

Modellierung vager natürlichsprachlicher Quantoren über Dialogspiele und Fuzzy Logik

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

eingereicht von

Matthias F. J. Hofer, MSc

Matrikelnummer 01476075

an der Fakultät für Informatik
der Technischen Universität Wien

Betreuung: AO. Prof. Dr. Techn. Chris Fermüller
Zweitbetreuung: Dr. Techn. Magdalena Ortiz

Diese Dissertation haben begutachtet:

Thomas Vetterlein

Carles Noguera

Chris Fermüller

Wien, 15. Februar 2019

Matthias F. J. Hofer

Modeling vague Natural Language Quantifiers via Dialogue Games and Fuzzy Logic

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der Technischen Wissenschaften

by

Matthias F. J. Hofer, MSc

Registration Number 01476075

to the Faculty of Informatics

at the TU Wien

Advisor: AO. Prof. Dr. Techn. Chris Fermüller

Second advisor: Dr. Techn. Magdalena Ortiz

The dissertation has been reviewed by:

Thomas Vetterlein

Carles Noguera

Chris Fermüller

Vienna, 15th February, 2019

Matthias F. J. Hofer

Kurzfassung

In der natürlichen Sprache (NL) werden Quantoren oft benutzt um Sachverhalte, wie zum Beispiel “Viele Leute mögen Fussball” und “Ungefähr die Hälfte der Menschheit ist weiblich”, darzustellen. Die Ausdrücke *viele* und *wenige* bedeuten, dass eine gewisse Menge relativ gesehen groß, beziehungsweise klein ist. Die Semantik dieser zwei Quantoren ist nicht endgültig festgelegt, sondern hängt insbesondere von Kontextinformationen ab. In ähnlicher Weise verhalten sich Quantoren wie *ungefähr die Hälfte* und *fast alle*, insofern als dass die Toleranzen die, bezüglich entsprechender Statements, akzeptabel sind, von Situation zu Situation verschieden sein können.

Fuzzy Logik wird oft herangezogen um solche NL Ausdrücke zu modellieren, insbesondere zeitgenössische t-norm basierte Mathematische Fuzzy Logik (MFL). Hintikka hat die klassische Logik (CL), und Giles hat die Łukasiewicz Logik, eine MFL, spiel semantisch ausgedrückt. Die Gemeinsamkeit ist dabei ein Zweispieler-Nullsummenspiel mit perfekter Information, wobei die zwei Spieler strategisch vorgehen. Fermüller und Roschger haben Giles’s Spiel um einen dritten nicht-strategischen Spieler erweitert und damit eingeführt was wir das Zufallszeugenauswahlprinzip nennen.

Dieses eben genannte Prinzip erlaubt es uns zwei weitere MFLen, Gödel und Produkt Logik, spiel semantisch auszudrücken. Das erreichen wir mittels propositionaler Quantoren, welche es uns ermöglichen den Delta-Operator zu modellieren, welcher im Wesentlichen ein Projektionsoperator ist, der unstetige Wahrheitsfunktionen hervorruft. Diesen benötigen wir um die Gödel Implikation in Giles’s System auszudrücken. Darüber hinaus modellieren wir die Multiplikation und Division von Wahrheitsfunktionen mittels dem propositionalen Quantor der auf dem Zufallszeugenauswahlprinzip basiert, dem Delta-Operator und dem propositionalen Existenzquantor. Das ermöglicht es uns die Konnektive der Produkt Logik zu definieren. Auf diesen Resultaten aufbauend, zeigen wir wie man alle MFLen die auf endlichen Darstellungen basieren in unserem System definieren kann.

Aufbauend auf der erweiterten Ausdrucksstärke, modellieren wir zusätzlich eine Vielzahl von NL Quantoren in unserem System. Dieses Unterfangen betreiben wir schrittweise, so dass die ordentliche Interpretierbarkeit von entsprechenden Aussagen garantiert bleibt. Zunächst modellieren wir semi-fuzzy Quantoren, das sind diejenigen die nur klassische Argumente annehmen, das heißt Prädikate die entweder (vollständig) wahr oder falsch sind. Diese werden dann in systematischer Art und Weise mittels Quantifier Fuzzification Mechanisms (QFMs) zu fully-fuzzy Quantoren erweitert.

Als abschließenden Beitrag dieser Arbeit definieren und testen (mittels Implementierung)

wir eine Abfragesprache die Quantoren basierend auf dem Zufallszeugenauswahlprinzip beinhaltet. Die Resultate zeigen, dass probabilistische Auswertungen nicht nur geeignet sind für Modelle vager NL Quantoren, sondern auch die Auswertungsdauer von großen Datenmengen verringern können.

Abstract

In natural language (NL), quantifiers are often used to make statements about states of affairs, like “Many people like football”, and “About half the people are female”. In particular, *many* and *few* express that some set of objects is relatively big, or small respectively. The semantics of those two quantifiers is not fixed once and for all, but rather depends on contextual information. Likewise, quantifiers like *about half* and *almost all* show a comparable behavior, as the tolerance margins that make corresponding statements acceptable can change from one situation to another.

Fuzzy logic is often used to model such NL constructs, in particular contemporary t-norm based mathematical fuzzy logics (MFLs). Hintikka expressed Classical Logic (CL) game semantically, and Giles expressed Łukasiewicz logic (\mathbb{L}), a MFL, game semantically. The shared underpinning is a two player zero sum game of perfect information, where the two players act strategically. Fermüller and Roschger have augmented Giles’s game by a third non-strategic player, thereby introducing what we call the *random witness selection principle* into the framework.

The latter principle allows us to also express two other MFLs, Gödel logic and Product logic, game semantically. We achieve this by allowing for propositional quantification, which enables us to model the Delta operator, which is basically a projection operator, evoking discontinuous truth functions. This is needed to express Gödel implication in Giles’s framework. Moreover, the propositional quantifier based on the random witness selection principle, together with the Delta operator and the existential propositional quantifier, allows us to model multiplication and division of truth functions, which we need to define the connectives of Product logic. Building on this result, we show how to define all MFLs that are finitely representable in our framework.

Furthermore, the gained expressibility is used to model a variety of NL quantifiers within the framework. This pursuit is conducted in a step-by-step manner, that guarantees neat interpretability of statements. First, we model semi-fuzzy quantifiers, i.e. quantifiers that can only take classical arguments, i.e. predicates that evaluate to either (definitely) true or false. Then we lift those to fully-fuzzy quantifiers in a systematic and principle guided way, by means of quantifier fuzzification mechanisms (QFMs).

As a final contribution of this thesis, we define and test, by means of an implementation, a full-fledged query language, featuring quantifiers based on the random witness selection principle. The results show that probabilistic evaluations not only are suitable to model vagueness in NL, but also increase efficiency in presence of large amounts of data.

