistic summaries of data using fuzzy logic. *International Journal of General System*. Vol. 30, no. 2, pp. 133–154.
- KIVIAT, Barbara, 2019. Credit scoring in the United States. *economic sociology_the european electronic newsletter*. Vol. 21, no. 1, pp. 33–42.
- KLIR, George; YUAN, Bo, 1995. *Fuzzy sets and fuzzy logic*. Prentice hall New Jersey.
- KLUYVER, Thomas; RAGAN-KELLEY, Benjamin; PÉREZ, Fernando; GRANGER, Brian; BUSSONNIER, Matthias; FREDERIC, Jonathan; KELLEY, Kyle; HAMRICK, Jessica; GROUT, Jason; CORLAY, Sylvain; IVANOV, Paul; AVILA, Damián; ABDALLA, Safia; WILLING, Carol, 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: LOIZIDES, F.; SCHMIDT, B. (eds.). *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pp. 87–90.
- MALKIEL, Burton G, 1989. Efficient market hypothesis. In: *Finance*. Springer, pp. 127–134.

- MAMDANI, Ebrahim H; ASSILIAN, Sedrak, 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies*. Vol. 7, no. 1, pp. 1–13.
- MARKOWITZ, Harry M, 2009. *Harry Markowitz: selected works*. World Scientific.
- MATLAB, 2010. *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc.
- MCCRAE, Robert R; COSTA JR, Paul T, 2008. The five-factor theory of personality.
- MINÁRIKOVÁ, E., 2021. Criteria for fuzzy rule-based systems and its applicability on examples. In: *Proceedings of the 24th International Scientific Conference for Doctoral Students and Post-Doctoral Scholars (EDAMBA), University of Economics in Bratislava*.
- MORAL, Alonso; CASTIELLO, Ciro; MAGDALENA, Luis; MENCAR, Corrado, 2021. *Explainable Fuzzy Systems*. Springer.
- MUSÍLEK, Petr, 1999. *Finanční trhy a investiční bankovníctví*. ETC. ISBN 80-86006-78-6.
- PAN, Carrie H; STATMAN, Meir, 2012. Questionnaires of risk tolerance, regret, overconfidence, and other investor propensities. *SCU Leavey School of Business Research Paper*. No. 10-05.
- RADA-VILELA, Juan, 2018. *The FuzzyLite Libraries for Fuzzy Logic Control*. Available also from: <https://fuzzylite.com/>.
- ROSS, Timothy J, 2005. *Fuzzy logic with engineering applications*. John Wiley & Sons. ISBN 9780470345764.
- VAN DE VENTER, Gerhard; MICHAYLUK, David; DAVEY, Geoff, 2012. A longitudinal study of financial risk tolerance. *Journal of Economic Psychology*. Vol. 33, no. 4, pp. 794–800.
- VAN ROSSUM, Guido; DRAKE, Fred L., 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN 1441412697.
- YAGER, Ronald R, 1982. A new approach to the summarization of data. *Information Sciences*. Vol. 28, no. 1, pp. 69–86.
- YOOK, Ken C; EVERETT, Robert, 2003. Assessing risk tolerance: Questioning the questionnaire method. *Journal of Financial Planning*. Vol. 16, no. 8, p. 48.

- ZADEH, Lotfi A, 1973. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on systems, Man, and Cybernetics*. No. 1, pp. 28–44.
- ZADEH, Lotfi A, 1983. Linguistic variables, approximate reasoning and dispositions. *Medical Informatics*. Vol. 8, no. 3, pp. 173–186.
- ZADEH, Lotfi A, 1988. Fuzzy logic. *Computer*. Vol. 21, no. 4, pp. 83–93.
- ZADEH, Lotfi A, 1996. Fuzzy sets. In: *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*. World Scientific, pp. 394–432.
- ZADEH, Lotfi Asker, 1975. The concept of a linguistic variable and its application to approximate reasoning—I. *Information sciences*. Vol. 8, no. 3, pp. 199–249.
- ZERMELO, Ernst, 1908. Untersuchungen über die Grundlagen der Mengenlehre. I. *Mathematische Annalen*. Vol. 65, no. 2, pp. 261–281.

List of Abbreviations

ETF Exchange-traded Fund. [22](#)

HCPI Harmonized Consumer Price Index. [32](#)

List of Annexes

Ageing Societies Essay Paper
Research Methods Proposal Paper
Application Implementation
Application Notebook

A
B
C
D

Annexes

Appendix A

**Ageing Societies – Alternatives to employment for older citizens:
Exploring challenges and opportunities from the point of view of the current young workforce as future retirees**

**Alternatives to employment for older citizens: exploring
challenges and opportunities from the point of view of the
current young workforce as future retirees**

Vit Chrubasik

Student ID U1972463

Individual Essay
BMO0263-QGN-2122: Ageing Societies

School of Business
University of Huddersfield
United Kingdom
December 2021

It is common knowledge that the rise of modern medicine and the decrease in natality is causing a substantial change in the age structure in the developed world. There is a higher proportion of people being on and beyond the post-productive threshold of 65 years of age, with the predictions of a continuing significant increase in the elderly populace (Gratton & Scott, 2016; van Dalen et al., 2010). The so-far prevailing approach to retirement being primarily considered a one-off event of leaving work and living off (commonly a state) pension would mean a lesser proportion of the people contributing to the economy and more people needing to be supported by the economy. This development is evidently macro-economically unsustainable (Cavendish, 2019). This is a complex interdisciplinary problem spanning multiple fields of social sciences. The frontier of our understanding of the problem is being created and expanded not only in research but also by the measures which are being proposed and taken on societal, organisational, and individual levels. The literature agrees that pursuing alternatives to employment for older citizens may be one of the solutions to the described issue. Cavendish (2019) makes additional cases for staying in work longer, including health benefits and longevity. Cavendish (2019) mentions the undesired effects on mental health. Here, the cause is attributed to the loss of sense of meaning, which is a known important psychological factor described by the seminal work of Frankl et al. (2006).

In this essay, we focus on future retirees. We weigh the focus towards the early stages of the productive age. We argue that this demographic is an important stakeholder in the topic, for they are at the risk of carrying a greater burden than the current old populace. We draw from Shultz and Wang (2011) and understand the retirement as a process, rather than an event in time. This process is also called bridge-retirement and entails a variety of (not only) economically productive activities. We present three ideas relating to the alternative to employment for older citizens taking into account the time current younger citizens have on their hands. We relax the rigid constraint of the post-productive age threshold of 65.

(i) Thanks to the effect of compounding, a young individual is in the best position to build a significant surplus of funds. A capital provider can then reallocate their funds to the economic units in need, i.e. become an investor or lender. They can then generate passive income from financial activities. (ii) An individual can also decide to pursue a venture of their own, either by starting a business or freelancing. The surplus of funds here can serve as equity and alleviate the risk of high leverage, which a retiree might not want to, or even could not expose themselves

to. An individual has a high chance of being perfectly capable of continuing working in either their current line of work or making changes in their career. (iii) Modern medicine along with the factors such as healthy lifestyle not only adds years to one's life but also adds life to one's years (Gratton & Scott, 2016; Zissimopoulos & Karoly, 2007). This means that the retirement age threshold becomes fuzzier and fuzzier, with the limit shifting towards later life.

We argue that informing the younger audience about the ideas and options regarding the alternatives to employment in the older age not only yields the fruits in the form of satisfied retirees in the future but also can spread the awareness about the current issues of the current older generations and therefore serve as an important building block in the bridge across the generation-gap of understanding and compassion. We discuss these ideas from the point of view of individuals, organisations, and society. We outline some opportunities and challenges for each level.

Investing is one of the essential ideas for young individuals in preparing for retirement. Consider this simple example. Given a fixed monthly payment C , an interest rate r , we can use a model of an ordinary annuity to calculate a result amount FV after t years of investing (Equation 1. We abstract away from risk and inflation for the sake of simplicity).

$$FV = C \cdot \frac{(1+r)^t - 1}{r} \quad (1)$$

Let $C = £100$, and $r = 6\%$ (a conservative rate for stocks investments). Twenty years of investing give £46,000, whereas thirty years of investing £105,000, and forty years of investing amount to almost £200,000. The role of time in this model is vital. Each additional year has an increased impact on the final amount of funds. Adjusting the other variables can then provide a young individual with a good amount of funds in the future with a bearable sacrifice of monthly income.

Private capital from individuals is able to supply loanable funds where traditional funding such as bank loans or initial public offerings would not be possible. An individual challenge stands out from the behavioural point of view. Building capital is a long-term endeavour that requires discipline, critical thinking, and frugality, all housed by the term of financial literacy. With the wage differences, each individual faces a different sacrifice of their present benefit for the future. There are also other variables, such as risk and liquidity, which are inherent to any investment project. This strategy becomes exponentially less viable for older individuals.

Some organisations may benefit from the supply of capital owned by individuals. With

an increase of angel investors (the term coined by the start-up culture), the funding for start-ups may become cheaper. There is an industry of financial services aimed towards retirement. The current marketing is oriented towards the traditional full retirement. An opportunity to reach a wider audience can be found in creating a market that is more oriented towards active retirement.

It is difficult to predict the emergent behaviour of the market with a substantially increased supply. Following the classical models, the price of capital should decrease. Even this oversimplification produces a challenge of raising the barrier to entry, which is already high. One consequence might be that the rate of return and risk will become infeasible for more and more young individuals.

There are many ways to build capital. The effect of compounding is a great model, understanding of which can serve as an initial pull factor to investing. However, this is an oversimplification, and successful investment sometimes requires more than a good level of financial literacy. The level of financial literacy itself is already low, as the meta-analysis from Miller et al. (2014) suggests.

Transition to self-employment may be a more attractive path for other future older workers. The starting capital requirement can be substantially lower than pure investment. Self-employment carries a degree of labour intensity. This active participation in the economy may help with the mental health issues connected to retirement (Cavendish, 2019). The early age can again benefit significantly in such transition. An individual who is aware of the possibility of such transition can search for opportunities during their working life and actively build-up starting capital. The research from Curran and Blackburn (2001) showed that not many people intended to pursue the self-employment path as part of the bridge-retirement. The contrast between studies by van Solinge (2014), and Moulton and Scott (2016) show that a proactive approach towards self-employment is preferable, as there a reactionary approach leads to unwanted work.

People coming from employment have lots of work experience and knowledge in the area of their field of work. This may be helpful in starting the venture in an area in which the individual was employed before retirement.

Expertise in the field is a sufficient condition for success, but entrepreneurial skills are necessary. Those skills may not be developed in conventional pre-retirement employment.

Meaning in life is proving to be necessary for mental well-being and is often tied to the working life (Cavendish, 2019). Terminating employment can therefore be a drastic measure with not

only economic consequences. When full-time employment may not be possible for valid reasons, an individual may seek quantitative change (fewer hours) or qualitative change (different job) or both (Wang et al., 2008). The organisational support of older employees remains an open question to the future with the current situation being in its early stages (van Dalen et al., 2010).

The three ideas discussed in this essay are not mutually exclusive. Investing in capital provides largely passive income. It can therefore be a supplement to the income stream.

As the younger audience has a long way to go to retirement, the organisational and societal levels become tougher to predict years in the future. We, therefore, argue that at this stage, a focus on the individual level should be emphasised.

We try to present the case to a broad young audience of individuals. We make some critical assumptions, which impose limitations on the essay.

Even though there are countries that have banned the mandatory retirement age (UK, Denmark, Poland in Europe), this is not the case for the majority (*Pensions at a Glance 2021*, 2021). This fact has to be considered, as it sets limitations to this essay, where we assume not retiring is a viable option.

This essay focuses on the alternatives providing economic benefits to the individual. It, therefore, leaves out other options such as volunteering, which offers well-being benefits such as fulfilment and a sense of meaning in the older age (Huang, 2019).

The ideas all assume a (sufficiently) healthy individual being able and willing to work in the future. The area of health and well-being alone presents challenges beyond the scope of this essay.

The case against full retirement is being made from different points of view. State pensions not being able to support an individual in retirement, and the physical and mental health implications are the two main motivations for an individual are two main drivers for individual motivation (extrinsic and intrinsic respectively) to be informed about the alternatives to pre-retirement work.

We outlined the ideas of building capital, self-employment and post-retirement work as the main approaches to be aware of, with proactivity being the overarching theme in all three of them. We also outlined Some of the opportunities and challenges to give the stakeholder initial ideas on how to approach this issue with respect to their current conditions and environment.

References

- Cavendish, C. (2019). The case for staying at work a decade longer — or even more. *Financial Times*.
- Curran, J., & Blackburn, R. A. (2001). Older people and the enterprise society: Age and self-employment propensities. *Work, employment and society*, 15(4), 889–902.
- Frankl, V. E., Lasch, I., Kushner, H. S., & Winslade, W. J. (2006). *Man's search for meaning*. Beacon Press.
- Gratton, L., & Scott, A. (2016). The 100-year life: Gift or curse? *London Business School review*, 27(3).
- Huang, L.-H. (2019). Well-being and volunteering: Evidence from aging societies in asia. *Social science & medicine* (1982), 229, 172–180.
- Miller, M., Reichelstein, J., Salas, C., & Zia, B. (2014). Can you help someone become financially capable? a meta-analysis of the literature. *A Meta-Analysis of the Literature (January 1, 2014)*. *World Bank Policy Research Working Paper*, (6745).
- Moulton, J. G., & Scott, J. C. (2016). Opportunity or necessity? disaggregating self-employment and entry at older ages. *Social forces*, 94(4), 1539–1566.
- Pensions at a glance 2021*. (2021). OECD. <https://doi.org/10.1787/ca401ebd-en>
- Shultz, K. S., & Wang, M. (2011). Psychological perspectives on the changing nature of retirement. *American Psychologist*, 66(3), 170.
- van Dalen, H. P., Henkens, C. J. I. M., & Schippers, J. (2010). How do employers cope with an ageing workforce? views from employers and employees. *Demographic research*, 22, 1015–1036.
- van Solinge, H. (2014). Who opts for self-employment after retirement? a longitudinal study in the netherlands. *European journal of ageing*, 11(3), 261–272.
- Wang, M., Zhan, Y., Liu, S., & Shultz, K. S. (2008). Antecedents of bridge employment: A longitudinal investigation. *Journal of applied psychology*, 93(4), 818–830.
- Zissimopoulos, J. M., & Karoly, L. A. (2007). Transitions to self-employment at older ages: The role of wealth, health, health insurance and other factors. *Labour economics*, 14(2), 269–295.

Dissemination Plan

The key stakeholders belong to the people freshly into their productive years or just before entering the productive years. Schools are the preferred places to conduct dissemination because most of the audience can be approached here. There are several groups of stakeholders found in schools. The groups are divided by grade. The belief is that the younger the audience is targeted, the wider audience can be reached; however, there is an adverse effect caused by the development in thinking and perception of information. More complex messages from this essay might not reach the audience correctly. There also is evidence from Miller et al. (2014) showing that unsolicited advice has little effect on decision making. We mention an industry of financial services in our idea of the capital building. Informing individuals in this industry may create a synergistic effect with the already existing motivation of making sales. Therefore, we propose carrying out traditional presentation seminars in the broker companies to reach the audience through a middle man.

Another middleman is the teachers, who can benefit from the published copy of this essay, pull important information or go deeper in researching the work this essay references.

Online media can also be used. Bite-sized videos are feasible, as those can be reused. An online platform such as YouTube is an appropriate candidate, as it enables targeting the right audience.

A dissemination plan aims to spark intrinsic motivation in a young audience. We highlight channels of dissemination along with an estimate of relative effects. The Table 1 shows the dissemination plan. Our estimated effects are relative, with one resembling the lowest effect and three the highest effect of carrying the dissemination through the channel.

Stakeholder	Channel	Effect
Senior High School Students	School seminars	1
Junior High School Students	School seminars	1
University Students	School seminars	1
Teachers	published copy of this paper	2
Financial Brokers	Presentations	3
Students and graduates	Bite-sized informational videos online	2

Table 1: The dissemination plan.

Appendix B

Querying and Data Summarization with Fuzzy Logic: Helping Pensions Advisors to Search for Clients based on Risk Profile and other Fuzzy Data – Research Project Proposal

**Querying and Data Summarization with Fuzzy Logic:
Helping Pensions Advisors to Search for Clients based on
Risk Profile and other Fuzzy Data**

Research Project Proposal

Vit Chrubasik

Student ID U1972463

Assignment Paper
BMO0116-QGA-2122: Research Methods

MSc Fintech
School of Business
University of Huddersfield
United Kingdom
June 2021

Contents

1 Background to the research 2

1.1 Research Aim 2

1.2 Research Objectives 2

2 Literature Review 2

2.1 Fuzzy Sets 2

2.2 Fuzzy Logic 2

2.3 Linguistic Variables 2

2.4 Linguistic Summaries 3

2.5 Fuzzy Querying and Fuzzy Databases 4

2.6 Relevant Applications 4

2.7 Conceptual Framework 4

3 Methodology 5

3.1 Research Approach 5

3.2 Data Collection 5

3.3 Sample 5

3.4 Data Analysis 5

List of Figures

1 Linguistic variable "Water Temperature." 3

2 "Is young" predicate membership function. 3

3 "Most" validity function. 4

4 Conceptual Framework 5

List of Abbreviations

SQL Structured Query Language. 4

1 Background to the research

The topic of pensions is ever-present throughout an individual's life. To take responsibility and control over securing well-being in their post-productive phase of life (Holzmann et al., 2008), individuals take on the role of private investors and use financial instruments offered to them on the market (Cipra, 2012). Because investing on one's own is a time-consuming endeavour, this work is offloaded to specialised financial services, such as funds of various kinds. Financial advisors gather data about clients and build risk profiles to match investors to suitable products.

The main problem with clients' data is their elusive quantifiability. One important example is *risk profile*. The term itself is vague and open to interpretation. Theories and models mainly deal with precise (crisp) values; more advanced models use probability theory. It is hence not being accounted for what is not (yet) known or for what is uncertain. Fuzzy logic build on top of fuzzy set theory (Zadeh, 1996, 1999b) may have a solution for capturing uncertain, vague, imprecise and otherwise fuzzy phenomena.

1.1 Research Aim

We (Chrubasik, will be published July 2022) explored the assessment of individual risk profiles through fuzzy logic. We came up with outputs expressed as fuzzy variables for each investor data as input. We intend to build on top of the aforementioned method. We aim to investigate the usage of fuzzy logic for matching financial advisors with clients, given fuzzy risk profile data. We aim to do so from the point of view of an advisor searching through data about clients. We aim to implement fuzzy logic in various parts of the matching process.

1.2 Research Objectives

1. Develop a method for querying data about potential clients when the query is fuzzy.
2. Develop a method for querying data about potential clients when the data is fuzzy.
3. Use linguistic summaries to describe data about potential clients.
4. Use the previous methods to create a priority list of potential clients to contact.
5. Juxtapose the methods using fuzzy logic to classical methods.

2 Literature Review

We queried the federated databases of Google Scholar and Summon. In relevant articles, we used backtracking to get to the relevant references.

2.1 Fuzzy Sets

In classical theory, given a set A , element x either is or is not a member of A . More formally:

$$x \in A \vee x \notin A$$

Zadeh (1996, originally 1965) extends the idea by introducing *degree of membership*. By defining degree of membership μ of element x equal to 0 if and only if $x \notin A$, and equal to 1 if and only if $x \in A$, Zadeh is able to define fuzzy membership as function $\mu : X \rightarrow [0, 1]$, where X is an arbitrary universe of discourse. Zadeh then defines fuzzy set \tilde{A} as a collection of elements from X , and a function assigning them memberships, or formally:

$$\tilde{A} := \{A \in X, \mu(x \in A) \mid \mu : X \rightarrow [0, 1]\}$$

2.2 Fuzzy Logic

Zadeh further used fuzzy set theory to construct the field of fuzzy logic (1999b). Statements of fuzzy logic have *degree of truth* in $[0, 1]$. Zadeh demonstrates that fuzzy logic provides a methodology to compute with words from natural language by understanding words as constraints on variables; e.g. in the predicate

$$p = \text{water is hot},$$

hot constrains the water temperature. For predicate $P(\text{is hot})$, we can construct a fuzzy set where $\mu : \text{Water Temperature} \rightarrow [0, 1]$. We thereby assign a truth value of hotness to each temperature. Computing with words is a suitable methodology in cases when we (i) *do not know rationale*, e.g. poorly defined probabilities and utilities. (ii) *do not need rationale* – there is tolerance for imprecision. (iii) *cannot solve rationale* – numerical computing is not enough. (iv) *cannot define rationale* – the concept is too complex to define precisely (Zadeh, 1999a).

2.3 Linguistic Variables

Linguistic variables (Zadeh, 1973) were created as a central tool in a framework for computationally dealing with highly complex and/or not well-defined systems (Zadeh, 1975a, 1983; Zadeh, 1975b). Linguistic variable LV (for graphical example see Figure 1) is a set $\{L, T, X, G, H\}$, $|LV| = 5$, where:

- L is the name of the variable
- T is a set of all linguistic labels of L
- X is the universe of discourse
- G is the syntactic rule to generate T
- H is a semantic rule, a fuzzy subset of X , giving each label $t \in T$ a meaning (polished definitions from Hudec, 2016; Ross, 2005)

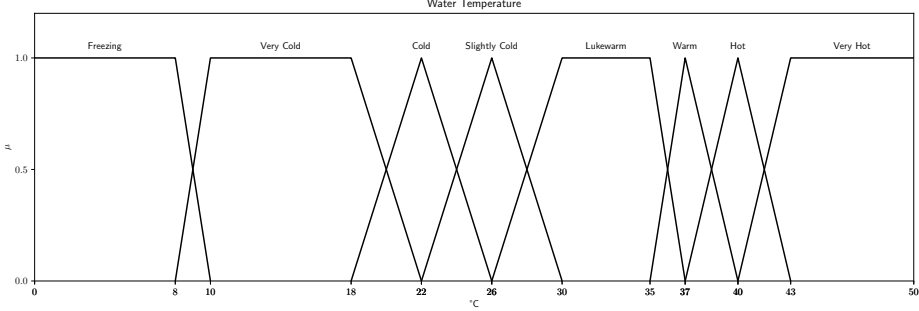


Figure 1: Linguistic variable "Water Temperature."

Linguistic variables are argued to be the tool to bridge real-world and computer data representation (Zadeh, 1983). They are the building block of a bridge connecting the world of human and computer reasoning. One substantial advantage of linguistic variables is the ability to update them. It is possible to update any part of the variable when the knowledge becomes less uncertain.

2.4 Linguistic Summaries

Data is meaningless unless described to the user. However, classical (statistical) methods produce numbers that often appear cryptic at first sight (Yager, 1982). Yager (1982) makes a case for linguistic summaries as an alternative to numeric quantities such as mean or variance. Linguistic summaries take on a form of a sentence of natural language. Given a data-set $D = \{d_1, d_2, \dots, d_n\}$, $n \in \mathbb{N}$, we can create linguistic summaries in the form:

Q elements of D have property P .

This is a fuzzy statement and has a truth value, also called validity $v \in [0, 1]$.

$$P : D \rightarrow \tilde{D}$$

$$Q(D) = \begin{cases} \frac{1}{n} \sum_{i=1}^n \mu_{\tilde{D}}(d_i) & \text{if } Q \text{ proportional} \\ \sum_{i=1}^n \mu_{\tilde{D}}(d_i) & \text{if } Q \text{ absolute} \end{cases}$$

$$v : Q \rightarrow [0, 1]$$

Example. Let D be the dataset of people's ages (duplicates allowed, people are distinct).

$$D = \{15, 13, 14, 33, 35, 20, 19, 8, 5\}$$

Most people are young.

$P : d \in D$ is young

$Q : \text{most}$

Given predicate "is young" P (similar to Yager, 1982):

$$P(d \in D) = \frac{1}{1 + e^{x-35}},$$

yielding membership function of shape in Figure 2,

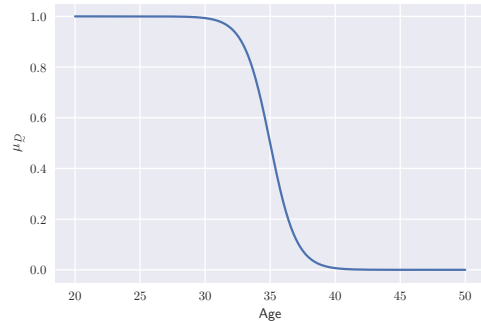


Figure 2: "Is young" predicate membership function.

computation of the predicate for D yields fuzzy set \tilde{D} (using standard notation of fuzzy sets $\tilde{A} := \left\{ +_{i=1}^n \frac{\mu_{\tilde{A}} a_i}{a_i} \right\}$; $+$ denotes fuzzy union)

$$P(d \in D) = D \approx \left\{ \frac{1}{15} + \frac{1}{13} + \frac{1}{14} + \frac{0.88}{33} + \frac{0.5}{35} + \frac{1}{20} + \frac{1}{19} + \frac{1}{8} + \frac{1}{5} \right\}$$

We compute the value of the quantifier Q :

$$\begin{aligned} \text{most of (proportional)} &= Q(D) \\ &= \frac{1}{n} \sum_{i=1}^n \mu_{(d_i)} \approx 0.93. \end{aligned}$$

Given the validity function (for shape see Figure 3)

$$v(Q) = \frac{1}{1 + e^{-30Q + 30 \cdot 0.7}},$$

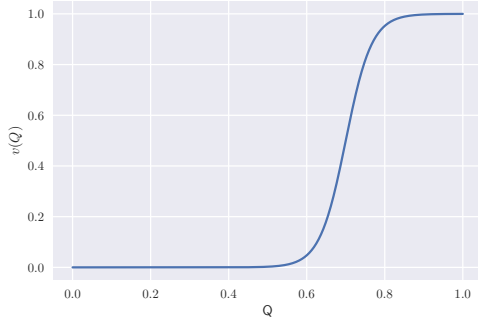


Figure 3: "Most" validity function.

We get $v(0.71) \approx 0.999 \doteq 1$. Hence we can say that most of people in D are young.

Using the classical methods, we could compute the mean (18), the median (20), and the standard deviation (≈ 9.65), but we would only end up with numbers. We would then be forced to make inferences based on opaque reasoning (1982).

It is evident from the example that this approach provides more freedom in talking about data than the classical methods. There is a question of how to come up with the membership functions. The uncertainty can also come from subjectivity (e.g. the quantifier "most"). In the case of Yager (1982), the function was plotted by drawing on a graphic tablet by a user.

2.5 Fuzzy Querying and Fuzzy Databases

Data accessing will be central in any proposed solution. Structured data, such as those about risk profiles are commonly stored in relational databases (Codd, 2002) and queried by **Structured Query Language (SQL)** (Chamberlin &

Boyce, 1974; Date, 1989). The uncertainty can occur both at the query phase and in the data (Hudec, 2009, 2016).

Classical querying requires precise inputs. Often, users' preferences are not precise. They often have to opt for arbitrary precise values. This fact means that the database is more likely to return empty or overabundant results (Bosc et al., 2008). The querying has to be then iteratively fine-tuned by the user, resulting in additional queries (Hudec, 2016).

SQL is already well-optimised for querying. Hence, it is only necessary to translate the condition to an SQL-compatible language (Hudec, 2016).

```
SELECT [distinct] <attributes>
FROM <relations>
WHERE <fuzzy condition>
```

There are many approaches to fuzzify SQL-like queries along with solutions (see Bosc and Pivert, 1995, 2000; Kacprzyk and Zadrozny, 2001; Tahani, 1977; Wang et al., 2007). Both the translation of queries and the representation of fuzzy data can be abstracted away into a black box in the scope of our project.

The relational model does not accommodate fuzzy data by default, because fuzzy sets are not atomic, and therefore collide with the First normal form (Codd, 1971, 1972). Hudec (2014) describes the extension of a traditional relational model by a fuzzy meta-model. Each fuzzy attribute is represented as a foreign key link to the fuzzy meta-model.

2.6 Relevant Applications

Hudec and Brokešová (2017) used querying to mine linguistic summaries from financial literacy questionnaires. Financial literacy is commonly considered a vital component of risk profile.

Vučetić and Hudec (2018) used fuzzy querying to create an engine for product searching (e-commerce).

These papers describe real-world applications of previously discussed methods. They can be interpreted as proofs of concept for our proposed project.

2.7 Conceptual Framework

We represent the framework (Figure 4) as a diagram of our proposed process with all the relevant techniques (keywords) from the literature review (Section 2). The Fuzzy Risk Assessment model (Chrubasik, 2022) can be represented as a black box in the context of this paper.

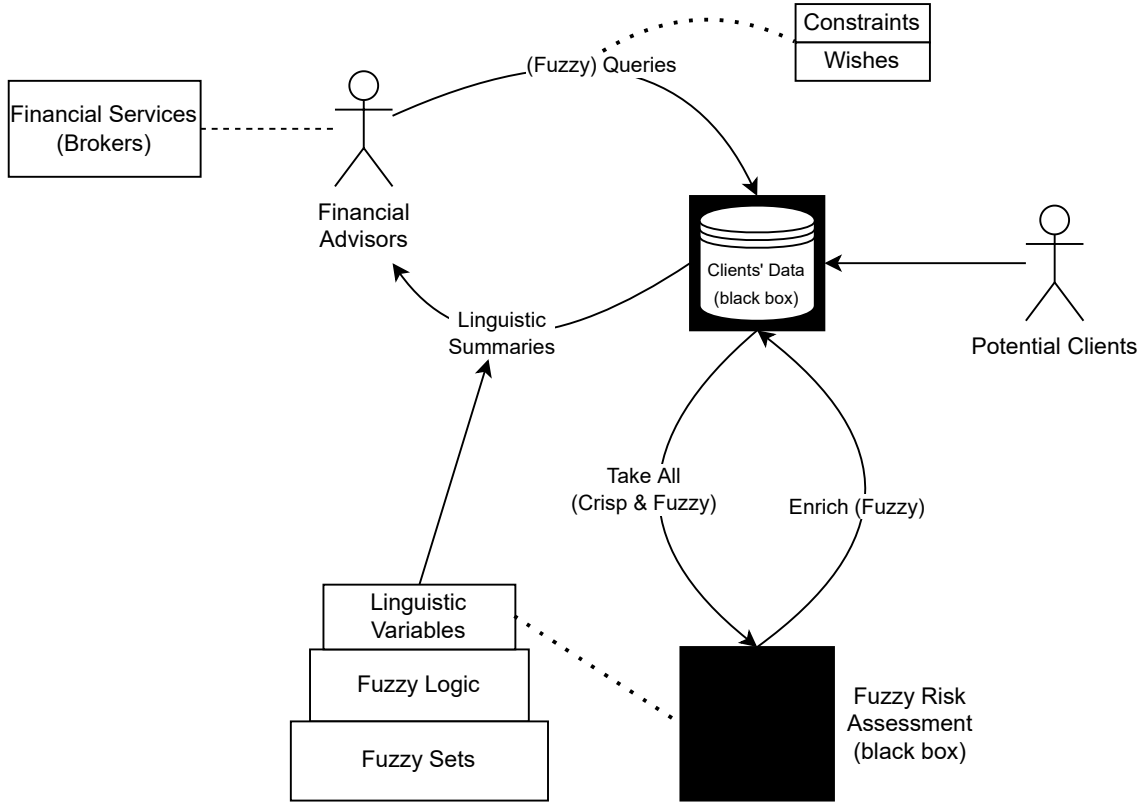


Figure 4: Conceptual Framework

3 Methodology

3.1 Research Approach

We choose a mix of quantitative and qualitative approaches. The validation of the fulfilment of our objectives demands the evaluation of the developed methods compared to the classical methods. The quantitative data would be the number of results in queries, the number of queries executed etc. The qualitative data would then come from the comparison of various aspects of our methods and the classical methods – readability, understandability etc.

3.2 Data Collection

The data for research will be generated from mock data. The only requirement for the data is for it to be resembling the real world. The dataset will represent a (potential) client database. Preferred attributes included will be age, marital status, employment, and income. In order to achieve our objective – querying fuzzy data – we will use the output from the prototype fuzzy inference system, which we outlined in Chrubasik (2022)

The generated data will take shape of natural language sentences and query statement results. The data will be produced by both our method and classical methods.

3.3 Sample

Kaggle Public Domain dataset will be used¹. The data will be enriched with a risk profile in the shape of fuzzy set suggesting asset allocation. The dataset is stripped of sensitive data, so ethical concerns are not raised.

In the case of real data, the legal landscape needs to be taken into account. Financial data becomes sensitive when attached to users' identities.

3.4 Data Analysis

The data will be stored in a relational database. A fitting schema will be designed to handle fuzzy data. Desired results will be interpreted in the natural language. Then, fuzzy queries will be applied to the data. The results will be summarised with linguistic summaries. Classical methods such as classical queries and descriptive statistics will be applied to the data for comparison.

References

- Bosc, P., Hadjali, A., & Pivert, O. (2008). Empty versus overabundant answers to flexible relational queries. *Fuzzy sets and systems*, 159(12), 1450–1467.

¹<https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets>

- Bosc, P., & Pivert, O. (1995). Sqlf: A relational database language for fuzzy querying. *IEEE transactions on Fuzzy Systems*, 3(1), 1–17.
- Bosc, P., & Pivert, O. (2000). Sqlf query functionality on top of a regular relational database management system. In *Knowledge management in fuzzy databases* (pp. 171–190). Springer.
- Chamberlin, D. D., & Boyce, R. F. (1974). Sequel: A structured english query language. *Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control*, 249–264.
- Chrubasik, V. (2022). *Assessing the risk profile of a retail investor with a focus on complementing retirement investments – a fuzzy logic approach*.
- Cipra, T. (2012). *Penze: Kvantitativní přístup*. Ekopress.
- Codd, E. F. (1971). Normalized data base structure: A brief tutorial. *Proceedings of the 1971 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control*, 1–17.
- Codd, E. F. (1972). Further normalization of the data base relational model. *Data base systems*, 6, 33–64.
- Codd, E. F. (2002). A relational model of data for large shared data banks. In *Software pioneers* (pp. 263–294). Springer.
- Date, C. J. (1989). *A guide to the sql standard*. Addison-Wesley Longman Publishing Co., Inc.
- Holzmann, R., Hinz, R. P., & Dorfman, M. (2008). Pension systems and reform conceptual framework. *World Bank Discussion Paper*, 824.
- Hudec, M. (2009). An approach to fuzzy database querying, analysis and realization. *Computer Science and Information Systems*, 6(2), 127–140.
- Hudec, M. (2014). Fuzzy data in traditional relational databases. *12th Symposium on Neural Network Applications in Electrical Engineering (NEUREL)*, 195–200.
- Hudec, M. (2016). Fuzziness in information systems. *Switzerland (CHE): Springer Nature*.
- Hudec, M., & Brokešová, Z. (2017). Mining and linguistically interpreting summaries from surveyed data related to financial literacy and behaviour. *International Conference on Data Management Technologies and Applications*, 67–83.
- Kacprzyk, J., & Zadrozny, S. (2001). Sqlf and fquery for access. *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*, 4, 2464–2469.
- Ross, T. J. (2005). *Fuzzy logic with engineering applications*. John Wiley & Sons.
- Tahani, V. (1977). A conceptual framework for fuzzy query processing—a step toward very intelligent database systems. *Information Processing & Management*, 13(5), 289–303.
- Vučetić, M., & Hudec, M. (2018). A fuzzy query engine for suggesting the products based on conformance and asymmetric conjunction. *Expert Systems with Applications*, 101, 143–158.
- Wang, T.-C., Lee, H.-D., & Chen, C.-M. (2007). Intelligent queries based on fuzzy set theory and sql. In *Information sciences 2007* (pp. 1426–1432). World Scientific.
- Yager, R. R. (1982). A new approach to the summarization of data. *Information Sciences*, 28(1), 69–86.
- Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on systems, Man, and Cybernetics*, (1), 28–44.
- Zadeh, L. A. (1975a). The concept of a linguistic variable and its application to approximate reasoning—ii. *Information sciences*, 8(4), 301–357.
- Zadeh, L. A. (1983). Linguistic variables, approximate reasoning and dispositions. *Medical Informatics*, 8(3), 173–186.
- Zadeh, L. A. (1996). Fuzzy sets. In *Fuzzy sets, fuzzy logic, and fuzzy systems: Selected papers by lotfi a zadeh* (pp. 394–432). World Scientific.
- Zadeh, L. A. (1999a). From computing with numbers to computing with words. from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on circuits and systems I: fundamental theory and applications*, 46(1), 105–119.
- Zadeh, L. A. (1999b). Fuzzy logic = computing with words. In *Computing with words in information/intelligent systems 1* (pp. 3–23). Springer.
- Zadeh, L. A. (1975b). The concept of a linguistic variable and its application to approximate reasoning—i. *Information sciences*, 8(3), 199–249.

Appendix C

Application Implementation

The application was implemented in Python3 (Harris et al., 2020) with the NumPy module for array/matrix computing (Harris et al., 2020). The visualizations were rendered using Matplotlib Python module (Hunter, 2007). The example implementation was implemented using Jupyter Notebooks (Kluyver et al., 2016). Code follows on the next page.


```

1 from fuzzy_inference.fuzzy_set import DEFAULT_RESOLUTION, FuzzySet
2 from fuzzy_inference.linguistic_variable import LinguisticVariable
3 from fuzzy_inference.rule import Rule
4 import numpy as np
5 import math
6 import matplotlib.pyplot as plt
7
8
9 class InferenceEngine:
10
11     def __init__(self) -> None:
12         self._inputvars: dict[str, LinguisticVariable] = {}
13         self._outputvars: dict[str, LinguisticVariable] = {}
14         self._rulebase: list[Rule] = []
15
16     @property
17     def rulebase(self):
18         return self._rulebase
19
20     @rulebase.setter
21     def rulebase(self, rulebase: list[Rule]) -> None:
22         self._rulebase = rulebase
23
24     @property
25     def inputvars(self) -> dict[str, LinguisticVariable]:
26         return self._inputvars
27
28     @inputvars.setter
29     def inputvars(self, inputvars: list[LinguisticVariable]) -> None:
30         self._inputvars = {l.name: l for l in inputvars}
31
32     @property
33     def outputvars(self) -> dict[str, LinguisticVariable]:
34         return self._outputvars
35
36     @outputvars.setter
37     def outputvars(self, outputvars: list[LinguisticVariable]) -> None:
38         self._outputvars = {l.name: l for l in outputvars}
39
40     def infer(self, measurements: dict):
41         for k, v in measurements.items():
42             self.inputvars[k].value = v
43
44         for rule in self._rulebase:
45             if rule.consequent is None:
46                 raise ValueError('consequent cannot be none at this stage.')
47             ceil: float = min([self.inputvars[ant[0]].fuzzified[ant[1]]
48                               for ant in rule.antecedents])
49             outvar: LinguisticVariable = self.outputvars[rule.consequent[0]]
50             # you do not actually need fuzzified, you need the initial fuzzy set
51
52             B = outvar.terms[rule.consequent[1]]
53             B_prime = FuzzySet.intersection(
54                 B, FuzzySet.uniform(ceil), (outvar.min, outvar.max))
55             outvar.b_prime[rule.consequent[1]] = B_prime
56
57     def output_fuzzy(self, resolution: int = DEFAULT_RESOLUTION) -> dict[str, FuzzySet]:
58         results = {}
59         for var_name, var in self.outputvars.items():
60             fuzz = FuzzySet.uniform(0)
61             for term_name, term in var.b_prime.items():
62                 fuzz = FuzzySet.union(
63                     fuzz, term, (var.min, var.max), resolution=DEFAULT_RESOLUTION)
64             # I need to perform the intesection operation on ALL the terms, not just one
65             results[var_name] = fuzz
66         return results
67
68     def defuzzify(self):
69         """uses center of maxima function to defuzzify output measures"""
70         results = {}
71         for var_name, var in self.outputvars.items():
72             fuzz = FuzzySet.uniform(0)
73             for term_name, term in var.b_prime.items():
74                 fuzz = FuzzySet.union(
75                     fuzz, term, (var.min, var.max), resolution=DEFAULT_RESOLUTION)
76             # I need to perform the intesection operation on ALL the terms, not just one
77             maximum = np.max(fuzz.vertices[:, 1])
78             maxima = np.where(fuzz.vertices[:, 1] == maximum)
79             max_left, max_right = maxima[0][0], maxima[0][-1]
80             x_max_index = (max_left + max_right)/2
81             lower_i = fuzz.vertices[math.floor(x_max_index)]
82             upper_i = fuzz.vertices[math.ceil(x_max_index)]
83             defuzz_x = (upper_i[0] + lower_i[0])/2
84             results[var_name] = defuzz_x
85         return results

```

inference_engine.py

```

1 from __future__ import annotations
2 from distutils.log import Log
3 from enum import Enum
4 from typing import Collection
5
6
7 class Rule:
8
9     def __init__(self) -> None:
10         self.antecedents: Collection[tuple[str, str]] = []
11         self.consequent: tuple[str, str] | None = None
12
13     def IF(self, antecedents: Collection[tuple[str, str]]) -> Rule:
14         self.antecedents = antecedents
15         return self
16
17     def THEN(self, consequent: tuple[str, str]) -> Rule:
18         self.consequent = consequent
19         return self

```

rule.py

```

1 from __future__ import annotations
2 from typing import Collection
3
4 from fuzzy_inference.fuzzy_set import FuzzySet
5
6 import matplotlib.pyplot as plt
7
8
9 class LinguisticVariable:
10     def __init__(self, name, terms: dict[str, FuzzySet], range: Collection = (0, 1), domain: str = 'x') -> None:
11         self.name = name
12         self.terms: dict[str, FuzzySet] = terms
13         self._a_prime: float | FuzzySet = 0
14         self.min, self.max = range
15         self.fuzzified: dict[str, float] = {} # for inputs
16         self.b_prime: dict[str, FuzzySet] = {k: FuzzySet.uniform(
17             0) for k, _ in self.terms.items()} # for fuzzy outputs
18         self.domain = domain
19
20     @property
21     def a_prime(self):
22         return self._a_prime
23
24     @a_prime.setter
25     def value(self, value: int | float | FuzzySet) -> None:
26         self._a_prime = value
27         if type(value) == float or type(value) == int:
28             self.fuzzified = self.fuzzify_crisp(self._a_prime) # type: ignore
29         elif type(value) == FuzzySet:
30             self.fuzzified = self.fuzzify_fuzzy(self._a_prime) # type: ignore
31         else:
32             raise ValueError(str(type(value)) + ' is not supported.')
33
34     def fuzzify_crisp(self, x: float) -> dict[str, float]:
35         fuz = {}
36         for k, v in self.terms.items():
37             fuz[k] = v.mu(x)
38         return fuz
39
40     def fuzzify_fuzzy(self, x: FuzzySet) -> dict[str, float]:
41         fuz = {}
42         for k, v in self.terms.items():
43             fuz[k] = FuzzySet.intersection(
44                 v, x, (self.min, self.max)).height
45         return fuz

```

linguistic_variable.py

```

1 from __future__ import annotations
2 from filecmp import DEFAULT_IGNORES
3 import math
4 import numpy as np
5
6 from typing import Collection
7
8
9 DEFAULT_RESOLUTION = 1000
10
11
12 class FuzzySet:
13     """
14     piecewise linear mu
15     helper methods construct fuzzy numbers
16     """
17
18     def __init__(self, vertices: Collection[Collection], height: float = 1, **kwargs) -> None:
19         self.vertices: np.ndarray = np.array(vertices)
20         self.height = height
21         self.kwargs = kwargs
22
23     def mu(self, x: float) -> float:
24         return np.interp([x], [t[0] for t in self.vertices], [t[1] for t in self.vertices])[0]
25
26     def mu_inv(self, mu: float) -> float:
27         return np.interp([mu], [t[1] for t in self.vertices], [t[0] for t in self.vertices])[0]
28
29     def discretize(self, range: tuple[float, float], resolution: int = DEFAULT_RESOLUTION):
30         return np.array([np.array([x, self.mu(x)]) for x in np.linspace(range[0], range[1], resolution)])
31
32     @staticmethod
33     def discrete(x: float) -> FuzzySet:
34         return FuzzySet(vertices=(
35             (x-.00000000000001, 0),
36             (x, 1),
37             (x+.00000000000001, 0)
38         ))
39
40     @staticmethod
41     def triangular(a, b, c) -> FuzzySet:
42         vertices = np.array([
43             (-np.inf, 0),
44             (a, 0),
45             (b, 1),
46             (c, 0),
47             (np.inf, 0)
48         ])
49         return FuzzySet(vertices)
50
51     @staticmethod
52     def trapezoidal(a, b, c, d) -> FuzzySet:
53         vertices = np.array([
54             (-np.inf, 0),
55             (a, 0),
56             (b, 1),
57             (c, 1),
58             (d, 0),
59             (np.inf, 0)
60         ])
61         return FuzzySet(vertices)
62
63     @staticmethod
64     def l_ramp(start, end) -> FuzzySet:
65         return FuzzySet(np.array([
66             (-np.inf, 1),
67             (start, 1),
68             (end, 0),
69             (np.inf, 0)
70         ]))
71
72     @staticmethod
73     def r_ramp(start, end) -> FuzzySet:
74         return FuzzySet(np.array([
75             (-np.inf, 0),
76             (start, 0),
77             (end, 1),
78             (np.inf, 1)
79         ]))

```

fuzzy_set.py – part 1

```

81 @staticmethod
82 def uniform(height: float) -> FuzzySet:
83     return FuzzySet(np.array([
84         (-np.inf, height),
85         (0, height),
86         (np.inf, height)
87     ]))
88
89 @staticmethod
90 def union(f1: FuzzySet, f2: FuzzySet, range: tuple[float, float], resolution: int = DEFAULT_RESOLUTION) ->
    FuzzySet:
91     x = np.linspace(range[0], range[1], resolution)
92     vertices = np.vstack((x, np.maximum(f1.discretize(range, resolution=resolution)[
93         :, 1], f2.discretize(range, resolution=resolution)[: , 1])))
94     f3 = FuzzySet(vertices)
95     f3.height = np.max(vertices[:, 1])
96     return f3
97
98 @staticmethod
99 def intersection(f1: FuzzySet, f2: FuzzySet, range: tuple[float, float], resolution: int = DEFAULT_RESOLUTION) ->
    FuzzySet:
100     x = np.linspace(range[0], range[1], resolution)
101     vertices = np.vstack((x, np.minimum(f1.discretize(range, resolution=resolution)[
102         :, 1], f2.discretize(range, resolution=resolution)[: , 1])))
103     f3 = FuzzySet(vertices)
104     f3.height = np.max(vertices[:, 1])
105     return f3
106
107 def constrain_range(self, range: tuple[float, float]):
108     arr = np.empty((len(self.vertices), 2), float)
109     for i, vertex in enumerate(self.vertices):
110         if vertex[0] == -np.inf:
111             arr[i] = np.array([range[0], vertex[1]])
112         elif vertex[0] == np.inf:
113             arr[i] = np.array([range[1], vertex[1]])
114         else:
115             arr[i] = vertex
116     return FuzzySet(arr)
117
118 def hedge(self, coeff: float, range: tuple[float, float] = (0, 1), resolution: int = DEFAULT_RESOLUTION) ->
    FuzzySet:
119     vertices = [np.array([v, self.mu(v)])
120                 for v in np.linspace(range[0], range[1], resolution)]
121     vertices = np.array([np.array([v[0], math.pow(v[1], coeff)])
122                         for v in vertices])
123     return FuzzySet(vertices)

```

fuzzy_set.py – part 2

```

1 from matplotlib.axes import Axes
2 from fuzzy_inference.fuzzy_set import FuzzySet
3 from fuzzy_inference.linguistic_variable import LinguisticVariable
4 import matplotlib.pyplot as plt
5 import numpy as np
6
7
8 def _term_name_x_pos(vertices: np.ndarray):
9     if len(vertices) == 4 + 2: # case trapezoidal
10         return (vertices[2][0] + vertices[3][0])/2
11     if len(vertices) == 3 + 2: # case triangular
12         return vertices[2][0]
13     if len(vertices) == 2 + 2: # case ramp
14         if vertices[0][1] > vertices[-1][1]: # case l_ramp
15             return (vertices[0][0] + vertices[1][0])/2
16         else: # case r_ramp
17             return (vertices[2][0] + vertices[3][0])/2
18
19
20 class LVVisualizer:
21
22     def __init__(self, lv: LinguisticVariable) -> None:
23         self.lv = lv
24
25     def visualize(self, ax: Axes):
26         for term_name, term in self.lv.terms.items():
27             vertices = term.constrain_range(
28                 (self.lv.min, self.lv.max)).vertices
29             ax.plot(vertices[:, 0], vertices[:, 1], c='k')
30             ax.text(_term_name_x_pos(vertices), 1.2,
31                 term_name, ha='center', va='center')
32             ax.set_ylim(0, 1.4)
33             ax.set_title(self.lv.name)
34             ax.set_xlabel(self.lv.domain)
35             ax.set_ylabel(r'$\mu$', rotation=0)
36         return ax
37
38
39 class FuzzySetVisualizer:
40     def __init__(self, f: FuzzySet, name: str = '', domain: str = 'x') -> None:
41         self.f = f
42         self.name = name
43         self.domain = domain
44
45     def visualize(self, ax: Axes, range: tuple[float, float] = (-np.inf, np.inf)):
46         vertices = self.f.constrain_range(
47             (range[0], range[1])).vertices
48         ax.plot(vertices[:, 0], vertices[:, 1], c='k')
49         if self.name != '':
50             ax.text(_term_name_x_pos(vertices), 1.2,
51                 self.name, ha='center', va='center')
52         ax.set_ylim(0, 1.4)
53         ax.set_xlabel(self.domain)
54         ax.set_ylabel(r'$\mu$', rotation=0)
55         return ax

```

matplotlib_visualizer.py – A Matplotlib helper for fuzzy sets and linguistic variables.

Appendix D

Application Notebook

Inference Notebook

July 6, 2022

```
[ ]: from fuzzy_inference.linguistic_variable import LinguisticVariable
from fuzzy_inference.fuzzy_set import FuzzySet
import matplotlib.pyplot as plt
from fuzzy_inference.matplotlib_visualizer import LVVizualizer, FuzzySetVizualizer
from fuzzy_inference.inference_engine import InferenceEngine
from fuzzy_inference.rule import Rule
```

0.1 Conscientiousness Inference

[illegible]

```

at_most_few = FuzzySet.l_ramp(3, 8)
number_of_late_payments = len(
    [x for x in obligation_adherence if not x])
validity = at_most_few.mu(number_of_late_payments)

# take into account questionnaire and hard data
late_payments = LinguisticVariable('No. late payments', {
    'At most few': FuzzySet.l_ramp(3, 8),
    'More than few': FuzzySet.r_ramp(3, 8)
}, (0, 20), 'No. late payments')
conscientiousness = LinguisticVariable('Conscientiousness', {
    'Low': FuzzySet.l_ramp(0.2, 0.3),
    'Moderate': FuzzySet.trapezoidal(0.2, 0.4, 0.6, 0.8),
    'High': FuzzySet.r_ramp(0.6, 0.8)
}, (0, 1), 'percentile')
neuroticism = LinguisticVariable('Neuroticism', {
    'Low': FuzzySet.l_ramp(0.2, 0.3),
    'Moderate': FuzzySet.trapezoidal(0.2, 0.4, 0.6, 0.8),
    'High': FuzzySet.r_ramp(0.6, 0.8)
}, (0, 1), 'percentile')

engine = InferenceEngine()
engine.inputvars = [late_payments, conscientiousness]
engine.outputvars = [conscientiousness]
engine.rulebase = [
    Rule().IF(
        (
            ('Conscientiousness', 'High'),
            ('No. late payments', 'At most few')
        )
    ).THEN(
        ('Conscientiousness', 'High')
    ),
    Rule().IF(
        (
            ('Conscientiousness', 'High'),
            ('No. late payments', 'More than few')
        )
    ).THEN(
        ('Conscientiousness', 'Low')
    ),
    Rule().IF(
        (
            ('Conscientiousness', 'Moderate'),
            ('No. late payments', 'More than few')
        )
    ).THEN(

```



```

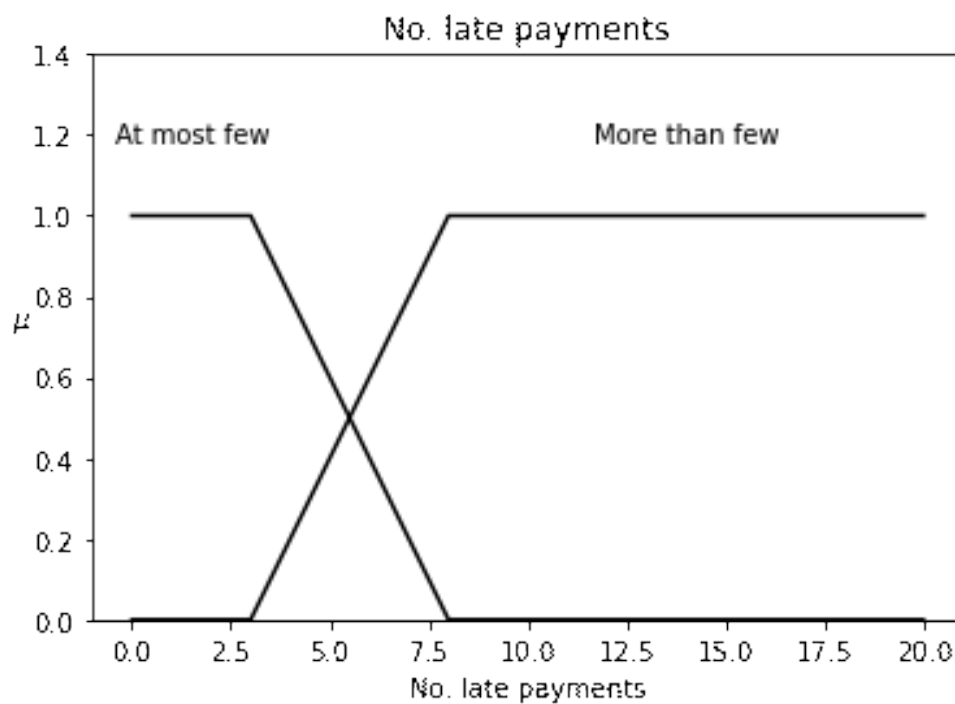
        ('Conscientiousness', 'Low')
    ),
    Rule().IF(
        (
            ('Conscientiousness', 'Low'),
            ('No. late payments', 'More than few')
        )
    ).THEN(
        ('Conscientiousness', 'Low')
    )
]

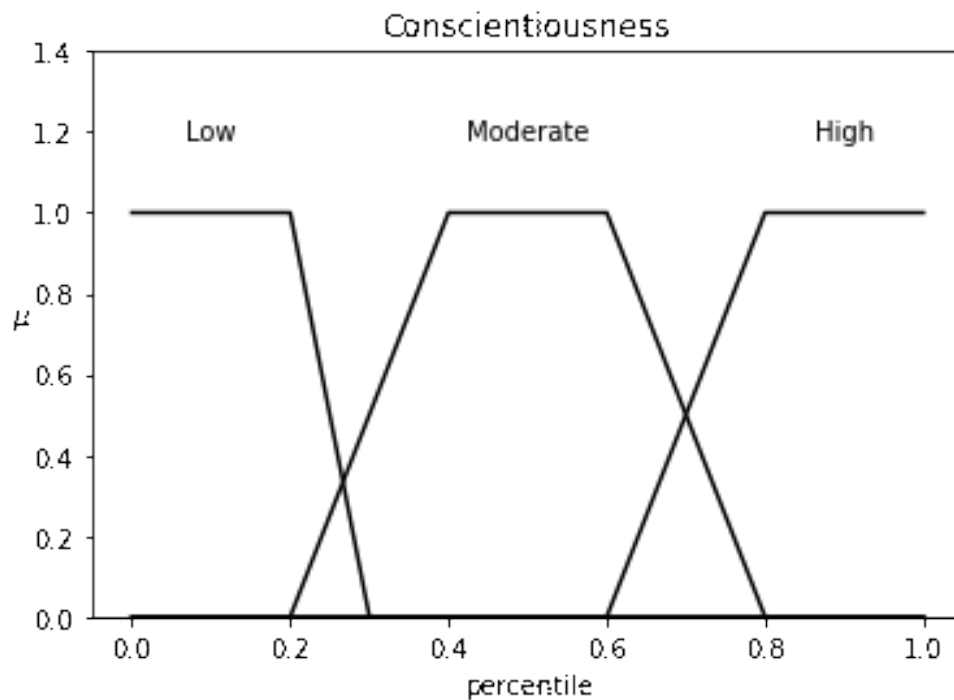
```

```

[ ]: lv_vizualizers = [LVVizualizer(lv) for _, lv in engine.inputvars.items()]
for viz in lv_vizualizers:
    fig, ax = plt.subplots()
    viz.vizualize(ax)

```





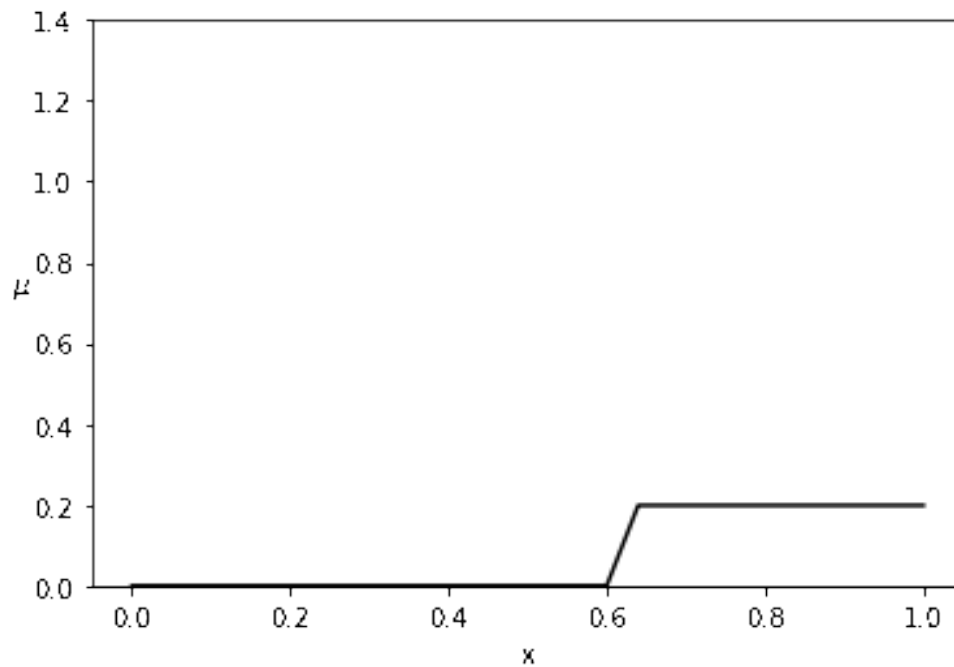
```
[ ]: engine.infer({
    'No. late payments': FuzzySet.triangular(7, 7, 8),
    'Conscientiousness': FuzzySet.triangular(0.8, 0.85, 0.9)
})
```

```
results = engine.defuzzify()
results
```

```
[ ]: {'Conscientiousness': 0.8203203203203203}
```

```
[ ]: measured_consc = engine.output_fuzzy()['Conscientiousness']
viz_consc = FuzzySetVizualizer(measured_consc)
fig, ax = plt.subplots()
viz_consc.vizualize(ax, (conscientiousness.min, conscientiousness.max))
```

```
[ ]: <AxesSubplot:xlabel='x', ylabel='$\mu$'>
```



```
[ ]: risk_capacity = LinguisticVariable('Risk Capacity', {
    'Low': FuzzySet.l_ramp(0.2, 0.3),
    'Moderate': FuzzySet.trapezoidal(0.2, 0.4, 0.6, 0.8),
    'High': FuzzySet.r_ramp(0.6, 0.8)
}, (0, 1), 'x')

risk_tolerance = LinguisticVariable('Risk Tolerance', {
    'Low': FuzzySet.l_ramp(0.2, 0.3),
    'Moderate': FuzzySet.trapezoidal(0.2, 0.4, 0.6, 0.8),
    'High': FuzzySet.r_ramp(0.6, 0.8)
}, (0, 1), 'x')

risk_requirement = LinguisticVariable('Risk Requirement', {
    'Low': FuzzySet.l_ramp(0.2, 0.3),
    'Moderate': FuzzySet.trapezoidal(0.2, 0.4, 0.6, 0.8),
    'High': FuzzySet.r_ramp(0.6, 0.8)
}, (0, 1), 'x')

risk_profile = LinguisticVariable('Risk Profile', {
    'Very conservative': FuzzySet.l_ramp(0.2, 0.3),
    'Conservative': FuzzySet.triangular(0.2, 0.3, 0.4),
    'Moderate': FuzzySet.trapezoidal(0.3, 0.4, 0.6, 0.7),
    'Aggressive': FuzzySet.triangular(0.6, 0.7, 0.8),
    'Very aggressive': FuzzySet.r_ramp(0.7, 0.8)
}, (0, 1), 'Proportion of risky assets.')
```

0.2 Risk Tolerance Inference

```
[ ]: rt_engine = InferenceEngine()
rt_engine.inputvars = [conscientiousness, neuroticism]
rt_engine.outputvars = [risk_tolerance]

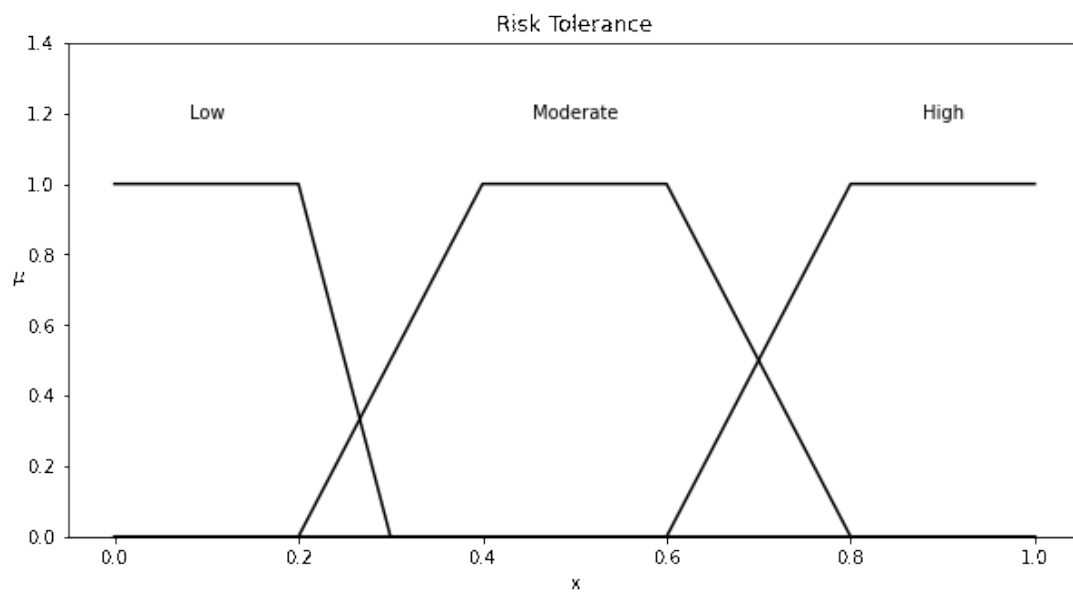
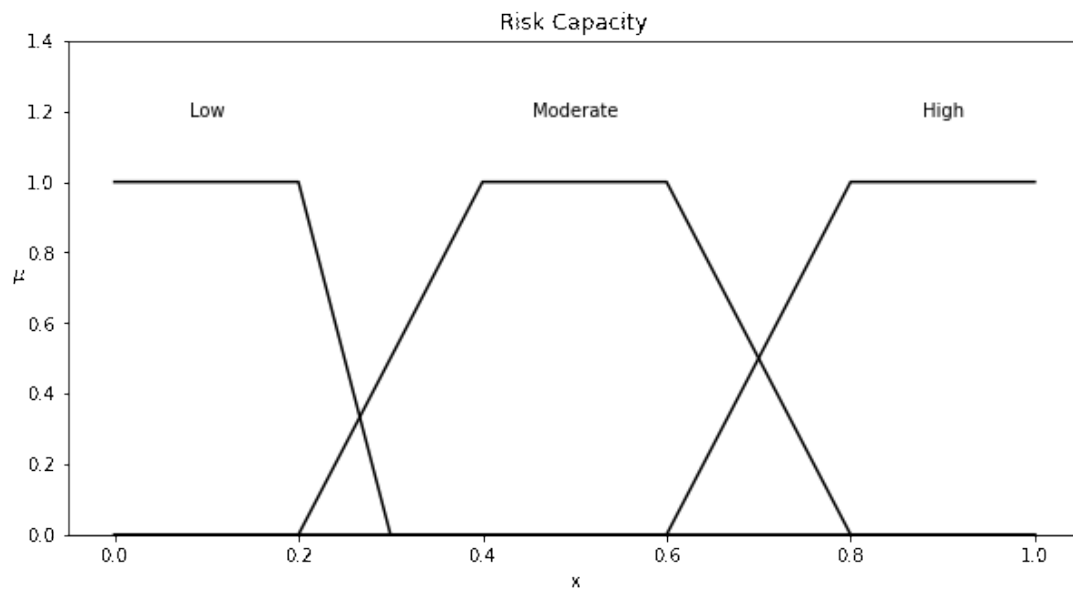
rt_engine.rulebase = [
    Rule().IF((
        ('Conscientiousness', 'Low'),
        ('Neuroticism', 'High'),
    )).THEN((
        ('Risk Tolerance', 'Low')
    )),
    Rule().IF((
        ('Conscientiousness', 'High'),
        ('Neuroticism', 'Low'),
    )).THEN((
        ('Risk Tolerance', 'High')
    )),
    Rule().IF((
        ('Conscientiousness', 'Moderate'),
        ('Neuroticism', 'Moderate'),
    )).THEN((
        ('Risk Tolerance', 'Moderate')
    )),
]

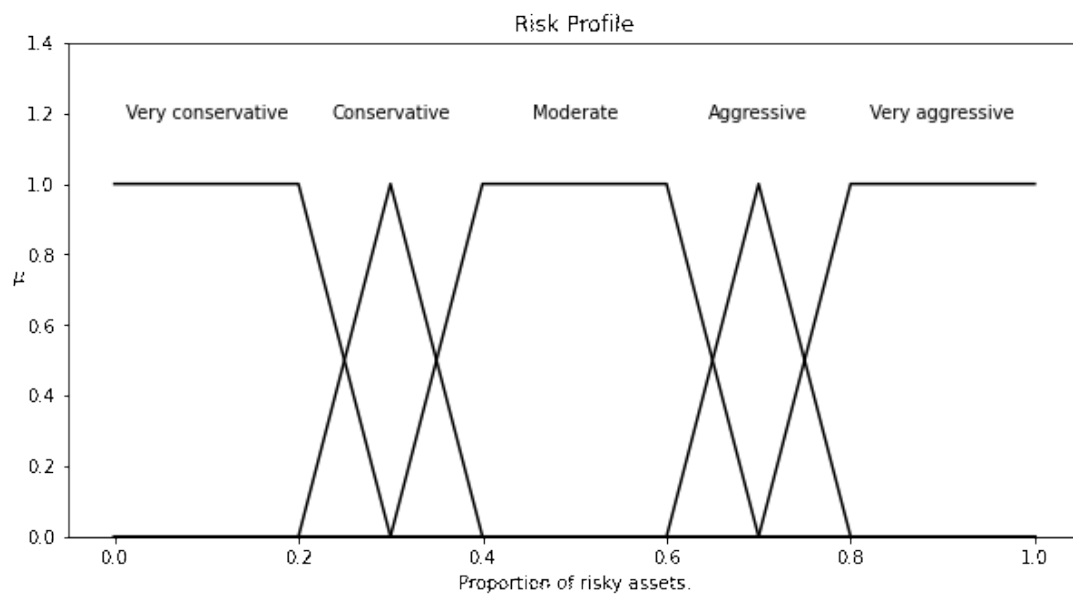
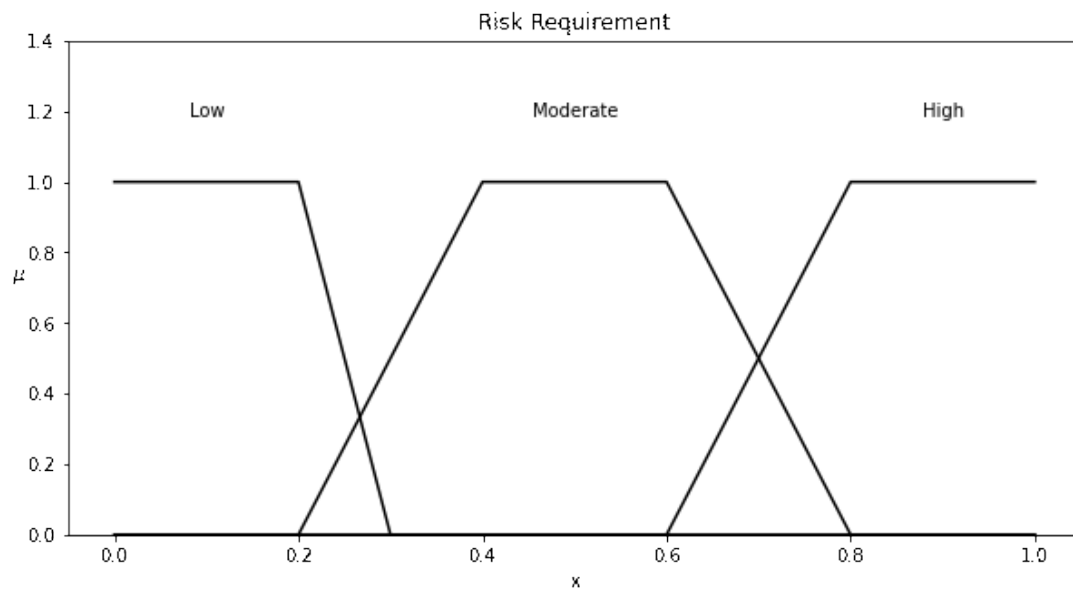
rt_engine.infer({
    'Conscientiousness': measured_consc,
    'Neuroticism': FuzzySet.triangular(0.6, 0.8, 0.9)
})

risk_tolerance_measured_fuzzy = rt_engine.output_fuzzy()['Risk Tolerance']
```

0.3 Risk Profile Inference

```
[ ]: lv_vizualizers = [risk_capacity, risk_tolerance, risk_requirement, risk_profile]
lv_vizualizers = [LVVizualizer(lv) for lv in lv_vizualizers]
for viz in lv_vizualizers:
    fig, ax = plt.subplots(figsize=(10, 5))
    viz.vizualize(ax)
```





```
[ ]: rp_engine = InferenceEngine()
rp_engine.inputvars = [risk_tolerance, risk_capacity, risk_requirement]
rp_engine.outputvars = [risk_profile]
rp_engine.rulebase = [
    Rule().IF((
        ('Risk Tolerance', 'Low'),
```

```

        ('Risk Capacity', 'Low'),
        ('Risk Requirement', 'Low'),
   )).THEN(
        ('Risk Profile', 'Very conservative')
    ),
    Rule().IF((
        ('Risk Tolerance', 'Low'),
        ('Risk Capacity', 'Low'),
        ('Risk Requirement', 'Moderate'),
   )).THEN(
        ('Risk Profile', 'Conservative')
    ),
    Rule().IF((
        ('Risk Tolerance', 'Low'),
        ('Risk Capacity', 'Moderate'),
        ('Risk Requirement', 'Low'),
   )).THEN(
        ('Risk Profile', 'Conservative')
    ),
    Rule().IF((
        ('Risk Tolerance', 'Moderate'),
        ('Risk Capacity', 'Low'),
        ('Risk Requirement', 'Low'),
   )).THEN(
        ('Risk Profile', 'Conservative')
    ),
    Rule().IF((
        ('Risk Tolerance', 'Moderate'),
        ('Risk Capacity', 'Moderate'),
        ('Risk Requirement', 'Moderate'),
   )).THEN(
        ('Risk Profile', 'Moderate')
    ),
    Rule().IF((
        ('Risk Tolerance', 'Moderate'),
        ('Risk Capacity', 'Moderate'),
        ('Risk Requirement', 'High'),
   )).THEN(
        ('Risk Profile', 'Aggressive')
    ),
    Rule().IF((
        ('Risk Tolerance', 'Moderate'),
        ('Risk Capacity', 'High'),
        ('Risk Requirement', 'Moderate'),
   )).THEN(
        ('Risk Profile', 'Aggressive')
    ),

```

```

Rule().IF((
    ('Risk Tolerance', 'High'),
    ('Risk Capacity', 'Moderate'),
    ('Risk Requirement', 'Moderate'),
)).THEN(
    ('Risk Profile', 'Aggressive')
),
Rule().IF((
    ('Risk Tolerance', 'High'),
    ('Risk Capacity', 'High'),
    ('Risk Requirement', 'High'),
)).THEN(
    ('Risk Profile', 'Aggressive')
),
]

very_low_hedge = FuzzySet.triangular(0, 0.1, 0.2).hedge(2)
somewhat_moderate = FuzzySet.triangular(0.3, 0.5, 0.4).hedge(0.5)
about_a_little_less_than_high = FuzzySet.triangular(0.6, 0.7, 0.8)

rp_engine.infer({
    'Risk Tolerance': risk_tolerance_measured_fuzzy,
    'Risk Capacity': very_low_hedge,
    'Risk Requirement': about_a_little_less_than_high,
})

results_rp = rp_engine.defuzzify()
results_rp

```

```
[ ]: {'Risk Profile': 0.5}
```