INSTITUT FÜR INFORMATIK
DER LUDWIG–MAXIMILIANS–UNIVERSITÄT MÜNCHEN

Bachelorarbeit

Asymptotische Eigenschaften wohlgeformter Fragmente von Domain-size Aware Markov Logic Networks in Statistical Relational AI

Yiming Wang
Matrikelnummer: 11787686

Aufgabensteller: Prof. Dr. François Bry
Betreuer: Dr. Felix Weitkämper, DPhil (Oxon)
Abgabetermin: 23.09.2020

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Hilfsmittel verwendet habe.

München, den 23.09.2020 **Yiming Wang**

Würdigung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich bei der Erstellung dieser Bachelorarbeit unterstützt und motiviert haben.

Mein hauptsächlicher Dank gilt meinem Betreuer Prof. Dr. Bry für seine kritischen sowie hilfreichen Anregungen während des gesamten Betreuungszeitraums. Die Freiheit, die er mir bei der Wahl des Themas gelassen hat, war nicht selbstverständlich. Außerdem möchte ich mich bei ihm für seine Unterstützung und Inspiration im ganzen Bachelorstudium inklusive der Bewerbungsphase des Masterstudiums bedanken.

Ich bedanke mich nachdrücklich bei Herrn Dr. Weitkämper (DPhil Oxon) für die ausgezeichnete Betreuung sowie die fachliche Unterstützung. Vor allem danke ich ihm, dass er mich in den schwierigen Zeiten motiviert hat, und immer als großes Vorbild fungiert hat. Die freundliche und zuverlässige Zusammenarbeit hat mir eine praxisnahe Forschung ermöglicht und wertvolle Einblicke gewährt.

Ebenfalls möchte ich mich bei meinem Kommilitonen Michael Vogt bedanken, der mir mit viel Geduld, Interesse und Hilfsbereitschaft zur Seite stand. Bedanken möchte ich mich für die zahlreichen interessanten Debatten und Ideen, die maßgeblich dazu beigetragen haben, dass diese Bachelorarbeit in dieser Form vorliegt.

Letztlich richte ich auch ein Dankeschön an Deutschland für das gebührenfreie Studium und für die Möglichkeit des Studiengangwechsels. Das Land hat mir trotz meiner schwierigen persönlichen Situation Bildung ermöglicht, die mir mein Heimatland damals verwehrt hat.

Zusammenfassung

Wir betrachten Markov Logic Networks (Kürzel MLN) und Relational Logistic Regression (Kürzel RLR) als zwei grundlegende Repräsentationsformalismen in der statistischen relationalen künstlichen Intelligenz (Kürzel StarAI), um gewichtete Formeln zu spezifizieren. Markov Logic Networks basieren auf ungerichteten Graphen, während Relational Logistic Regression gerichtete azyklische Graphen darstellen.

Markov Logic Networks bekommen trotz ihrer funktionierenden Wahrscheinlichkeitseinschätzung in der statistischen relationalen künstliche Intelligenz oft die Kritik, dass bei Vergrößerung der Domänen die Wahrscheinlichkeiten gegen feste oder extreme Werte konvergieren können, wobei das vordefinierte Gewicht die Wahrscheinlichkeiten nicht mehr beeinflusst. Das untergräbt die intendierte Wichtigkeit der Formel, die durch das Gewicht repräsentiert werden soll.

Mithilfe der Domain-size Aware Relational Logistic Regression (Kürzel DA-RLR) wird in dieser Arbeit bewiesen, dass die asymptotischen Wahrscheinlichkeiten in einigen Fragmenten der Domain-size Aware Markov Logic Networks (Kürzel DA-MLN) gewichtsabhängig asymptotisch sind. Somit beeinflusst das vordefinierte Gewicht einer Formel immer die Wahrscheinlichkeit, auch wenn die Domänengöße gegen unendlich geht.

Schlüsselwörter: Domain-size Aware Markow Logic Networks · Asymptotische Eigenschaften · Relational Logistic Regression · Skalierung nach Domänengröße · Bayes'sche Netze · Statistische Relationale Künstliche Intelligenz (StarAI)

Inhaltsverzeichnis

1 Einführung . . . 6
1.1 Statistische Relationale Künstliche Intelligenz (StarAI) . . . 6
1.2 Motivation: Modellierung der COVID19 Epidemie . . . 6
1.3 Ziel der Arbeit . . . 7
1.4 Aktuelle Forschung zur Skalierbarkeit . . . 8
2 (Domain-size Aware) Markov Logic Networks und (Domain-size Aware) Relational Logistic Regression 9
2.1 Notation wichtiger Begriffe der Prädikatenlogik Erster Stufe . . . 9
2.2 Markov Logic Networks . . . 9
2.3 Domain-size Aware Markov Logic Networks . . . 10
2.4 Relational Logistic Regression . . . 11
2.5 Domain-size Aware Relational Logistic Regression . . . 11
2.6 Starkes Gesetz der Großen Zahlen . . . 12
3 Problem Asymptotischer Wahrscheinlichkeiten in DA-MLN . . . 13
3.1 Charakterisierung von Wahrscheinlichkeiten in MLN und DA-MLN . . . 13
3.2 Problem der Aggregationsfunktion in DA-MLN - Aggregat-Effekt . . . 14
4 Inspiration von Asymptotischen Wahrscheinlichkeiten in DA-RLR . . . 16
4.1 Grund der exakten gewichtsabhängigen Skalierung in (DA-)RLR . . . 16
4.2 Notwendige Beschränkung über Verbindungsvektor von DA-MLN . . . 18
4.3 Gegenbeispiel von DA-MLN ohne identische Einträge im Verbindungsvektor . . . 19
5 Beweis der Gewichtsabhängigen Konvergenz der Wahrscheinlichkeiten in DA-MLN mit $\wedge$-Quantor mithilfe vom Satz von Beppo Levi . . . 21
5.1 Falls $\omega > 0$ . . . 21
5.2 Falls $\omega < 0$ . . . 27
6 Beweis der Gewichtsabhängigen Konvergenz der Wahrscheinlichkeiten in DA-MLN mit $\vee$-Quantor mithilfe vom Fixpunktsatz von Banach . . . 32
6.1 Für alle $\omega \in \mathbb{R}$ . . . 32
6.2 Fixpunkt von S. Banach . . . 33
7 Anwendungsfälle für die gewichtsabhängigen Wahrscheinlichkeiten von DA-MLN . . . 36
7.1 Verbreitung von Krebs und Rauchen im Freundeskreis . . . 36
7.2 Hyperlinks und Klassifizierung von Uni-Webseiten . . . 38
7.3 Arbeitsbeziehungen in Filmindustrie . . . 39
8 Zusammenfassung und Einblicke in die zukünftige Forschung . . . 41
Literatur . . . 42

1 Einführung

1.1 Statistische Relationale Künstliche Intelligenz (StarAI)

In den vergangenen Jahren hat sich die statistische relationale künstliche Intelligenz zu einem vielversprechenden Ansatz entwickelt, da StarAI die Lernfähigkeit der klassischen symbolischen Künstlichen Intelligenz (Kürzel KI) mit der Anpassungsfähigkeit der modernen statistischen KI kombiniert. Allerdings ist noch nicht klar, wie sich StarAI unter verschiedenen Domänengrößen verhält. Dies ist besonders interessant für StarAI, da das Gewicht der statistischen relationalen Formalismen in vielen Anwendungsfällen dazu dienen sollte, die Wahrscheinlichkeit einer Formel widerzuspiegeln.

Darüber hinaus hat sich die Skalierbarkeit als ein wesentliches Hindernis für den breiten Einsatz der StarAI in verschiedenen Anwendungsfällen erwiesen. Das Thema der Skalierung über die Domänengröße hinweg hat daher einige Aufmerksamkeit erhalten. Einige allgemeine Verhaltensmuster sind bis jetzt schon bekannt.

2014 haben Poole et al. das Skalierungsverhalten sowohl von Markov Logic Networks (Kürzel MLN) als auch von Relational Logistic Regression (Kürzel RLR) in kleinen Klassen von Formeln charakterisiert, auf denen sich die Wahrscheinlichkeitsberechnung als asymptotisch unabhängig von den definierten oder gelernten Parametern erweisen. 2018 haben Jäger und Schulte einige notwendige Bedingungen genannt, unter denen die Domänengröße keine Rolle bei der Wahrscheinlichkeitsberechnung verschiedener StarAI-Ansätze spielt. 2019 haben Mittal et al. die Charakterisierung von Poole et al. mit erheblichem analytischen und numerischen Aufwand erweitert und teilweise korrigiert. Ein Vorschlag zur Abschwächung der Abhängigkeit der Domänengröße in MLN durch Skalierung der Gewichte wurde auch präsentiert. Die Skalierungsparameter sind mit den Formeln verbunden und entsprechen der Größe der Domänen. Der resultierende Formalismus wurde "Domain-size Aware Markov Logic Networks" (Kürzel DA-MLN) genannt. 2020 hat Weitkämper anhand einiger Beispiele gezeigt, dass das Problem der Aggregationsfunktion entsteht, wenn man ein einziges Gewicht für Formeln wählt, deren Literale unterschiedliche Anzahl von Verbindungen haben.

Eine systematische Untersuchung dieser Abhängigkeit in DA-MLN steht jedoch noch aus. In dieser Arbeit wird eine Charakterisierung von ein paar Beispielfällen von DA-MLN präsentiert, die zeigen, dass die asymptotischen Wahrscheinlichkeiten bestimmter Formelklassen in DA-MLN von den gelieferten Gewichten abhängig sind.

1.2 Motivation: Modellierung der COVID19 Epidemie

Eine Reihe von Anwendungsfällen in der KI beruhen auf der relationalen Struktur der Domänen. Die Markov-Logik-Netze (genaue Definition siehe Abschnitt 2.2) sind ein beliebtes Modell vom StarAI, das die zugrundeliegenden Domänen als eine Menge gewichteter Formeln darstellt und die Wahrscheinlichkeit solcher Formeln einschätzt.

Die Markov-Logik wurde erfolgreich auf eine Vielzahl von KI-Anwendungen angewandt, etwa auf Bereiche wie Computer Vision, NLP, Biologie und Robotik. [D.L09] Trotz der Anwendbarkeit der Markov-Logik bleibt eine bedeutende Herausforderung offen.

Denken Sie an das Problem der Epidemievorhersage in der COVID19-Krise in einer Stadt. Angesichts der Bevölkerung in einer Stadt stehen einige Informationen, ob jede Person krank ist oder nicht, der Forschung zur Verfügung. Wir möchten vorhersagen, ob es eine Epidemie gibt. Die Domäne wird durch eine einzige MLN-Formel modelliert $\omega : sick(x) \implies epidemic$, wobei ω das Gewicht der Formel ist, und die Wichtigkeit dieser Formel präsentiert. Das Gewicht ω kann von Trainingsdaten gelernt werden.

Abbildung 1 [Mittal et al. 2019][D.L09] zeigt, dass die Wahrscheinlichkeit einer Epidemie 0,75 beträgt, wenn die Domänengröße 3 ist. Hier beträgt das Gewicht ω den Wert 1.

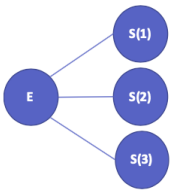

Abb. 1. Domänengröße $|D_x| = 3, \mathcal{P}(\text{Epidemie ist 100\% sicher}) \approx 0,75$, S:sick, E:Epidemie

Abbildung 1 [Mittal et al. 2019][D.L09] zeigt, dass die Wahrscheinlichkeit einer Epidemie schnell zu 1 tendiert, wenn die Domänengröße auf 15 zunimmt. Das Gewicht $\omega = 1$ ist unverändert. Aber tatsächlich gilt dies für jedes fixierte positive Gewicht ω. Diese Aussage ist komplett unrealistisch, dass die Verbreitung einer Epidemie eine 100% Wahrscheinlichkeit hat, wenn nur 15 Personen daran erkrankt sind, obwohl das Gewicht ω hier ziemlich klein ist.

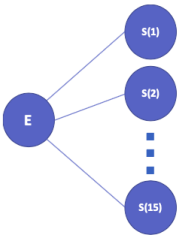

Abb. 2. Domänengröße $|D_x| = 15, \mathcal{P}(\text{Epidemie ist 100\% sicher}) \approx 1$, S:sick, E:Epidemie

Wenn wir beispielsweise COVID-19 mit anderen Epidemien vergleichen wollen, dann ist das Gewicht ω jeder Epidemie auch von höchster Bedeutung, da das Gewicht ω die Wichtigkeit einer Epidemie darstellt, z.B. wie schlimm die Symptomen sind, oder wie schwer die Epidemie zu verheilen ist.

Wie wir das Gewicht einer Epidemie ins Spiel bringen können, untersuchen wir im Rahmen dieser Bachelorarbeit.

1.3 Ziel der Arbeit

Wie wir im Abschnitt 1.2 besprochen haben, sind die gewichtsabhängigen asymptotischen Wahrscheinlichkeiten von Domain-size Aware Markov Logic Networks für viele Anwendungsgebiete notwendig, insbesondere wenn die Domänengröße gegen unendlich geht, z.B. bei der Modellierung der Epedimieverbreitung unter Millionen Bevölkerung.

Nach Untersuchung anderer statistischen relationalen Modellen stellen wir fest, dass Domain-size Aware Relational Logistic Regressions diese gewünschten gewichtsabhängigen asymptotischen Wahrscheinlichkeiten haben. Eine natürliche Überlegung wäre, ob wir Domain-size Aware Markov Logic Networks in Domain-size Aware Relational Logistic Regressions mit zusätzlichen Beschränkungen umwandeln können, um diese Eigenschaften zu erwerben. Daraus entsteht diese Arbeit mit vier Fragmenten von DA-MLN als erster Schritt dieser Untersuchung.

Die Anwendungsfälle dieser gewichtsabhängigen asymptotischen Wahrscheinlichkeiten von DA-MLN sind für uns von großem Interesse. Im letzten Teil dieser Arbeit schauen wir noch drei Anwendungsfälle aus der echten Welt an, um eine Verbindung zwischen theoretischen Forschungsfortschritten und praktischen Vorteilen zu erstellen.

1.4 Aktuelle Forschung zur Skalierbarkeit

Die Problematik der sich ändernden Domänengrößen wurde schon sehr früh bei der Entwicklung der StarAI bemerkt. Bereits 1998 diskutierte Jäger die Existenz asymptotischer Wahrscheinlichkeiten in relationalen Bayes'schen Netzen (ein weiterer Ansatz für gerichtete Netze) und stellte eine Verbindung zur Theorie endlicher Modelle und Theorie unendlicher Modelle her.

Jäger charakterisierte jedoch weder die Wahrscheinlichkeiten, zu denen relationale Bayes'sche Netze konvergieren, noch die Bedingungen, unter denen diese Wahrscheinlichkeiten von Gewichten abhängen. Dies wurde erstmals explizit als ein Problem isoliert, als Jain et al. 2010 die Adaptive Markov Logic Networks (AMLN) als einen Lösungsvorschlag präsentierten. Dies war der erste Vorschlag, die Gewichtung der Formeln eines MLN mit zunehmender Größe der Domäne zu variieren.

Weitkämper (2020) hat einen weiteren Ansatz, der als eine RLR-Übersetzung der AMLN betrachtet werden kann, aufgebracht. Er isolierte eine bestimmte Skalierungsfunktion und untersuchte ihre technischen Eigenschaften und ihr asymptotisches Verhalten eingehend. Außerdem schlug er vor, die Wahl des Skalierungsfaktors als ein semantisches Problem zu betrachten, und legte den Schwerpunkt mehr auf die Interpretation der Daten als die Daten selbst.

Eine detaillierte Analyse des Verhaltens der unkorrigierten Modelle wurde von Poole et al. (2014) mit Bezug auf MLN und den von Kasemi et al. (2014) eingeführten neuen Rahmen der RLR durchgeführt. Er unternahm eine Studie über das asymptotische Verhalten der MLN und bewies 0-1 Gesetze für eine bestimmte Klasse von MLN (korrigiert von Mittal et al. (2019)). Außerdem zeigten Jaeger und Schulte (2018), dass eine schmale Klasse von MLN (entsprechend den σ-determinierten MLN, in denen die unendlichen Modelle definiert sind (Singla und Domingos, 2007)) well-behaved sind und unter einer Zunahme der Domänengröße in der Tat invariant sind.

Mit großer Relevanz zu unserer Arbeit haben Mittal et al. (2019) die Klassifikation von Poole et al. (2014) verfeinert und eine mutmaßliche Lösung für das Problem der Wahrscheinlichkeiten mit ändernden Dömanengrößen vorgeschlagen. In ihrem Modell des DA-MLN, das in Abschnitt 2.2 formell vorgestellt wird, ändern sich die Gewichte einer Formel explizit nach der Anzahl der Verbindungen zwischen Literalen. Sie zeigten, dass für bestimmte Kombinationen von MLN und Abfragen die asymptotischen Wahrscheinlichkeiten unregelmäßig von den Gewichten abhängen, wenn sie als DA-MLN betrachtet werden. Allerdings wenn sie als einfache MLN betrachtet werden, gibt diese Abhängigkeiten von Gewichten nicht mehr.

In dieser Arbeit wird gezeigt, dass mithilfe von gerichteten Modellen wie RLR die aus unserer Sicht größte technische Schwäche des DA-MLN, nämlich den Aggregat-Effekt bei bestimmten DA-MLN Fragmenten überwinden wird. Es wird gezeigt, dass bei Anwendung auf den RLR-Ansatz die Skalierung der Gewichtung ein gewichtsabhängiges Verhalten bei zunehmender Domänengröße garantiert.

2 (Domain-size Aware) Markov Logic Networks und (Domain-size Aware) Relational Logistic Regression

In diesem Abschnitt werden wir kurz die Syntax und Semantik von (Domain-size Aware) Markov Logic Network und von (Domain-size Aware) Relational Logistic Regression vorstellen. Beide Ansätze kombinieren statistische mit logischen Informationen, und beide verwenden dazu Gewichte. (DA-)MLN basieren jedoch auf ungerichteten grafischen Modellen (Markov Netze), während die RLR auf gerichteten graphischen Modellen (Bayes'sche Netze) basieren. Eine aufschlussreiche Diskussion über diese Unterscheidung und ihre Bedeutung in der Statistischen Relationalen KI findet sich in der Arbeit von de Raedt et al., (2016).

2.1 Notation wichtiger Begriffe der Prädikatenlogik Erster Stufe

Die Signatur einer Theorie der Logik erster Ordnung (First Order logic, in Kürze FOL) besteht aus Konstanten-, Variablen-, Funktions- und Prädikatensymbole. Wir werden eine echte Untermenge der FOL verwenden.

Ein Term ist eine Konstante oder eine logische Variable. Wir werden die logischen Variablen mit kleinen englischen Buchstaben bezeichnen.

Die Domäne einer logischen Variablen x wird durch D_x bezeichnet.

Ein Prädikat erster Ordnung nimmt Terme als seine Argumente und definiert eine Boolesche Beziehung zwischen den Termen.

Ein propositionales Prädikat ist ein Prädikat, das keine Argumente benötigt.

Ein Atom ist ein Prädikatensymbol, das auf ein Tupel von Termen angewandt wird.

Ein Literal ist ein Atom oder seine Negation.

Formeln werden wie folgt rekursiv definiert: (a) Ein Literal ist eine Formel, (b) Die Negation einer mit f bezeichneten Formel $\neg f$ ist ebenfalls eine Formel, (c) Wenn f_1 und f_2 zwei Formeln sind, dann sind $(f_1 \wedge f_2)$ und $(f_1 \vee f_2)$ ebenfalls Formeln.

Eine Belegung (auf englisch Grounding) eines Atoms ist ein Atom, dessen Terme alle belegt sind.

Eine Belegung einer Formel ist eine Formel, deren Atome alle belegt sind.

Im weiteren Verlauf des Aufsatzes werden wir eine logische Variable einfach als Variable bezeichnen. Außerdem werden wir mit typisierten Theorien arbeiten, bei denen eine Untergruppe von Variablen zum gleichen Typ gehören kann, in welchem Fall ihre Domänen gemeinsam genutzt werden. Wir werden $\mathfrak{X}$ verwenden, um die Menge aller Variablen zu bezeichnen, die in den Formeln einer MLN vorkommen.

2.2 Markov Logic Networks

Da Markov Logic Networks in der Literatur bereits ausgiebig diskutiert wurden, werden wir diesen Unterabschnitt hauptsächlich dazu verwenden, einige Notationen einzuführen. Eine ausführlichere Diskussion findet sich im Originalbeitrag von Richardson und Domingos (2006) oder in de Raedt et al. (2016). Daher werden wir auch die Definition von MLN auf einen Rahmen beschränken, der groß genug ist, um alle Beispiele in dieser Arbeit und in den in Abschnitt 1.4 erwähnten Arbeiten aufzunehmen.

Definition 1. *Sei $\mathcal{L}$ eine (potenziell) multi-sortierte relationale Signatur. Ein Markov Logic Network T über $\mathcal{L}$ ist eine Sammlung von Paaren (φ, w) wo φ eine quantorenfreie $\mathcal{L}$-Formel ist mit $w \in \mathbb{R}$. Wir nennen w das Gewicht von φ in T.*

Markov Logic Networks wurden von Richardson und Domingos (2006) eingeführt und haben seither auf dem Gebiet der StarAI großen Einfluss ausgeübt. Ihre Semantik basiert auf ungerichteten Netzen, was bedeutet, dass jedes Literal in einer Formel jedes andere auf dynamische Art und Weise beeinflussen kann. Wir können eine solche Semantik für ein MLN T erhalten, indem wir eine (endliche) Domäne für jede Art von $\mathcal{L}$ wählen. Wir definieren die Semantik als Wahrscheinlichkeitsverteilung über alle $\mathcal{L}$-Strukturen auf den gegebenen Domänen wie folgt:

Definition 2. *Gegeben einer Auswahl von Domänen für $\mathcal{L}$, definiert ein MLN T über $\mathcal{L}$ eine Wahrscheinlichkeitsverteilung auf den möglichen $\mathcal{L}$-Strukturen auf den gewählten Domänen wie folgt: Sei $\mathfrak{X}$ eine $\mathcal{L}$-Struktur auf den gegebenen Domänen. Dann ist:*

$$\mathcal{P}_{T,D}(X = \mathfrak{X}) = \frac{1}{Z}\exp\left(\sum_i w_i n_i(\mathfrak{X})\right)$$

wo $n_i(\mathfrak{X})$ die Zahl der wahren Belegungen von φ_i in $\mathfrak{X}$ ist, w_i das Gewicht von φ_i ist und Z eine Normalisierungskonstante ist, um sicherzustellen, dass sich alle Wahrscheinlichkeiten auf 1 summieren. Da die Wahrscheinlichkeiten nur von den Domänengrößen abhängen, können wir auch schreiben $\mathcal{P}_{T,n}$ für Domänen der Größe n. Wir nennen weiter $w_T := \sum_i w_i n_i(\mathfrak{X})$ das Gewicht von $\mathfrak{X}$.

2.3 Domain-size Aware Markov Logic Networks

Mittal et al. (2019) haben Gewichtsskalierungen bei MLN eingeführt, um die Effekte variierender Domänengrößen auszugleichen. Sie nennen den sich daraus ergebenden Formalismus Domain-size Aware Markov Logic Networks (DA-MLN) und wir werden hier die wichtigsten Definitionen aus ihrer Arbeit wiederholen.

Definition 3. *Ein Domain-size Aware Markov Logic Network (DA-MLN) ist durch die gleichen Daten wie ein MLN gegeben (siehe Definition 1).*

Um die Semantik an die sich ändernde Domänengröße anzupassen, verwenden Mittal et al. (2019) das Konzept eines Verbindungsvektors.

Definition 4. *Sei φ eine Formel. Sei Ψ die Menge der Literale von φ und sei weiter für jedes $\psi \in \Psi$ V_ψ die Menge der freien Variablen in φ, die nicht in ψ vorkommen. Sei für jede Variable x die zugehörige Domäne D_x. Dann ist der Verbindungsvektor von φ die Menge*

$$\left\{ \prod_{x \in V_\phi} |D_x| \;\middle|\; \phi \in \Psi \right\}$$

Der Verbindungsvektor hält fest, mit wie vielen Tupeln sich jedes Literal verbinden kann. In der Formel $P(x) \wedge Q(x,y) \wedge R(z)$ zum Beispiel kann sich das Literal $P(x)$ mit $|D_y| \cdot |D_z|$ Tupeln verbinden. Das Literal $Q(x,y)$ kann sich mit $|D_z|$, das Literal $R(z)$ mit $|D_x| \cdot |D_z|$ verbinden. Das Problem ist jetzt, diese Information zu einem einzigen Skalierungsfaktor zu aggregieren. Mittal et al. (2019) nutzen das Maximum der Einträge des Verbindungsvektors, schlagen aber auch vor andere Optionen zu evaluieren.

Definition 5. *Gegeben eine Wahl von Domänen für $\mathcal{L}$, definiert ein DA-MLN T über $\mathcal{L}$ eine Wahrscheinlichkeitsverteilung auf den möglichen $\mathcal{L}$-Strukturen auf den gewählten Domänen wie folgt: Sei $\mathfrak{X}$ eine $\mathcal{L}$-Struktur auf den gegebenen Domänen. Dann ist:*

$$\mathcal{P}_{T,D}(X = \mathfrak{X}) = \frac{1}{Z}\exp\left(\sum_i \frac{w_i n_i(\mathfrak{X})}{C_{i,D}}\right)$$

wo $n_i(\mathfrak{X})$ die Zahl der wahren Belegungen von φ_i in $\mathfrak{X}$ ist, w_i das Gewicht von φ_i, $C_{i,D}$ das Maximum der Einträge des Verbindungsvektors von φ_i, und Z eine Normalisierungskonstante, um sicherzustellen, dass sich alle Wahrscheinlichkeiten auf 1 summieren.

Mittal et al. (2019) geben einige Beispiele an, die illustrieren, wie der Wechsel von einem MLN zu einem DA-MLN das asymptotische Verhalten verändert.

2.4 Relational Logistic Regression

Relational Logistic Regression (RLR) unterscheidet sich von MLN insofern, als dass es auf gerichteten statt ungerichteten Modellen basiert. Anstatt beliebige gewichtete Formeln zu erlauben, benennt man zunächst ein Relational Belief Network (RBN). Dies ist ein gerichteter azyklischer Graph dessen Knoten mit Atomen aus $\mathcal{L}$ beschriftet sind. Ein RLR-Modell über diesem RBN besteht aus einer zusätzlichen Beschriftung jedes Knotens wie folgt:

Definition 6. *Eine Relational Logistic Regression T über einer relationalen Signatur $\mathcal{L}$ besteht aus einem gerichteten azyklischen Graph G dessen Knoten O beschriftet sind mit einem Paar $(\varphi, (V, \psi, w)_i)$, wobei φ ein $\mathcal{L}$-Atom ist und $(V, \psi, w)_i$ eine Menge von Tripeln, sodass V_i eine endliche Menge von Variablensymbolen ist, ψ_i eine quantorenfreie Formel und $w_i \in \mathbb{R}$. Weiter wird Folgendes verlangt:*

1. *Jedes $\mathcal{L}$-Atom ist die Beschriftung genau eines Knotens*
2. *Keines der Variablensymbole aus φ ist in V_i*
3. *ψ_i ist eine boolesche Kombination der ersten Beschriftungen der Elternknoten von O, und jede Variable in ψ_i erscheint entweder in φ oder ist in V_i. Wenn O ein Blatt ist, dann gibt es nur ein (1) Tripel mit $\psi ='' x = x''$ für eine Variable x.*

Während die Semantik eines MLN in Definition 2 durch eine einzige Wahrscheinlichkeitsverteilung gegeben war, gab Weitkämper (2020) die Semantik eines RLR durch Induktion über den zugrundeliegenden RBN an. Genauer verwendete er Induktion über den längsten Pfad von einem Blatt zu einem gegebenen Knoten, was er den Index eines Knotens nannte. Um die Schreibweise zu vereinfachen, fügte er einige Notationen ein bevor er die Semantik eines RLR definierte:

Definition 7. *Die* sigmoid*-Funktion ist definiert durch* $\mathrm{sigmoid}(k) := \frac{\exp(k)}{\exp(k)+1}$. *Wir schreiben weiter $\vec{a} \in D$ für ein Tupel aus D (statt $\vec{a} \in D^n$) und nutzen den Ausdruck 1_ψ für die Funktion die 1 zurückliefert falls ψ erfüllt ist und 0 sonst. Zu guter Letzt verwenden wir die Konvention, für die Belegung von ψ Die sigmoid-Funktion ist definiert durch* $\mathrm{sigmoid}(k) := \frac{\exp(k)}{\exp(k)+1}$. *Wir schreiben weiter $\vec{a} \in D$ für ein Tupel aus D (statt $\vec{a} \in D^n$) und nutzen den Ausdruck 1_ψ für die Funktion die 1 zurückliefert falls ψ erfüllt ist und 0 sonst. Zu guter Letzt verwenden wir die Konvention, für die Belegung von ψ, die der Variable $\vec{a}$ den Wert $\vec{x}$ zuweist, $\psi(\vec{a}/\vec{x})$ zu schreiben. Für eine kurze und informelle Einführung in MLN und RLR and ihre Beziehung zueinander, siehe Kaezemi et al. (2014b).*

2.5 Domain-size Aware Relational Logistic Regression

Nun werden wir das Prinzip des DA-MLN an den gewichteten gerichteten Modellansatz der RLR anpassen. Weitkämper (2020) zeigte, dass er, da es nur ein Literal gibt, das direkt durch eine beliebige Verbindung (das Kinderliteral der Kante) beeinflusst wird, das Problem der Aggregation eines Verbindungsvektors vermeiden konnte und direkt durch die Domänengrößen der freien Variablen in der entsprechenden Variablenmenge skalieren konnte. Formal definierte er:

Definition 8. *Ein DA-RLR T ist durch die gleichen Daten gegeben wie eine RLR (siehe Definition 6). Die Semantik unterscheidet sich jedoch wie folgt mit der Definition 6:*
Sei D eine mehrfach sortierte Domäne, und sei D_x der Definitionsbereich einer Variable x. Für jede Menge von Variablensymbolen V lasse $|D|_V := \Pi_{x \in V} |D_x|$.

In der Induktion ergibt sich die Wahrscheinlichkeit einer $\mathcal{L}_{n+1}$-Struktur aus der Wahrscheinlichkeit ihres $\mathcal{L}_n$-Reduktionsfaktors multipliziert mit der Wahrscheinlichkeit der Variablenbelegung der Atome in $\frac{\mathcal{L}_{n+1}}{\mathcal{L}_n}$. Diese sind gegeben durch

$$\mathcal{P}(q(\vec{x})) = \text{sigmoid}\left(\sum_i \frac{\omega_i}{|D|_{V_i}} \sum_{\vec{a} \in D} 1_{\psi_i(\vec{a}/V_i)}\right)$$

Die Definition hängt nun explizit von den Domänengrößen ab, so wie ein DA-MLN von seinen Domänengrößen abhängt. Allerdings stellt die Zahl $|D|_{V_i}$ im Nenner nun die möglichen Verbindungen nur für das Atom $q(\vec{x})$ dar - da dies das einzige Kind dieses Verbindungsgraphen ist, gibt es keinen Verbindungsvektor und keine Aggregationsfunktion. Für diese Darstellung bewies er, dass jeder Clusterpunkt der Wahrscheinlichkeit einer beliebigen Formel mit den gewählten Gewichten variiert, wenn sich die Domänengrößen unendlich nähern. Darüber hinaus kann das Grenzverhalten in Form eines speziellen Graphen, nämlich die Relational Belief Networks beschreiben, die wir Proportional Relational Belief Network (PRBN) nennen und in Abschnitt 4.1 genauer besprechen werden.

2.6 Starkes Gesetz der Großen Zahlen

Bevor wir zu konkreten Berechnungen übergehen, werden wir kurz ein Lemma aufzeichnen, das wir an mehreren Stellen benötigen werden, nämlich das starke Gesetz der großen Zahlen.

Das starke Gesetz der großen Zahlen besagt, dass sich die relative Häufigkeit eines Zufallsergebnisses in der Regel um die theoretische Wahrscheinlichkeit eines Zufallsergebnisses zu 100% stabilisiert, wenn das zugrundeliegende Zufallsexperiment immer wieder unter denselben Voraussetzungen durchgeführt wird.

Lemma 1. *[Bau96, Kapitel III]*

Ist eine Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen gegeben, so sagt man, dass diese Folge dem starken Gesetz der großen Zahlen genügt, wenn der Mittelwert der skalierten Zufallsvariablen $\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathrm{E}(X_i))$ fast sicher gegen 0 konvergiert. Das bedeutet, dass $P(\lim_{n \to \infty} \overline{X}_n = 0) = 1$ ist.

Definition 9. *Fast Sichere Konvergenz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und (M, d) ein separabler, metrischer Raum (wie zum Beispiel $\mathbb{R}^n$) versehen mit der Borelschen σ-Algebra $\mathcal{B}(M)$ sowie X, X_n Zufallsvariablen von $(\Omega, \mathcal{A}, P)$ nach $(M, \mathcal{B}(M))$. Die Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ konvergiert dann fast sicher oder P-fast sicher gegen X, wenn eine Menge $N \in \mathcal{A}$ existiert mit $P(N) = 0$ und $\lim_{n \to \infty} d(X(\omega), X_n(\omega)) = 0$ für alle $\omega \in \Omega \setminus N$. Man schreibt dann auch $X_n \xrightarrow{f.s.} X$, $X_n \xrightarrow{P\text{-}f.s.} X$ oder $X_n \to X$ P-f.s. Alternativ findet sich für reelle Zufallsvariablen auch die Formulierung, dass die Zufallsvariablen genau dann fast sicher konvergieren, wenn $P\left(\{\omega \in \Omega \colon \lim_{n \to \infty} X_n(\omega) = X(\omega)\}\right) = 1$ ist.*

3 Problem Asymptotischer Wahrscheinlichkeiten in DA-MLN

Wir haben in den vorhergehenden Abschnitten gesehen, dass DA-MLN die Abhängigkeit des Grenzverhaltens von den Gewichten einer Formel verstärken. In diesem Abschnitt bauen wir auf der Arbeit von Mittal et al. (2019) auf und stellen einige Semantik von Weitkämper(2020) vor, die in den späteren Beweisen verwendet wird. Um diesen Ableitungen eine klare Struktur zu geben, werden wir zunächst Wahrscheinlichkeitskerne für die Markov Logic Networks vorstellen und einige grundlegende Fakten über ihr Grenzverhalten nachweisen.

3.1 Charakterisierung von Wahrscheinlichkeiten in MLN und DA-MLN

Um die Wahrscheinlichkeit zu bestimmen, dass eine bestimmte atomare Formel wahr ist, berechnen wir das Wahrscheinlichkeitsmaß der Menge aller Welten, in denen diese Formel gilt. Dies kann auch bequem als Integral geschrieben werden

$$\mathcal{P}(R(\vec{x})) = \int_{\mu T,n} 1_{(R\vec{x})}$$

Wir möchten die Indikatorfunktion in Form von Integral durch eine Funktion ersetzen, die wir analytisch in Form von Gewichten des MLN-Modells ausdrücken können. Das können wir erreichen, indem wir den gewichteten Mittelwert der Indikatorfunktionen zwischen zwei Modellen betrachten, die sich nur im Wert von $R(\vec{x})$ unterscheiden.

Seien $\mathfrak{X}$ eine Struktur und $\vec{x} \in \mathfrak{X}$ ein Tupel. Dann sei $\mathfrak{X}_{R(\vec{x})}$ die Struktur, die sich potenziell von $\mathfrak{X}$ nur in dem Fall, dass $R(\vec{x})$ in $\mathfrak{X}_{R(\vec{x})}$ gilt, unabhängig davon, ob es in $\mathfrak{X}$ gilt. $\mathfrak{X}_{\neg R(\vec{x})}$ ist analog definiert. Auf diese Art und Weise teilt sich die Klasse aller Strukturen einer gegebenen Domänengröße zu gleichen Teilen in Strukturen der Form $\mathfrak{X}_{R(\vec{x})}$ und Strukturen der Form $\mathfrak{X}_{\neg R(\vec{x})}$, und es besteht eine natürliche Eins-zu-Eins-Entsprechung zwischen Strukturen jedes Typs. Daher können wir das obige Integral berechnen, indem wir stattdessen das gewichtete Mittel der Indikatorfunktion sowohl für $\mathfrak{X}_{R(\vec{x})}$ als auch für $\mathfrak{X}_{\neg R(\vec{x})}$ berücksichtigen. Dies lässt sich in Form einer Hilfsfunktion beschreiben, die die Abhängigkeit der Gewichte vom Wahrheitswert von $R(\vec{x})$ widerspiegelt:

Definition 10. *[Wei20, S.4] Sei $\mathfrak{X}$ eine Struktur der Domänengröße n und sei T ein MLN mit dem Gewicht ω. Dann ergibt sich*

$$\delta^{T_\omega,n}_{R(\vec{x})} := \sum_i \omega_i n_i \left(\mathfrak{X}_{R(\vec{x})}\right) - \sum_i \omega_i n_i \left(\mathfrak{X}_{\neg R(\vec{x})}\right)$$

Deshalb können wir die Wahrscheinlichkeit von $R(\vec{x})$ als ein Integral über δ ausdrücken, das direkt mit den Gewichten im MLN verbunden ist

$$\mathcal{P}(R(\vec{x})) = \int_{\mu T_\omega,n} \text{sigmoid}\left(\delta^{T_\omega,n}_{R(\vec{x}))}\right)$$

Wir werden das Prinzip des MLN an den Modellansatz des DA-MLN anpassen, und wir werden sehen, dass wir, da es endlich viele Verbindungen zwischen den Literalen gibt, das Gewicht einer Formel entsprechend durch den Verbindungsvektor anpassen müssen.

Definition 11. *Ein DA-MLN T ist durch die gleichen Daten gegeben wie ein MLN (siehe Definition 9). Sei D eine mehrfach sortierte Domäne, x eine Variable, und sei D_v der Verbindungsvektor für alle Variablen einer Formel. Für jede Menge von Variablensymbolen V sei $|D|_V := \pi_{x\in V}|D_v|$.*

Dann ergibt sich aus der letzten Definition über MLN:

$$\delta^{T_\omega,n}_{R(\vec{x})} := \sum_i \frac{\omega_i}{|D|_{V_i}} n_i \left(\mathfrak{X}_{R(\vec{x})}\right) - \sum_i \frac{\omega_i}{|D|_{V_i}} n_i \left(\mathfrak{X}_{\neg R(\vec{x})}\right)$$

Die Wahrscheinlichkeit von $R(\vec{x})$ lässt sich genau so wie in der Definition 3 ausdrücken.

Die Definition 4 hängt nun explizit von den Domänengrößen des DA-MLN ab. Allerdings stellt die Zahl $|D|_{V_i}$ im Nenner nun das Maximum der Verbindungsanzahl ausgehend von allen Literalen in der Formel dar. Da viele Literale weniger ausgehende Kanten oder Verbindungen haben, gibt es den sogenannten Aggregat-Effekt.

3.2 Problem der Aggregationsfunktion in DA-MLN - Aggregat-Effekt

In diesem Abschnitt besprechen wir einige Formeln als Beispiele, wo MLN die Anzahl der Verbindungen eines Literals, nämlich $Q(x)$, "überkompensiert". Das nennt man auch Aggregat-Effekt. Somit ist die asymptotische Wahrscheinlichkeit dieses MLN mit unendlicher Domänengrößen nicht gewichtsabhängig.

Proposition 1. *[Wei20, S.6] Sei T_ω das MLN, das durch die Formel $P \wedge Q(x) \wedge R(x,y)$ gegeben ist: $\omega, \omega > 0, |D_x| = |D_y| = n$. Dann ist die asymptotische Wahrscheinlichkeit von $Q(x)$ 1, unabhängig vom Gewicht ω.*

Proposition 2. *[Wei20, S.6] Sei T_ω das DA-MLN, das durch die Formel $P \wedge Q(x) \wedge R(x,y)$ gegeben ist: $\omega, \omega > 0, |D_x| = |D_y| = n$. Dann ist die asymptotische Wahrscheinlichkeit von $Q(x)$ $\frac{1}{2}$, unabhängig vom Gewicht ω.*

(Genaue Berechnung siehe Weitkämper (2020) Seite 6.)

Der Unterschied im Vergleich zur Proposition 1 kommt daher, dass das DA-MLN in der Proposition 2 das Maximum n^2 vom Verbindungsvektor $(n^2, n, 1)$ als Skalierungsfaktor hat. Somit werden die Einflussgrößen von Verbindungen einiger Literale durch das Teilen des Skalierungsfaktors "überkompensiert". Daraus entsteht der Aggregat-Effekt von den nicht exakt ausgeglichenen Verbindungen.

Mittal et al. (2019) verwendeten bei der Einführung der DA-MLN die Aggregationsfunktion *max* als pragmatische Wahl mit guten formalen Eigenschaften, die sich auch in der Praxis bewährt hat. Sie wiesen jedoch auch darauf hin, dass die Untersuchung von anderen Wahlmöglichkeiten für weitere Forschung hochinteressant sein könnte, und erwähnten dabei ausdrücklich die Funktion *sum*. Daher möchten wir kurz erörtern, inwieweit die in diesem Abschnitt aufgeworfenen Fragen von der genauen verwendeten Aggregationsfunktion abhängen.

Es ist klar, dass der Grad der Skalierungsfaktor wichtiger als die genaue Zahl (siehe Proposition 1 und 2) ist, da das asymptotische Verhalten in den Fällen degeneriert ist, in denen die Grenzen entweder auf 0 oder 1 liegen. Wenn die Maximumfunktion als Aggregationsfunktion verwendet wird, ist der Grad gleich dem höchsten Grad in allen Einträgen des Verbindungsvektors. Diese Skalierung ist jedoch ungeeignet, um das Verhalten der anderen Literale in der Formel zu untersuchen, da die Skalierung die Anzahl der Verbindungen "überkompensiert". Genau dies geschieht in den hier besprochenen Beispielen 1 und 2. Daher würde eine Änderung der Aggregationsfunktion auf *sum* die Probleme eher verschärfen als abschwächen, da sie dann noch weiter herunterskalieren würde.

Stattdessen mag es plausibel erscheinen, einen Begriff des Mittelwerts zu verwenden, wie Weitkämper (2020) erwähnte. Da es mit einer multiplikativen Skalierung zu tun haben, wäre der arithmetische Mittelwert keine natürliche Wahl, da der arithmetische Mittelwert im gleichen Grad wie das Maximum liegen würde.

Stattdessen könnte man den geometrischen Mittelwert als eine Aggregationsfunktion versuchen, laut Weitkämper (2020). Dies wird das asymptotische Verhalten der Literale an den Extremen nicht regulieren - tatsächlich, wenn man das Standardbeispiel von $R(x)$ betrachtet. Ein Beispiel zur Unterstützung dieser Aussage nannte

Weitkämper (2020): $R(x) \to P : \omega$. Der geometrische Mittelwert als Skalierungsfaktor würde die Verbindungen des $R(x)$ "überkompensieren" und die Verbindungen des P "unterkompensieren". Es ist theoretisch immer noch nicht optimal, und da der Grad der Verbindungsanzahl der Literale unterschiedlich ist. Die Konvergenz zu den degenerierten Wahrscheinlichkeiten würde jedoch für alle Literale verlangsamt, und dies könnte für praktische Anwendungen relevanter sein.

Wahrscheinlich würden wir nach weiteren Forschungen feststellen, dass keine einzelne Aggregationsfunktion für alle Literale einer Formel immer optimal sein könnte. Für die zukünftige Forschung wäre aber trotzdem interessant, die Beweis- und Berechnungsergebnisse aus dieser Arbeit mit solchen Aggregationsfunktionen zu probieren.

Wir werden nun entlang einer anderen Richtung fortfahren und versuchen herauszufinden, welche Bedingungen den DA-MLN eine exakte Skalierung ermöglichen können. Eine exakte Skalierung ist grundsätzlich schwer bei DA-MLN, da sie in Form von ungerichteten Graphen dargestellt werden können, bei denen sich alle Literale durch Verbindungen gegenseitig beeinflussen. Es ist anders bei (DA-)RLR, da es gerichtetes Modell ist, in dem eine Verbindung nur ein einziges Literal beeinflusst und man ein Literal direkt entlang der möglichen Verbindungen dieses Literals skalieren kann.

4 Inspiration von Asymptotischen Wahrscheinlichkeiten in DA-RLR

Wir haben in den vorhergehenden Abschnitten gesehen, dass DA-MLN mit Verbindungsvektor die Gewichtsabhängigkeit des Grenzverhaltens der (atomaren) Formeln trotz des negativen Einflusses des Aggregat-Effekts verstärken.

Genauer gesagt haben wir gesehen, dass das Problem daher rührt, dass man ein einzelnes Gewicht für Formeln wählen muss, deren Literale unterschiedliche Anzahl an Verbindungen haben. Die Einführung des Verbindungsvektors schwächt diesen Aggregat-Effekt ab, aber beseitigt ihn nicht vollständig. Wir haben auch gesehen, dass MLN ein gewichtsunabhängiges Grenzverhalten haben können.

Das (DA-)RLR-Modell, welches durch gerichtete Graphen dargestellt werden kann, hat uns zu einer möglichen Lösung inspiriert, um den Aggregat-Effekt in bestimmten DA-MLN vollständig zu beseitigen. Wir schauen uns zuerst an, warum (DA-)RLR eine exakte gewichtsabhängige Skalierung haben.

4.1 Grund der exakten gewichtsabhängigen Skalierung in (DA-)RLR

Die gängige Methode um einen gerichteten Graphen darzustellen ist das sogenannte Bayes'sche Netz, ein gängiger Begriff im Bereich der Logik und Informatik. Allerdings ist die Darstellung als Bayes'sches Netz technisch sehr aufwendig, wenn es um Schleifen im Relational Belief Network (RBN) geht. Daher gehen wir ab diesem Abschnitt davon aus, dass das RBN keine solchen Schleifen hat. Pearl (1988) hat den Begriff Polytree für ein solches Netz geprägt, da Schleifen bei der Bewertung der Belief-Verbreitung in allgemeinen Bayes'schen Netzen zusätzliche Schwierigkeiten bereiten.

Bei der Konstruktion einer geeigneten graphischen Darstellung für das Grenzverhalten der DA-RLR müssen wir daher die Berechnung der Wahrscheinlichkeiten in Bayes'schen Netzen anpassen, um den oben genannten Unterschied zu verdeutlichen. Dazu wird der Begriff Proportional Relational Belief Network (PBRN) verwendet.

Definition 12. *Ein Polytree wird als Proportional Relational Belief Network (PRBN) über eine Sprache $\mathcal{L}$ bezeichnet, wenn es sich um ein Relational Belief Network über die Sprache $\mathcal{L}_P$ handelt, das aus der Signatur von $\mathcal{L}_P$ besteht. Es gibt genügend viele generische Konstanten für jede Arität in jeder Domäne, die in einem Relationssymbol in $\mathcal{L}_P$ vorkommt, mit folgenden Eigenschaften:*

Wenn $R(\vec{x}, y)$ ein Atom von $\mathcal{L}_P$ ist und y_0 eine geeignete generische Konstante ist, dann erhält man die Eltern des Knotens $R(\vec{x}, y)$, indem y in allen Eltern von $R(\vec{x}, y)$ durch y_0 ersetzt wird.

Anmerkung 1. Die Eigenschaft in Definition 12 bestimmt eindeutig ein PRBN, die ein gegebenes RBN über die Sprache $\mathcal{L}$ erweitert.

Betrachten wir ein Beispiel mit der Formel $P \wedge Q(x) \wedge R(x, y)$.

Beispiel 1. Es gibt verschiedene Konfigurationen zur Darstellung von $P \wedge Q(x) \wedge R(x, y)$ als (DA-)RLR. Da wir an dem Verhalten von $Q(x)$ interessiert sind, betrachten wir dies als die abhängige Formel, sodass wir die Formel $P \wedge R(x, y) : \omega$ für $Q(x)$ und dieses RBN erhalten:

Abb. 3. RBN zur Darstellung der RLR mit Betracht auf $Q(x)$

Bei der Erstellung des zugehörigen PRBN werden an jeder Stelle Substitutionen von generischen Konstanten vorgenommen. $Q(x_0)$ wird Eltern P und $R(x_0, y)$ haben:

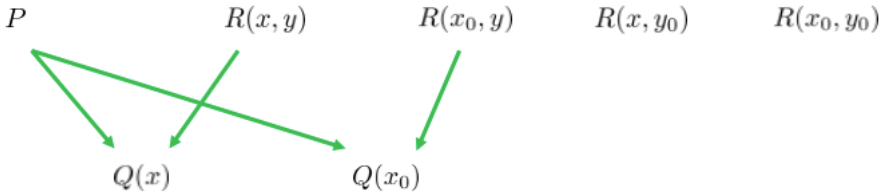

Abb. 4. PRBN zur Darstellung der RLR mit Betracht auf $Q(x)$

Anmerkung 2. Das resultierende PRBN ist ein Polytree, genau dann, wenn das RBN ein Polytree ist.

In den zwei obigen Graphen repräsentieren die grünen Pfeile jeweils die Anzahl an Verbindungen zwischen zwei Literalen. Ein Pfeil geht jeweils abwärts von einem Literal, und aufwärts zu einem anderen Literal.

Zur Berechnung der Wahrscheinlichkeiten in (DA-)RLR skaliert man immer gemäß die Anzahl von Verbindungen, die auf der anderen Seite des Pfeils liegen, also auf der Quelle des Pfeils. Da die Beeinflussung zwischen zwei Literalen in einem (DA-)RLR immer deutlich durch einen Pfeil mit Aufwärtsrichtung markiert ist, ist sie eindeutig einseitig. Diese Einseitigkeit des Einflusses gewährleistet eine exakte, und unabhängige Skalierung jedes Literals in einer Formel, da jedes Literal durch seine eigene Verbindungsanzahl herunterskaliert werden kann, unabhängig von anderen Literalen in der Formel.

Das sehen wir auch an der Definition 8, wo die Verbindungsanzahl eines Variablensymbols immer einzeln als $|D|_V := \Pi_{x \in V} |D_x|$ definiert ist, unabhängig von anderen Literalen in der Formel. Die gewichtsabhängige Wahrscheinlichkeit ist gegeben durch:

$$\mathcal{P}(q(\vec{x})) = \text{sigmoid}\left(\sum_i \frac{\omega_i}{|D|_{V_i}} \sum_{\vec{a} \in D} 1_{\psi_i(\vec{a}/V_i)}\right)$$

wobei die Verbindungsanzahl $|D|_{V_i}$, die dem Domänengrößenprodukt entspricht, im Nenner innerhalb der ersten Summe steht, und für jedes andere Literal variieren kann. Die Zahl $|D|_{V_i}$ im Nenner stellt nun die möglichen Verbindungen nur für das Literal $Q(\vec{x})$ dar - da dies das einzige Kind dieses Verbindungsgraphen ist, gibt es keinen Verbindungsvektor und keine Aggregationsfunktion.

Wir werden DA-MLN nun in Form von ungerichteten Graphen darstellen, bei denen sich alle Literale gegenseitig beeinflussen. Danach analysieren wir den Skalierungsunterschied zwischen (DA-)RLR und DA-MLN, und wie wir diesen Unterschied durch zusätzliche Beschränkungen über den Verbindungsvektor minimieren, oder sogar komplett beseitigen können.

4.2 Notwendige Beschränkung über Verbindungsvektor von DA-MLN

Der Unterschied in DA-MLN im Vergleich zu DA-RLR scheint zu sein, dass sich der Einfluss von Verbindungen in gerichteten Graphen nur in eine Richtung auswirkt, während dies in ungerichteten Graphen in beide Richtungen geht. Das heißt, dass der Einfluss zwischen zwei verbundenen Literalen dynamisch und gegenseitig ist.

Da wir in DA-MLN für die Skalierung einer gesamten Formel nur eine Zahl nehmen dürfen, die laut Mittal et al. (2019) der maximale Eintrag im Verbindungsvektor sein sollte, gibt keine exakte Kompensation für ein Literal einer Verbindung, wenn die beiden Literale der Verbindung unterschiedliche Verbindungsanzahl (Einträge im Verbindungsvektor) haben.

Das Problem gibt es nicht bei DA-RLR, da die Pfeilrichtung aussagt, in welcher Richtung man skaliert. Das heißt, dass man nicht beide Literale zugleich herunterskalieren muss, sondern nur in der Richtung des Pfeils ein Literal.

Die Intuition ist, wenn man ein DA-MLN hat, und es skalieren muss, und wenn die Einträge im Verbindungsvektor nicht dieselben sind, dann kann man irgendein Literal nicht exakt herunterskalieren, weil man sich für einen Eintrag im Verbindungsvektor entscheiden muss.

Ein intuitiver Gedanke zur exakten Skalierung wäre, die DA-MLN mit zusätzlichen Beschränkungen zu gestalten, damit es zwischen zwei verbunden Literalen einer Formel ein- und ausgehend gleich viele Verbindungen gibt. Genauer gesagt muss das DA-MLN, das eine exakte Skalierung in der Wahrscheinlichkeitsberechnung hat, einen Verbindungsvektor mit identischen Einträgen haben. Nur so kann die Skalierung in beide Richtungen exakt sein.

Proposition 3. *Sei T eine DA-MLN über einer relationalen Signatur $\mathcal{L}$ mit einer einzigen Formel $q(\vec{x}) : \omega$, wobei ω das Gewicht der Formel ist. Wenn der Verbindungsvektor $\vec{V}$ aus gleichen Einträgen mit Wert n besteht, wobei n nach der Definition von Verbindungsvektor ein Domönengrößenprodukt ist und von den einzelnen Variablendomänengrößen abhängen soll, ist die optimale Wahl für die Aggregationsfunktion n. Diese Aggregationsfunktion ermöglicht eine exakte Skalierung für alle Literale in der Formel.*

Beweis. Die Eigenschaft, dass der Verbindungsvektor $\vec{V}$ aus gleichen Einträgen mit Wert n besteht, bedeutet, dass jedes Literal gleich viele ein- und ausgehende Verbindungen hat. Natürlich ist die Aggregationsfunktion gleich zur konstanten Funktion n, egal ob man diese nach dem *max*-, nach dem *min*-Prinzip oder nach dem geometrischen Mittelwert auswählt. (Im Abschnitt 3.2 haben wir bereits besprochen, warum es nicht nach dem *sum*-Prinzip oder nach dem arithmetischen Mittelwert ausgewählt werden soll. Weitere Wahlkriterien müssen im Einzelfall gesondert besprochen werden.)

Die Tatsache, dass die Aggregationsfunktion genau der Verbindungsanzahl jedes Literals entspricht, gewährleistet eine exakte Skalierung, wobei jedes Literal durch das richtige Domänengrößenprodukt herunterskaliert wird. Man kann es sich so vorstellen, dass eine Einflussgröße entlang einer Verbindung auf- und abwärts geht, und beide verbundenen Literale sind jeweils gleich stark davon betroffen (die Einflussgröße kann man sich übrigens vorstellen wie die Stärke des Einflusses auf ein Literal). Die zugrundeliegende Aggregationsfunktion bildet in dem Sinne eine Einflussgröße auf beide Literale einer Verbindung, die beide Literale exakt herunterskalieren kann.

Anhand eines Beispiels kann man das Gleichgewicht der gegenseitigen Beeinflussung besser veranschaulichen.

Sei $\mathcal{T}$ ein DA-MLN mit einer einzelnen Formel $P(x, y) \wedge Q(y, z) \wedge R(x, z) : \omega$, wobei ω das Gewicht für die Formel ist. Für die Domänengröße gilt $|D_z| = |D_x| = |D_y| = n(\in \mathbb{N})$. Dieses DA-MLN erfüllt die Anforderung dieser Proposition, da der Verbindungsvektor $\vec{V} = (|D_z|, |D_x|, |D_y|) = (n, n, n)$ ist.

Es gibt verschiedene Konfigurationen zur Darstellung dieses DA-MLN mit der Formel $P(x,y) \wedge Q(y,z) \wedge R(x,z)$. Hier wird angenommen, dass man sich für das Verhalten vom $Q(y,z)$ interessiert. Daher betrachten wir dieses Literal als unabhängige atomare Formel.

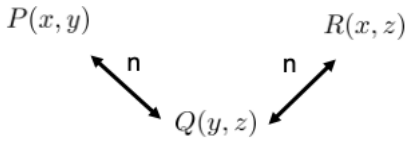

Abb. 5. Ungerichteter Graph des DA-MLN mit der Formel $P(x,y) \wedge Q(y,z) \wedge R(x,z)$

Anhand dieses Graphen sehen wir, dass $P(x,y)$ und $R(x,z)$ unabhängig sind, wenn die Pfeile von $P(x,y)$ und $R(x,z)$ nach $Q(y,z)$ gehen. Wenn aber die Pfeile von $Q(y,z)$ wiederum nach $P(x,y)$ und $R(x,z)$ gehen, sind $P(x,y)$ und $R(x,z)$ nicht mehr unabhängig. Wenn z.B. $P(x,y)$ eine höhere Wahrscheinlichkeit hat, wird $Q(y,z)$ entsprechend wahrscheinlicher wahr, somit wird auch $R(x,z)$ mit entsprechend größerer Wahrscheinlichkeit wahr. In dieser Hinsicht unterscheidet sich dieses DA-MLN von dem DA-RLR unter gleichen Bedingungen.

Aber da $P(x,y)$, $R(x,z)$ und $Q(y,z)$ gleich viele Verbindungen zwischeneinander haben, sind alle gegenseitigen Einflüsse gleich groß. Wenn jedes Literal in der Wahrscheinlichkeitsberechnung in Definition 10 durch n, also die Aggregationsfunktion geteilt wird, wird jedes Literal um den gleichen Anteil herunterskaliert, was auch genau die von jedem Literal benötigte Einflussgröße ist, da sie der Verbindungsanzahl jedes Literals entspricht.

Damit ist gezeigt, dass die Voraussetzung identischer Einträge im Verbindungsvektor eines DA-MLN notwendig ist, um eine exakte Skalierung zu ermöglichen und den Aggregat-Effekt vollständig zu beseitigen.

4.3 Gegenbeispiel von DA-MLN ohne identische Einträge im Verbindungsvektor

Im vorhergehenden Abschnitt haben wir ausführlich besprochen, welche Voraussetzungen ein DA-MLN erfüllen muss, um allen Literalen einer Formel eine exakte Skalierung zu ermöglichen, und den Aggregat-Effekt komplett zu beseitigen. Wenn man diesen Aggregat-Effekt nicht mehr hat, kann man somit vermuten, dass solche DA-MLN gewichtsabhängige asymptotische Wahrscheinlichkeiten haben, wenn die Domänengrößen gegen unendlich gehen, welche (DA-)RLR auch haben.

Um die Beschränkung des Verbindungsvektors ausdrücklich darzustellen, ist ein Gegenbeispiel zur Veranschaulichung sinnvoll.

Beispiel 2. Sei T ein DA-MLN über einer relationalen Signatur $\mathcal{L}$ mit einer einzigen Formel $P(x) \wedge Q(y,z) \wedge R(z) : \omega$. Für die Menge der Variablensymbole V sei $|D|_V = \prod_{x \in V} |D_x|$, wobei $|D_x| = |D_y| = |D_z| = \{1,2\}$ gilt. Somit gilt $|D_x| = |D_y| = |D_z| = 2$.

Der Skalierungsfaktor $\vec{V}$ besteht in diesem Fall aus den folgenden Einträgen:

$$\vec{V}_{P(x)} = |D_y| * |D_z| = 4$$

$$\vec{V}_{Q(x,y)} = |D_z| = 2$$

$$\vec{V}_{R(z)} = |D_x| * |D_y| = 4$$

Hier wird angenommen, dass man sich für das Verhalten vom $Q(y, z)$ interessiert. Daher betrachten wir dieses Literal als unabhängige atomare Formel.

Anhand des PBRN unten sieht man, dass die Anzahl der Verbindungen ausgehend von $P(x)$ und $R(z)$ unterschiedlich ist (Anzahl grüner Kanten ungleich Anzahl roter Kanten). Wenn wir nun den Verbindungsvektoreintrag $\vec{V}_{P(x)} = 4$ als Aggregationsfunktion verwenden würden, würde das Literal $Q(y, z)$ "überkompensiert", da $Q(y, z)$ nur um Verbindungsvektoreintrag $\vec{V}_{Q(y,z)} = 2$ kompensiert werden muss. Dasselbe würde auch passieren wenn man den Verbindungsvektoreintrag $\vec{V}_{R(z)} = 4$ als Aggregationsfunktion verwenden würde.

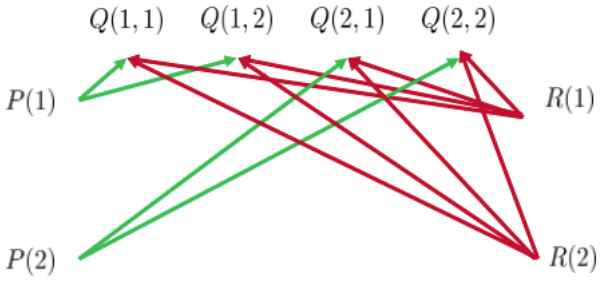

Abb. 6. PRBN des DA-RLR mit Betracht auf $Q(y, z)$

Analog würde eine "Unterkompensation" geschehen, wenn man den Verbindungsvektoreintrag $\vec{V}_{Q(y,z)} = 2$ als Aggregationsfunktion nähme. Dann würden $P(x)$ und $R(z)$ "unterkompensiert", da sie eigentlich um Faktor $\vec{V}_{P(x)} = \vec{V}_{R(z)} = 4$ herunterskaliert werden müssten, was nicht mehr anhand des obigen PRBN dargestellt wird.

Egal für welche Aggregationsfunktion man sich in diesem DA-MLN entscheidet, es gibt es immer einen Aggregat-Effekt, den man nicht komplett beseitigen kann.

Nun ist die Beschränkung für DA-MLN ohne Aggregat-Effekt gut erörtert worden - es wird die gleiche Verbindungsanzahl zwischen allen Literalen benötigt, was zu identischen Einträgen im Verbindungsvektor führt. Im nächsten Abschnitt werden wir formal beweisen, dass einige solche DA-MLN Fragmente gewichtsabhängige asymptotische Wahrscheinlichkeiten haben, wenn die Domänengröße gegen unendlich geht.

5 Beweis der Gewichtsabhängigen Konvergenz der Wahrscheinlichkeiten in DA-MLN mit ∧-Quantor mithilfe vom Satz von Beppo Levi

Wir werden nun einige grundlegende DA-MLN Fragmente als Muster nehmen und beweisen, dass sie gewichtsabhängige asymptische Wahrscheinlichkeiten haben, wenn die Domänengrößen gegen unendlich gehen.

Die ersten zwei Fragmente behandeln DA-MLN mit ∧-Quantoren. Sie unterscheiden sich nur im Definitionsbereich des Gewichts ω. Da die zwei Fälle gegenseitig ergänzend sind, werden wir am Ende dieses Abschnitts sehen, dass das Resultat eigentlich für alle Gewichte $\omega \in \mathbb{R}$ gilt.

5.1 Falls $\omega > 0$

Theorem 1. *Sei T_ω, n ein (möglicherweise mehrfach sortiertes) DA-MLN mit einer einzigen Formel $P(x) \wedge Q(y) : \omega$, $\omega > 0$. Sei $\mathfrak{X}$ eine Struktur auf der Domänengröße n. Für die Domänengröße gilt hier $|D_x| = |D_y| = n (\in \mathbb{N})$.*

Wenn die Domänengröße n gegen unendlich geht, konvergieren die Wahrscheinlichkeiten von $P(x)$ und $Q(y)$ gegen einen gewichtsabhängigen Wert $x (\in (0, 1))$.

Beweis. Da n das größte Element im Verbindungsvektor $\overrightarrow{V}_{T,n} = (n, n)$ ist, skalieren wir mit n. Außerdem sei $|\cdot|$ die Anzahl an Belegungen, die diese Formel erfüllen.

Da $\omega > 0$ gegeben ist, gilt für die Wahrscheinlichkeit $\mathcal{P}(Q(y))$:

$$\frac{1}{2} \leq \mathcal{P}(Q(y)) \leq 1 \tag{1}$$

Analog gilt natürlich auch:

$$\frac{1}{2} \leq \mathcal{P}(P(x)) \leq 1$$

Wir betrachten in diesem Beweis erstens die Wahrscheinlichkeit von $Q(Y)$.

Hier kann man das starke Gesetz der großen Zahlen anwenden (siehe Abschnitt 2.6). Da die Domänengröße n hier anschaulich die Durchführung n unabhängiger Zufallsexperimente bedeutet, kann man sagen, dass die relative Häufigkeit, dass $P(x_0)$ mit $x_0 \in D_x$ wahr ist, sich um die theoretische Wahrscheinlichkeit von $P(x)$ stabilisiert. Daraus ergibt sich dieses Intervall.

$$\frac{n}{2} \leq |P(x)| \leq n \tag{2}$$

Trivialerweise gilt dann auch:

$$\frac{\omega}{n}|P(x)| \geq \frac{\omega}{n} \cdot \frac{n}{2} = \frac{\omega}{2}$$

$$\frac{\omega}{n}|P(x)| \leq \frac{\omega}{n} \cdot n = \omega$$

Insgesamt gilt:

$$\frac{\omega}{n}|P(x)| \in \left[\frac{\omega}{2}, \omega\right] \text{ wobei } \omega > 0, n \in \mathbb{N}. \tag{3}$$

Nach Definition 10 aus Abschnitt 3.1 gilt für die Wahrscheinlichkeit von $Q(y)$ folgende Gleichung:

$$\begin{aligned}\mathcal{P}(Q(y)) &\overset{\text{Def } 8}{=} \int_{\mu T_\omega, n} \text{sigmoid}\left(\delta^{T,n}_{Q(y)}\right) \\ &\overset{\text{Def } \delta}{=} \int_{\mu T_\omega, n} \text{sigmoid}\left(\sum_i \frac{\omega_i}{n} n_i\left(\mathfrak{X}_{Q_{(y)}}\right) - \sum_i \frac{\omega_i}{n} n_i\left(\mathfrak{X}_{\neg Q_{(y)}}\right)\right) \\ &= \int_{\mu T_\omega, n} \text{sigmoid}\left(\frac{\omega}{n}|P(x)|\right) \end{aligned} \tag{4}$$

Erste Schleife der Wahrscheinlichkeitsberechnung

Da wir das Intervall schon von $\frac{\omega}{n}|P(x)|$ aus der Gleichung 3 wissen, setzen wir dieses Intervall direkt in die Gleichung 4 ein:

$$\int_{\mu T_\omega, n} \text{sigmoid}\left(\frac{\omega}{2}\right) \leq \mathcal{P}(Q(y)) \leq \int_{\mu T_\omega, n} \text{sigmoid}\left(\omega\right) \tag{5}$$

Hier kann man das starke Gesetz der großen Zahlen, das bereits in Abschnitt 2.6 besprochen wurde, im Rahmen der Gleichung 4 anwenden. Der exakte Anteil von Belegungen, die $Q(y)$ erfüllen, nähert sich der Wahrscheinlichkeit von $Q(y)$ für fast alle $\mathfrak{X}$ an. Dadurch fällt das Integral-Zeichen in der Ungleichung 11 weg, also:

$$\text{sigmoid}\left(\frac{\omega}{2}\right) \leq \mathcal{P}(Q(y)) \leq \text{sigmoid}(\omega) \tag{6}$$

Davon nehmen wir zuerst die obere Schranke $\text{sigmoid}(\omega)$ zur weiteren Berechnung. Da die Funktion sigmoid als $\text{sigmoid}(k) := \frac{\exp(k)}{\exp(k)+1}$ definiert ist, gilt:

$$\mathcal{P}(Q(y)) \leq \text{sigmoid}(\omega) \overset{\text{Def.}}{=} \frac{\exp(\omega)}{\exp(\omega)+1}$$

Wir wissen, dass die sigmoid-Funktion die zwei Asymptoten $y = 1$ und $y = 0$ hat. (Den Beweis lassen wir hier weg, da er einfach in gängiger Literatur zu finden ist.) Eine graphische Darstellung von sigmoid findet man unten.

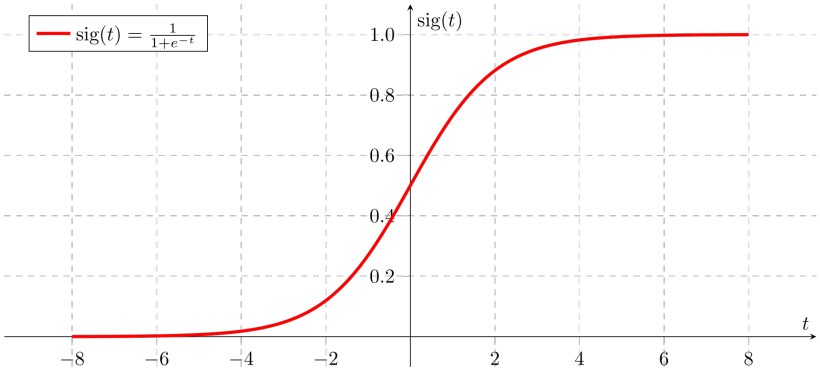

Abb. 7. Graph von sigmoid-Funktion ($sig(t) := sigmoid(t), t \in [-8, 8]$)

Wegen der oberen Asymptote $y = 1$ von sigmoid gilt:

$$\mathcal{P}(Q(y)) \leq \text{sigmoid}\left(\frac{\omega}{2}\right) < 1 \tag{7}$$

Für die untere Schranke $\text{sigmoid}(\frac{\omega}{2})$ gilt nach der Definition von sigmoid:

$$\mathcal{P}(Q(y)) \geq \text{sigmoid}(\frac{\omega}{2}) \overset{\text{Def.}}{=} \frac{\exp(\frac{\omega}{2})}{\exp(\frac{\omega}{2}) + 1}$$

Da $\omega > 0$, ist $\frac{\omega}{2} > 0$. Es gilt nach der streng wachsenden Monotonie der Exponentialfunktion mit e: $\exp(\frac{\omega}{2}) > \exp(0) = 1$. Es gilt für die untere Schranke folgende Ungleichung, denn sigmoid wächst streng monoton.

$$\mathcal{P}(Q(y)) \geq \frac{\exp(\frac{\omega}{2})}{\exp(\frac{\omega}{2}) + 1} > \frac{1}{1+1} = \frac{1}{2} \tag{8}$$

Setzen wir die Gleichungen 7 und 8 zusammen bekommen wir diese Ungleichung:

$$\frac{1}{2} < \text{sigmoid}\left(\frac{\omega}{2}\right) \leq \mathcal{P}(Q(y)) \leq \text{sigmoid}\left(\omega\right) < 1 \tag{9}$$

Da $P(x)$ und $Q(y)$ komplett symmetrisch in der Formel $P(x) \wedge Q(y)$ stehen, kann die Ungleichung 9 problemlos auf $\mathcal{P}(P(x))$ übertragen werden, nämlich:

$$\frac{1}{2} < \text{sigmoid}\left(\frac{\omega}{2}\right) \leq \mathcal{P}(P(x)) \leq \text{sigmoid}\left(\omega\right) < 1$$

Zweite Schleife der Wahrscheinlichkeitsberechnung

Da wir $\mathcal{P}(P(x)) \in \left(\text{sigmoid}\left(\frac{\omega}{2}\right), \text{sigmoid}\left(\omega\right)\right)$ schon wissen, können wir es nochmal in Gleichung 4 einsetzen, um ein noch genaueres Intervall für $\mathcal{P}(Q(y))$ und natürlich $\mathcal{P}(P(x))$ zu bekommen.

Hierzu verwenden wir zuerst das starke Gesetz der großen Zahlen (siehe Abschnitt 2.6) genau so, wie wir es in der Ungleichung 6.1 verwendet haben. Ganz analog bekommen wir:

$$\frac{\omega}{n}|P(x)| \in \left[\omega \cdot \text{sigmoid}\left(\frac{\omega}{2}\right), \omega \cdot \text{sigmoid}(\omega)\right] \text{ wobei } \omega > 0, n \in \mathbb{N}. \tag{10}$$

Nach der Einsetzung in die Gleichung 4 erhalten wir die folgende Ungleichung:

$$\int_{\mu T_\omega, n} \text{sigmoid}\left(\omega \cdot \text{sigmoid}\left(\frac{\omega}{2}\right)\right) \leq \mathcal{P}(Q(y)) \leq \int_{\mu T_\omega, n} \text{sigmoid}\left(\omega \cdot \text{sigmoid}(\omega)\right) \tag{11}$$

Das starke Gesetz der großen Zahlen hilft uns hier nochmal, das Integralzeichen wegzuräumen, so wie in der Ungleichung 6. Danach erhalten wir das genauere und schmälere Intervall von $\mathcal{P}(Q(y))$.

$$\mathcal{P}(Q(y)) = \left[\text{sigmoid}\left(\omega \cdot \text{sigmoid}(\frac{\omega}{2})\right), \text{sigmoid}(\omega \cdot \text{sigmoid}(\omega))\right] \tag{12}$$

Nun wollen wir das Intervall 12 mit dem Ergebnis von $\mathcal{P}(Q(y))$ aus der ersten Schleife (Ungleichung 9), nämlich $\mathcal{P}(P(x)) \in \left(\text{sigmoid}\left(\frac{\omega}{2}\right), \text{sigmoid}\left(\omega\right)\right)$ vergleichen.

Für die untere Schranke der beiden Intervalle müssen wir $\omega \cdot \text{sigmoid}(\frac{\omega}{2})$ und $\frac{\omega}{2}$ vergleichen, da sigmoid streng monoton wächst und die beiden Terme innerhalb des sigmoid()-Ausdruckes stehen. Es ist offensichtlich $\frac{\omega}{2} = \frac{1}{2} \cdot \omega$ und $\omega \cdot \text{sigmoid}(\frac{\omega}{2}) = \text{sigmoid}(\frac{\omega}{2}) \cdot \omega$. Das heißt, dass die beiden ω-Terme sich nur im Koeffizienten, der jeweils vor ω steht, unterscheiden. Wenn wir die beiden Koeffizienten miteinander vergleichen, fällt schnell auf, dass $\frac{1}{2} < \text{sigmoid}(\frac{\omega}{2})$ gilt, denn $\omega > 0$ und die sigmoid-Funktion geht durch $(0, 0.5)$. Insgesamt gilt für die untere Schranke:

$$\text{sigmoid}\left(\frac{\omega}{2}\right) < \text{sigmoid}\left(\omega \cdot \text{sigmoid}\left(\frac{\omega}{2}\right)\right) \leq \mathcal{P}(Q(y)) \tag{13}$$

Für die obere Schranke der beiden Intervalle müssen wir $\omega \cdot \text{sigmoid}(\omega)$ und ω vergleichen, da sigmoid streng monoton wächst und die beiden Terme innerhalb des sigmoid()-Ausdruckes stehen. Es ist offensichtlich $\omega = 1 \cdot \omega$ und $\omega \cdot \text{sigmoid}(\omega) = \text{sigmoid}(\omega) \cdot \omega$. Genauso wie für den Vergleich der unteren Schranken sehen wir 1 und sigmoid(ω) als Koeffizienten jeweils vor ω. Da $y = 1$ die obere Asymptote der sigmoid-Funktion ist, gilt trivialerweise $\text{sigmoid}(\omega) < 1$. Daraus ergibt sich:

$$\mathcal{P}(Q(y)) \leq \text{sigmoid}(\omega \cdot \text{sigmoid}(\omega)) < \text{sigmoid}(\omega) \tag{14}$$

Verallgemeinerung der Wahrscheinlichkeitsberechnungsschleifen

Anhand der Ungleichungen 13 und 14 sehen wir, dass das Intervall für $\mathcal{P}(Q(y))$ im Vergleich zur ersten Schleife schmäler ist, und die Schranken sich um sigmoid($\omega \cdot (\textit{letzte Schranke})$) geändert sind.

Wir können tatsächlich diese Änderungen ab der zweiten Schleife auf allen Schleifen verallgemeinern.

Sei Term $\mathcal{O}_n$ die obere Schranke des Intervalls der n. Schleife, die wir bereits aus der letzten Schleife wissen. Ähnlich definieren wir Term $\mathcal{U}_n$ für die untere Schranke. Dann gilt am Beginn der n. Schleife schon $\mathcal{P}(Q(y)) \in [\mathcal{U}_n, \mathcal{O}_n]$.

Nach der Gleichung 4 bekommen wir

$$\int_{\mu T_\omega, n} \text{sigmoid}\left(\omega \cdot \mathcal{U}_n\right) \leq \mathcal{P}(Q(y)) \leq \int_{\mu T_\omega, n} \text{sigmoid}\left(\omega \cdot \mathcal{O}_n\right) \tag{15}$$

Das starke Gesetz der großen Zahlen hilft uns hier nochmal, das Integralzeichen wegzuräumen, so wie in der Ungleichung 6. Danach erhalten wir das genauere und schmälere Intervall von $\mathcal{P}(Q(y))$.

$$\mathcal{P}(Q(y)) = [\text{sigmoid}\left(\omega \cdot \mathcal{U}_n\right), \text{sigmoid}(\omega \cdot \mathcal{O}_n)] \tag{16}$$

Wegen der Symmetrie der Formel $P(x) \wedge Q(y)$ gilt das Intervall 16 wieder für $P(x)$, und wirkt dann weiter auf $Q(y)$ selber. So beginnt die nächste Schleife $n + 1$.

Nun bleiben die Fragen, ob die neue untere Schranke immer größer als die letzte untere Schranke ist, und ob die neue obere Schranke immer kleiner als die letzte obere Schranke ist. Außerdem ist unklar, ob sich die obere und die untere Schranke nie überschneiden. Dazu können wir Induktion als Beweismethode nehmen.

Induktionsbeweis für die obere Schranke des Intervalls für $\mathcal{P}(Q(y))$

Die Behauptung wäre, dass $\mathcal{O}_{n+1} < \mathcal{O}_n$ immer gilt, wobei $n \in \mathbb{N}$ ist.

Beweis. Induktionsanfang$(n = 1) : \mathcal{O}_1 = \text{sigmoid}(\omega),\ \mathcal{O}_2 = \text{sigmoid}(\omega \cdot \text{sigmoid}(\omega))$. $\mathcal{O}_2 < \mathcal{O}_1$ gilt haben wir schon in der zweiten Schleife der Wahrscheinlichkeitsberechnung bewiesen.

Induktionsschritt: Es gelte die Aussage für ein $n \in \mathbb{N}$, d.h. es gelte $\mathcal{O}_{n+1} < \mathcal{O}_n$. Zu zeigen ist die Aussagen für $n + 1$, also

$$\mathcal{O}_{n+2} < \mathcal{O}_{n+1} \tag{17}$$

$$\mathcal{O}_{n+2} = \text{sigmoid}(\omega \cdot \mathcal{O}_{n+1}) \tag{18}$$

$$\mathcal{O}_{n+1} = \text{sigmoid}(\omega \cdot \mathcal{O}_n) \tag{19}$$

Wir möchten nun $\mathcal{O}_{n+2}$ und $\mathcal{O}_{n+1}$ vergleichen. Um die beiden Ausdrücke zu vergleichen müssen wir $\mathcal{O}_{n+1}$ und $\mathcal{O}_n$ vergleichen, da sigmoid streng monoton wächst und die beiden Terme innerhalb des sigmoid()-Ausdruckes 18 und 19 stehen. Da wir im Induktionsschritt schon $\mathcal{O}_{n+1} < \mathcal{O}_n$ gelten lassen, können wir aufgrund der streng wachsenden Monotonie von sigmoid-Funktion direkt sagen, dass Ungleichung 17 gilt.

Damit haben wir gezeigt, dass die obere Schranke des Intervalls nach jeder Schleife der Wahrscheinlichkeitsberechnung immer kleiner als die letzte obere Schranke ist.

Mithilfe der Induktion können wir ein ähnliches Resultat für die untere Schranke zeigen.

Induktionsbeweis für die untere Schranke des Intervalls für $\mathcal{P}(Q(y))$

Die Behauptung wäre, dass $\mathcal{U}_{n+1} > \mathcal{U}_n$ immer gilt, wobei $n \in \mathbb{N}$ ist.

Beweis. Induktionsanfang$(n = 1) : \mathcal{U}_1 = \text{sigmoid}(\frac{\omega}{2}),\ \mathcal{U}_2 = \text{sigmoid}(\omega \cdot \text{sigmoid}(\frac{\omega}{2}))$. $\mathcal{U}_2 > \mathcal{U}_1$ gilt haben wir schon in der zweiten Schleife der Wahrscheinlichkeitsberechnung bewiesen.

Induktionsschritt: Es gelte die Aussage für ein $n \in \mathbb{N}$, d.h. es gelte $\mathcal{U}_{n+1} > \mathcal{U}_n$. Zu zeigen ist die Aussagen für $n+1$, also

$$\mathcal{U}_{n+2} > \mathcal{U}_{n+1} \tag{20}$$

$$\mathcal{U}_{n+2} = \text{sigmoid}(\omega \cdot \mathcal{U}_{n+1}) \tag{21}$$

$$\mathcal{U}_{n+1} = \text{sigmoid}(\omega \cdot \mathcal{U}_n) \tag{22}$$

Wir möchten nun $\mathcal{U}_{n+2}$ und $\mathcal{U}_{n+1}$ vergleichen. Um die beiden Ausdrücke zu vergleichen müssen wir $\mathcal{U}_{n+1}$ und $\mathcal{U}_n$ vergleichen, da sigmoid streng monoton wächst und die beiden Terme innerhalb des sigmoid()-Ausdruckes 21 und 22 stehen. Da wir im Induktionsschritt schon $\mathcal{U}_{n+1} > \mathcal{U}_n$ gelten lassen, können wir aufgrund der streng wachsenden Monotonie von sigmoid-Funktion direkt sagen, dass Ungleichung 20 gilt.

Damit haben wir gezeigt, dass die untere Schranke des Intervalls nach jeder Schleife der Wahrscheinlichkeitsberechnung immer größer als die letzte untere Schranke ist.

Verschmelzung beider Grenzwerte der Schranken
Nun müssen wir nur noch zeigen, dass die Grenzwerte beider Schranken gleich sind, indem wir zeigen, dass folgende Gleichung gilt, unabhängig vom Gewicht ω.

$$\lim_{n \to \infty} (\mathcal{O}_n - \mathcal{U}_n) = 0 \tag{23}$$

Beweis. Um die Gleichung 23 zu zeigen, brauchen wir den Satz von der monotonen Konvergenz (auch als Satz von Beppo Levi genannt).

Theorem 2. *Wenn eine Folge von reellen Zahlen monoton fallend ist und nach unten begrenzt ist, dann ist ihr Infimum die Grenze.*

Um zu zeigen, dass $\mathcal{O}_n - \mathcal{U}_n$ nach unten begrenzt auf 0 ist, nutzen wir nochmal die Induktion an.

Induktionsanfang $(n = 1) : \mathcal{O}_1 - \mathcal{U}_1 = \text{sigmoid}(\omega) - \text{sigmoid}\left(\frac{\omega}{2}\right) > 0$

Induktionsschritt: Nun nehmen wir an, dass $\mathcal{O}_n - \mathcal{U}_n > 0$ gilt.

Für $\mathcal{O}_{n+1} - \mathcal{U}_{n+1}$ gilt folgende Gleichung:

$$\mathcal{O}_{n+1} - \mathcal{U}_{n+1} = \text{sigmoid}(\omega \cdot \mathcal{O}_n) - \text{sigmoid}(\omega \cdot \mathcal{U}_n) > 0 \tag{24}$$

Der Beweis dazu ist einfach. Da $\text{sigmoid}(\omega \cdot \mathcal{O}_n)$ und $\text{sigmoid}(\omega \cdot \mathcal{U}_n)$ unterscheiden sich nur im Kern, nämlich steckt in $\text{sigmoid}(\omega \cdot \mathcal{O}_n)$ ein $\text{sigmoid}(\omega)$ und in $\text{sigmoid}(\omega \cdot \mathcal{U}_n)$ ein $\text{sigmoid}(\frac{\omega}{2})$, und die Verschachtelung von den beiden Ausdrücken komplett identisch sind, gilt die Gleichung 24 durch die mehrfachen Verwendungen der streng wachsenden Monotonie der sigmoid-Funktion.

Die Gleichung 24 sagt uns, dass diese Folge der Schrankendifferenz nach unten begrenzt ist. Da sie auch monoton fallend ist, welches schon durch die zwei vorherigen Induktionen bewiesen ist, kann man sagen, dass ihr Infimum auch ihr Grenzwert ist. Wir nennen diesen Grenzwert a.

Nun rechnen wir die Grenzwerte beider Seite der Gleichung 24 aus. Da die beiden Seiten sich um zwei Darstellungen derselben Zahlenfolge handeln, und jede Zahlenfolge höchstens gegen einen Grenzwert konvergieren

kann, haben die beiden Seiten auch denselben Grenzwert. Am Ende der Berechnung werden wir noch sehen, dass a, der Grenzwert der Schrankendifferenz, für alle $\omega \in \mathbb{R}$ gleich ist. Die Grenzwertberechnung ergibt sich wie folgt:

$$\begin{aligned}
a &= \lim_{n\to\infty} (\text{sigmoid}(\omega \cdot \mathcal{O}_n) - \text{sigmoid}(\omega \cdot \mathcal{U}_n)) \\
&= \lim_{n\to\infty} \left(\frac{\exp(\omega\ \mathcal{O}_n)}{\exp(\omega\ \mathcal{O}_n)+1} - \frac{\exp(\omega\ \mathcal{U}_n)}{\exp(\omega\ \mathcal{U}_n)+1} \right) \\
&= \lim_{n\to\infty} \left(\frac{\exp(\omega\ \mathcal{O}_n)\cdot(\exp(\omega\ \mathcal{U}_n)+1) - \exp(\omega\ \mathcal{U}_n)\cdot(\exp(\omega\ \mathcal{O}_n)+1)}{\exp(\omega\ \mathcal{O}_n)\cdot\exp(\omega\ \mathcal{U}_n) + \exp(\omega\ \mathcal{O}_n) + \exp(\omega\ \mathcal{U}_n) + 1} \right) \\
&= \lim_{n\to\infty} \left(\frac{\exp(\omega\ \ (\mathcal{O}_n + \mathcal{U}_n)) + \exp(\omega\ \mathcal{O}_n) - \exp(\omega\ \ (\mathcal{O}_n + \mathcal{U}_n)) - \exp(\omega\ \mathcal{U}_n)}{\exp(\omega\ \ (\mathcal{O}_n + \mathcal{U}_n)) + \exp(\omega\ \mathcal{O}_n) + \exp(\omega\ \mathcal{U}_n) + 1} \right) \\
&= \lim_{n\to\infty} \left(\frac{\exp(\omega\ \mathcal{O}_n) - \exp(\omega\ \mathcal{U}_n)}{\exp(\omega\ \ (\mathcal{O}_n + \mathcal{U}_n)) + \exp(\omega\ \mathcal{O}_n) + \exp(\omega\ \mathcal{U}_n) + 1} \right)
\end{aligned} \tag{25}$$

Da $\exp(\omega\ \mathcal{U}_n) > 0$ wegen der unteren Asymptote $x = 0$ von Exponentialfunktion mit e gilt, können wir zugleich den Nenner und den Zähler in Gleichung 25 durch $\exp(\omega\ \mathcal{U}_n)$ teilen.

$$\begin{aligned}
a &= \lim_{n\to\infty} \left(\frac{\frac{\exp(\omega\ \mathcal{O}_n)}{\exp(\omega\ \mathcal{U}_n)} - 1}{\frac{\exp(\omega\ \ (\mathcal{O}_n - \mathcal{U}_n))}{\exp(\omega\ \mathcal{U}_n)} + \frac{\exp(\omega\ \mathcal{O}_n)}{\exp(\omega\ \mathcal{U}_n)} + 1 + \frac{1}{\exp(\omega\ \mathcal{U}_n)}} \right) \\
&= \lim_{n\to\infty} \left(\frac{\exp(\omega\ (\mathcal{O}_n - \mathcal{U}_n)) - 1}{\exp(\omega\ \mathcal{O}_n) + \exp(\omega\ (\mathcal{O}_n - \mathcal{U}_n)) + 1 + \frac{1}{\exp(\omega\ \mathcal{U}_n)}} \right) \\
&= \frac{\exp(\omega \cdot a) - 1}{\exp(\omega\ \mathcal{O}_n) + \exp(\omega \cdot a) + 1 + \frac{1}{\exp(\omega\ \mathcal{U}_n)}}
\end{aligned} \tag{26}$$

Hier ist die einzige Lösung $a = 0$, und dies gilt nicht nur für alle $\omega > 0$, sondern auch für alle $\omega \in \mathbb{R}$. Das sieht man leicht, wenn man $a = 0$ zurück in die Gleichung 26 einsetzt. (Dieses Resultat werden wir im Abschnitt 5.2 auch brauchen.)

Wir können nach der formalen Definition der asymptotischen Funktion sagen, dass $\mathcal{P}(Q(y))$ asymptotisch ist, wenn die Domänengröße nach unendlich geht, solange das Gewicht $\omega > 0$ ist. Außerdem hängen die Grenzwerte von $\mathcal{O}_n$ und $\mathcal{U}_n$ rein vom Gewicht ω ab, da ω die einzige Variable in den beiden Ausdrücken $\mathcal{O}_n$ und $\mathcal{U}_n$ ist. Somit ist die Konvergenz auch gewichtsabhängig.

Wegen der Symmetrie der Formel $P(x) \wedge Q(y)$, lässt sich diese gewichtsabhängige asymptotische Wahrscheinlichkeit von $Q(y)$ eins zu eins auf $P(x)$ übertragen.

5.2 Falls $\omega < 0$

Nun wollen wir den ergänzenden Fall $\omega < 0$ behandeln.

Theorem 3. *Sei T_ω, n ein DA-MLN mit einer einzigen Formel $P(x) \wedge Q(y) : \omega$, $\omega > 0$. Für die Domänengröße gilt hier $|D_x| = |D_y| = n(\in \mathbb{N})$. Sei $\mathfrak{X}$ eine Struktur auf der Domänengröße n.*

Wenn die Domänengröße n gegen unendlich geht, konvergieren die Wahrscheinlichkeiten von $P(x)$ und $Q(y)$ gegen einen gewichtsabhängigen Wert $x(\in (0, 1))$.

Beweis. Da n das größte Element im Verbindungsvektor $\overrightarrow{V}_{T,n} = (n, n)$ ist, skalieren wir mit n. Außerdem sei $|\cdot|$ die Anzahl an Belegungen, die diese Formel erfüllen.

Da $\omega < 0$ gegeben ist, gilt für die Wahrscheinlichkeit $\mathcal{P}(Q(y))$:

$$\mathcal{P}(Q(y)) \in \left[0, \frac{1}{2}\right] \tag{27}$$

Analog gilt natürlich auch:

$$\mathcal{P}(P(x)) \in \left[0, \frac{1}{2}\right]$$

Wir betrachten in diesem Beweis erstens die Wahrscheinlichkeit von $Q(y)$ (für $P(x)$ verläuft der Beweis weitgehend analog wegen der Symmetrie von $P(x) \wedge Q(y)$).

Analog wie in der Gleichung 6.1 kann man das starke Gesetz der großen Zahlen verwenden (siehe Abschnitt 2.6). Es ergibt sich:

$$\frac{\omega}{n}|P(x)| \leq \frac{\omega}{n} \cdot 0 = 0$$

$$\frac{\omega}{n}|P(x)| \geq \frac{\omega}{n} \cdot \frac{n}{2} = \frac{\omega}{2}$$

Hier unterscheiden sich wegen $\omega < 0$ die zwei Ungleichungen symmetrisch vom ersten Fall.

Insgesamt gilt:

$$\frac{\omega}{n}|P(x)| \in \left[\frac{\omega}{2}, 0\right] \text{ wobei } \omega < 0, n \in \mathbb{N}. \tag{28}$$

Nach der Definition 10 aus dem Abschnitt 3.1 ergibt sich für die Wahrscheinlichkeit von $Q(y)$ wieder folgendes (siehe Gleichung 4):

$$\mathcal{P}(Q(y)) = \int_{\mu T_\omega, n} \text{sigmoid}\left(\frac{\omega}{n}|P(x)|\right) \tag{29}$$

Erste Schleife der Wahrscheinlichkeitsberechnung

Da wir schon das Intervall von $\frac{\omega}{n}|P(x)|$ aus der Gleichung 28 wissen, kann man dies direkt in die Gleichung 29 einsetzen:

$$\int_{\mu T_\omega, n} \text{sigmoid}\left(\frac{\omega}{2}\right) \leq \mathcal{P}(Q(y)) \leq \int_{\mu T_\omega, n} \text{sigmoid}\,(0) \tag{30}$$

Analog zur Gleichung 6 kann man das starke Gesetz der großen Zahlen (siehe Abschnitt 2.6) nochmal im Rahmen der Gleichung 29 anwenden, um das Integral-Zeichen in der Ungleichung 30 wegzuräumen, also:

$$\text{sigmoid}\left(\frac{\omega}{2}\right) \leq \mathcal{P}(Q(y)) \leq \text{sigmoid}\,(0)$$

Davon nehmen wir zuerst die untere Schranke zur weiteren Berechnung. Es gilt wegen der unteren Asymptote $y = 0$ der sigmoid-Funktion für die untere Schranke sigmoid $\left(\frac{\omega}{2}\right)$:

$$\mathcal{P}(Q(y)) > 0 \tag{31}$$

Für die obere Schranke sigmoid(0) gilt:

$$\mathcal{P}(Q(y)) \leq \text{sigmoid}(0) \overset{\text{Def.}}{=} \frac{\exp(0)}{\exp(0)+1} = \frac{1}{1+1} = \frac{1}{2} \tag{32}$$

wegen der Eigenschaft der Exponentialfunktion mit e: $\exp(0) = 1$.

Setzen wir die Gleichungen 31 und 32 zusammen bekommen wir das Intervall für $\mathcal{P}(Q(y))$, nämlich

$$0 < \text{sigmoid}(\frac{\omega}{2}) \leq \mathcal{P}(Q(y)) \leq \text{sigmoid}(0) = \frac{1}{2} \tag{33}$$

Da $P(x)$ und $Q(y)$ komplett symmetrisch in der Formel $P(x) \wedge Q(y)$ stehen, kann die Ungleichung 33 problemlos auf $\mathcal{P}(P(x))$ übertragen werden, nämlich:

$$0 < \text{sigmoid}\left(\frac{\omega}{2}\right) \leq \mathcal{P}(P(x)) \leq \text{sigmoid}(0) = \frac{1}{2} \tag{34}$$

Zweite Schleife der Wahrscheinlichkeitsberechnung

Da wir $\mathcal{P}(P(x)) \in \left(\text{sigmoid}\left(\frac{\omega}{2}\right), \text{sigmoid}\left(0\right)\right)$ schon wissen, können wir dieses Intervall nochmal in Gleichung 29 einsetzen, um ein noch genaueres Intervall für $\mathcal{P}(Q(y))$ zu bekommen.

Hier verwenden wir zuerst das starke Gesetz der großen Zahlen (siehe Abschnitt 2.6) genau so, wie wir es in der Ungleichung 6.1 verwendet haben. Ganz analog erhalten wir:

$$\frac{\omega}{n}|P(x)| \in \left[\omega \cdot \text{sigmoid}\left(\frac{\omega}{2}\right), \omega \cdot \text{sigmoid}(0)\right] \text{ wobei } \omega < 0 \tag{35}$$

Ganz analog wie in Abschnitt 5.1 nach Definition 10 und der Verwendung des starken Gesetzes der großen Zahlen bekommen wir das zweite und schmälere Intervall über $\mathcal{P}(Q(y))$.

$$\mathcal{P}(Q(y)) = \left[\text{sigmoid}\left(\omega \cdot \text{sigmoid}(\frac{\omega}{2})\right), \text{sigmoid}(\omega \cdot \text{sigmoid}(0))\right] \tag{36}$$

Nun wollen wir das Intervall 36 mit dem Ergebnis von $\mathcal{P}(Q(y))$ aus der ersten Schleife (Ungleichung 34), nämlich $\mathcal{P}(Q(y)) \in \left(\text{sigmoid}\left(\frac{\omega}{2}\right), \text{sigmoid}\left(0\right)\right)$ vergleichen.

Für die untere Schranke der beiden Intervalle müssen wir $\omega \cdot \text{sigmoid}(\frac{\omega}{2})$ und $\frac{\omega}{2}$ vergleichen, da sigmoid streng monoton wächst und die beiden Terme innerhalb des sigmoid()-Ausdruckes stehen. Analog zum Abschnitt 5.1 fällt schnell auf, dass $\frac{\omega}{2} < \omega \cdot \text{sigmoid}(\frac{\omega}{2})$ gilt, denn $\omega < 0$ und $\text{sigmoid}(\frac{\omega}{2}) \in \left(0, \frac{1}{2}\right)$.

Für die obere Schranke der beiden Intervalle müssen wir $\omega \cdot \text{sigmoid}(0)$ und 0 vergleichen, da sigmoid streng monoton wächst und die beiden Terme innerhalb des sigmoid()-Ausdruckes stehen. Es ist offensichtlich $\omega \cdot \text{sigmoid}(0) < 0$, da $\omega < 0$ gilt und sigmoid-Funktion immer größer als 0 ist.

Verallgemeinerung der Wahrscheinlichkeitsberechnungsschleifen

Komplett identisch wie im Abschnitt 5.1.

Nun bleiben die Fragen, ob die neue untere Schranke immer größer als die letzte untere Schranke ist, und die neue obere Schranke immer kleiner als die letzte obere Schranke ist. Außerdem ist unklar, ob sich die obere und die untere Schranke nie überschneiden. Dies werden wir nun zeigen.

Induktionsbeweis für die obere Schranke des Intervalls für $\mathcal{P}(Q(y))$

Sei $\mathcal{O}_n$ die obere Schranke des Intervalls der n. Schleife. Wir behaupten, dass $\mathcal{O}_{n+1} < \mathcal{O}_n$ für alle $n \in \mathbb{N}^+$ gilt.

Beweis. Induktionsanfang($n = 1$) : $\mathcal{O}_1 = \text{sigmoid}(0)$, $\mathcal{O}_2 = \text{sigmoid}(\omega \cdot \text{sigmoid}(0))$. $\mathcal{O}_2 < \mathcal{O}_1$ gilt, wie wir schon in der zweiten Schleife der Wahrscheinlichkeitsberechnung gesehen haben.

Induktionsschritt: Es gelte die Aussage für ein $n \in \mathbb{N}$, d.h. es gelte $\mathcal{O}_{n+1} < \mathcal{O}_n$. Zu zeigen ist die Aussage für $n + 1$, also

$$\mathcal{O}_{n+2} < \mathcal{O}_{n+1} \tag{37}$$

$$\mathcal{O}_{n+2} = \text{sigmoid}(\omega \cdot \mathcal{O}_{n+1}) \tag{38}$$

$$\mathcal{O}_{n+1} = \text{sigmoid}(\omega \cdot \mathcal{O}_n) \tag{39}$$

Der Beweis dazu ist analog wie im Abschnitt 5.1 durchführbar.

Somit haben wir gezeigt, dass die obere Schranke des Intervalls nach jeder Schleife der Wahrscheinlichkeitsberechnung immer kleiner als die letzte obere Schranke ist.

Für die untere Schranke können wir das Resultat analog zeigen.

Induktionsbeweis für die untere Schranke des Intervalls für $\mathcal{P}(Q(y))$

Sei $\mathcal{U}_n$ die untere Schranke des Intervalls der n. Schleife. Wir behaupten, dass $\mathcal{U}_{n+1} < \mathcal{U}_n$ für alle $n \in \mathbb{N}^+$ gilt.

Beweis. Induktionsanfang($n = 1$) : $\mathcal{U}_1 = \text{sigmoid}(\frac{\omega}{2})$, $\mathcal{U}_2 = \text{sigmoid}(\omega \cdot \text{sigmoid}(\frac{\omega}{2}))$. $\mathcal{U}_2 < \mathcal{U}_1$ gilt, wie wir schon in der zweiten Schleife der Wahrscheinlichkeitsberechnung gesehen haben.

Induktionsschritt: Es gelte die Aussage für ein $n \in \mathbb{N}$, d.h. es gelte $\mathcal{U}_{n+1} < \mathcal{U}_n$. Zu zeigen ist die Aussage für $n + 1$, also

$$\mathcal{U}_{n+2} < \mathcal{U}_{n+1} \tag{40}$$

$$\mathcal{U}_{n+2} = \text{sigmoid}(\omega \cdot \mathcal{U}_{n+1}) \tag{41}$$

$$\mathcal{U}_{n+1} = \text{sigmoid}(\omega \cdot \mathcal{U}_n) \tag{42}$$

Der Beweis dazu ist analog wie im Abschnitt 5.1 durchführbar.

Somit haben wir gezeigt, dass die untere Schranke des Intervalls nach jeder Schleife der Wahrscheinlichkeitsberechnung immer größer als die letzte untere Schranke ist.

Verschmelzung beider Grenzwerte der Schranken

Nun müssen wir nur noch zeigen, dass die Grenzwerte beider Schranken gleich sind, indem wir zeigen, dass die Gleichung 23 auch für den Fall $\omega < 0$ gilt.

Beweis. Um Gleichung 23 zu zeigen, brauchen wir den Satz von der monotonen Konvergenz (Theorem 2 im Abschnitt 5.1).

Um zu zeigen, dass $\mathcal{O}_n - \mathcal{U}_n$ nach unten begrenzt auf 0 ist, führen wie die Induktion analog durch wie im Abschnitt 5.1. Der einzige Unterschied in dieser Induktion im Vergleich mit der Induktion aus dem Abschnitt 5.1 ist, dass in $\text{sigmoid}(\omega \cdot \mathcal{O}_n)$ ein $\text{sigmoid}\left(\frac{\omega}{2}\right)$ und in $\text{sigmoid}(\omega \cdot \mathcal{U}_n)$ ein $\text{sigmoid}(0)$ ist. Durch die mehrfachen Verwendungen der streng wachsenden Monotonie der sigmoid-Funktion gilt die Gleichung 24 auch für den Fall $\omega < 0$.

Die Berechnung des Grenzwerts beider Seiten der Gleichung 24 ist komplett identisch wie im Abschnitt 5.1, da die Berechnung im Abschnitt 5.1 allgemein für die Schrankendifferenz gilt, unabhängig vom Gewicht ω. Die einzige Lösung ist auch $a = 0$. Dieses Resultat haben wir schon im Abschnitt 5.1 nach der Gleichung 26 schon besprochen.

Wir können nach der formalen Definition der asymptotischen Funktion sagen, dass $\mathcal{P}(Q(y))$ asymptotisch ist, wenn die Domänengröße nach unendlich geht, solange das Gewicht $\omega < 0$ ist. Außerdem hängen die Grenzwerte von $\mathcal{O}_n$ und $\mathcal{U}_n$ rein vom Gewicht ω ab, da ω die einzige Variable in den beiden Ausdrücken $\mathcal{O}_n$ und $\mathcal{U}_n$ ist. Somit ist die Konvergenz auch gewichtsabhängig.

Wegen der Symmetrie der Formel $P(x) \wedge Q(y)$, lässt sich diese gewichtsabhängige asymptotische Wahrscheinlichkeit von $Q(y)$ eins zu eins auf $P(x)$ übertragen.

Zusammenfassung

Da im Abschnitt 5.1 wir das gleiche Resultat schon für den Fall $\omega > 0$ gezeigt haben, und $\omega = 0$ einem puren Zufallsexperiment ohne Gewichtung entspricht, können wir nun sagen, dass $\mathcal{P}(Q(y))$ und $\mathcal{P}(P(x))$ für die Formel $P(x) \wedge Q(y)$ in einem DA-MLN immer gewichtsabhängig konvergiert, und es gilt für alle $\omega \in \mathbb{R}$.

6 Beweis der Gewichtsabhängigen Konvergenz der Wahrscheinlichkeiten in DA-MLN mit ∨-Quantor mithilfe vom Fixpunktsatz von Banach

Im letzten Abschnitt haben wir mithilfe vom Satz von Beppo Levi gezeigt, dass die Wahrscheinlichkeiten einzelner Literale in der Formel $P(x) \wedge Q(y)$ gewichtsabhängig konvergieren, wenn die Domänengröße n gegen unendlich geht. Nun interessieren wir uns für einen ähnlichen Fall, nämlich die Wahrscheinlichkeiten von Literalen in der Formel $P(x) \vee Q(y)$.

Tatsächlich kann man das gleiche Resultat auch für die neue Formel beweisen. Das Resultat gilt auch für alle Gewichte $\omega \in \mathbb{R}$. Allerdings ist der Beweis sehr viel aufwendiger, wenn es mithilfe vom Satz von Beppo Levi gehen muss, da die oberen und unteren Schranken gegenseitig in den Wahrscheinlichkeitsberechnungsschleifen beeinflussen, und sich nicht mehr so einfach rekursiv darstellen lassen.

Aus dem Grund werfen wir einen Blick in die Funktionalanalysis ein, und versuchen nun die gewichtsabhängige Konvergenz der Wahrscheinlichkeiten mithilfe vom Fixpunktsatz von Banach zu beweisen.

6.1 Für alle $\omega \in \mathbb{R}$

Theorem 4. *Sei T_ω, n ein DA-MLN mit einer einzigen Formel $P(x) \vee Q(y) : \omega$, $\omega \in \mathbb{R}$. Für die Domänengröße gilt hier $|D_x| = |D_y| = n (\in \mathbb{N})$. Sei $\mathfrak{X}$ eine Struktur auf der Domänengröße n.*

Wenn die Domänengröße n gegen unendlich geht, konvergieren die Wahrscheinlichkeiten von $P(x)$ und $Q(y)$ gegen einen gewichtsabhängigen Wert $x (\in (0, 1))$.

Beweis. Da n das größte Element im Verbindungsvektor $\overrightarrow{V}_{T,n} = (n, n)$ ist, skalieren wir mit n. Außerdem sei $|\cdot|$ die Anzahl an Belegungen, die diese Formel erfüllen.

Wir betrachten in diesem Beweis erstens die Wahrscheinlichkeit von $Q(Y)$.

Nach Definition 10 aus dem Abschnitt 3.1 gilt für die Wahrscheinlichkeit von $Q(y)$ folgende Gleichungen:

$$\begin{aligned}
\mathcal{P}(Q(y)) &\overset{\text{Def } 8}{=} \int_{\mu T_\omega, n} \text{sigmoid}\left(\delta^{T,n}_{Q(y)}\right) \\
&\overset{\text{Def } \delta}{=} \int_{\mu T_\omega, n} \text{sigmoid}\left(\sum_i \frac{\omega_i}{n} n_i \left(\mathfrak{X}_{Q_{(y)}}\right) - \sum_i \frac{\omega_i}{n} n_i \left(\mathfrak{X}_{\neg Q_{(y)}}\right)\right) \\
&= \int_{\mu T_\omega, n} \text{sigmoid}\left(\frac{\omega}{n} |\neg P(x)|\right) \qquad (43)
\end{aligned}$$

Man kann das starke Gesetz der großen Zahlen (siehe Abschnitt 2.6) hier andersrum anwenden, um $|\neg P(x)|$ anders darzustellen.

Da die Domänengröße n hier anschaulich die Durchführung n unabhängiger Zufallsexperimente bedeutet, kann man sagen, dass die relative Häufigkeit, dass $P(x_0)$ mit $x_0 \in D_x$ falsch ist, sich um die theoretische Wahrscheinlichkeit von $\neg P(x)$ stabilisiert. Daraus ergibt sich diese Gleichung, wenn n groß genug ist.

$$\mathcal{P}(\neg P(x)) = n \cdot |\neg P(x_0)|$$

Wir setzen die obige Gleichung in die Gleichung 43 ein.

$$\mathcal{P}(Q(y)) = \int_{\mu T_\omega, n} \text{sigmoid}\left(\omega \cdot \mathcal{P}(\neg P(x))\right)$$

Da für die Wahrscheinlichkeitsfunktion einer Formel $f(\overrightarrow{a})$ gilt,

$$\mathcal{P}(f(\overrightarrow{a})) + \mathcal{P}(\neg f(\overrightarrow{a})) = 1$$

folgt unmittelbar auch für das Literal $P(x)$:

$$\mathcal{P}(P(x)) + \mathcal{P}(\neg P(x)) = 1$$
$$\mathcal{P}(\neg P(x)) = 1 - \mathcal{P}(P(x))$$

Daher gilt weiterhin für die Gleichung 43 folgendes:

$$\mathcal{P}(Q(y)) = \int_{\mu T_\omega, n} \text{sigmoid}\left(\omega \cdot (1 - \mathcal{P}(P(x)))\right) \tag{44}$$

Hier kann man das starke Gesetz der großen Zahlen (siehe Abschnitt 2.6) im Rahmen der Gleichung 44 anwenden. Der exakte Anteil von Belegungen, die $Q(y)$ erfüllen, nähert sich der Wahrscheinlichkeit von $Q(y)$ für fast alle $\mathfrak{X}$ an. Dadurch fällt das Integral-Zeichen in der Gleichung 44 weg, also:

$$\mathcal{P}(Q(y)) = \text{sigmoid}\left(\omega \cdot (1 - \mathcal{P}(P(x)))\right) \tag{45}$$

Da $P(x)$ und $Q(y)$ komplett symmetrisch in der Formel $P(x) \vee Q(y)$ stehen, kann die Gleichung 45 problemlos auf $\mathcal{P}(P(x))$ übertragen werden, nämlich:

$$\mathcal{P}(P(x)) = \text{sigmoid}\left(\omega \cdot (1 - \mathcal{P}(Q(y)))\right) \tag{46}$$

Setzen wir die Gleichung 46 in die Gleichung 45 ein, bekommen wir eine rekursive Darstellung über $\mathcal{P}(Q(y))$.

$$\mathcal{P}(Q(y)) = \text{sigmoid}\left(\omega \cdot (1 - \text{sigmoid}\left(\omega \cdot (1 - \mathcal{P}(Q(y)))\right))\right) \tag{47}$$

Hier können wir den Fixpunktsatz von Banach verwenden, um zu zeigen, dass $\mathcal{P}(Q(y))$ gegen einen eindeutig bestimmten Fixpunkt konvergiert.

6.2 Fixpunkt von S. Banach

Wir führen zunächst einige Begriffe ein.

Definition 13. *Eine Teilmenge U eines normierten Raumes X heißt abgeschlossen, wenn sie alle Grenzelemente konvergenter Folgen aus U enthält.*

Per Definition des Vollständigkeits-Begriffes gilt dann

Theorem 5. *Jede abgeschlossene Teilmenge einer vollständigen Teilmenge eines normierten Raumes ist vollständig.*

Definition 14. *Sei U Teilmenge eines normierten Raumes X. Ein Operator $\varphi : U \to X$ heißt Kontraktionsoperator, falls eine Zahl k ∈[0,1) existiert, sodass*

$$d\left(\varphi(x), \varphi(y)\right) \leq k \cdot d(x, y), \ \forall x, y \in U$$

Jede derartige Zahl k heißt Kontraktionszahl von φ.

Man beachte, dass der Operator φ nicht zwingend linear sein muss. Aus der Ungleichung in dieser Definition folgt:

Theorem 6. *Ein Kontraktionsoperator ist stetig.*

Theorem 7. *Jedes Element x eines normierten Raumes (X,d) mit der Eigenschaft*

$$\varphi(x) = x$$

heißt Fixpunkt des Operators φ: $U \subset X \to X$.

Wir können nun den zentralen Satz dieses Abschnittes formulieren.

Theorem 8. *[Wer11] (Fixpunktsatz von Banach)*
Gegeben seien ein vollständiger metrischer Raum (X, d) und eine nicht leere, abgeschlossene Menge $U \subset X$. Sei φ: $U \to U$ eine Kontraktion mit Kontraktionszahl $k \in [0, 1)$. Das bedeutet, es gilt $d\left(\varphi(x), \varphi(y)\right) \leq k \cdot d(x, y)$ für alle $x, y \in U$.

Dann besitzt φ genau einen Fixpunkt, d.h. es existiert genau ein $x \in U$ mit $\varphi(x) = x$.

Außerdem sei die Folge $(x_n)_{n\in\mathbb{N}}$ iterativ definiert durch $x_{n+1} = \varphi(x_n)$ für einen beliebigen Startwert x_0 aus U.

Unter den obigen Voraussetzungen gilt: Es existiert genau ein $\tilde{x} \in U$, sodass $\varphi(\tilde{x}) = \tilde{x}$ ist. Für alle $x_0 \in U$ gilt außerdem $\lim_{n\to\infty} x_n = \tilde{x}$.

Die Abbildung φ besitzt also einen eindeutig bestimmten Fixpunkt und dieser stimmt für alle Startwerte der oben angegebenen Iterationsvorschrift mit dem Grenzwert der Iteration überein.

Wir definieren die Funktion $\varphi : x \to \text{sigmoid}\left(\omega \cdot (1 - \text{sigmoid}\left(\omega \cdot (1 - x)\right))\right), \mathbb{R} \longmapsto \mathbb{R}$, die auf der rekursiven Darstellung von $\mathcal{P}(Q(y))$ basiert (siehe Gleichung 47).

Sei $(\mathbb{R}, d)$ ein Raum. Mathepedia 2010 zeigte, dass man das Cauchysche Konvergenzkriterium nutzen kann, um zu zeigen, dass die reellen Zahlen $\mathbb{R}$ mit der Betragsmetrik $d(x, y) = |x - y|$ ein vollständiger metrischer Raum sind [Mat10]. Zugleich sind die reellen Zahlen $\mathbb{R}$ trivialerweise nicht leer und abgeschlossen.

Des Weiteren müsste $|\varphi(x) - \varphi(y)| \leq k \cdot |x - y|$ mit $k \in [0, 1)$ für alle $x, y \in \mathbb{R}$ gelten. Für alle $\omega \in [-2, 2]$ kann man es leicht graphisch beweise, dass es so ein $k \in [0, 1)$ gibt. Für weitere $\omega \in \mathbb{R} \setminus [-2, 2]$ fehlt uns noch ein formeller Beweis. Die Klärung dieses offenen Punktes ist von Interesse und kann Gegenstand zukünftiger Forschung sein. Angenommen, die Funktion φ findet unter dem variierenden Gewicht ω eine Kontraktionszahl $k(\in [0, 1))$.

Da $\varphi(x)$ eine Wahrscheinlichkeitsfunktion in DA-MLN ist und innerhalb $(0, 1)$ liegen muss, existiert mindestens ein $k(\in (0, 1))$ als Kontraktionszahl.

Da wir alle Bedingungen vom Fixpunktsatz von Banach erfüllen, können wir sagen, dass es einen eindeutig bestimmten Fixpunkt für die Funktion φ gibt. Der Fixpunkt $\tilde{x}$ wird durch folgende Gleichung bestimmt.

$$\tilde{x} = \text{sigmoid}\left(\omega \cdot (1 - \text{sigmoid}\left(\omega \cdot (1 - \tilde{x})\right))\right) \tag{48}$$

Außerdem sei die Folge $(x_n)_{n\in\mathbb{N}}$ iterativ definiert durch $x_{n+1} = \varphi(x_n)$ für einen beliebigen Startwert x_0 aus U. Es existiert genau ein $\tilde{x} \in U$, sodass $\varphi(\tilde{x}) = \tilde{x}$ ist. Für alle $x_0 \in U$ gilt außerdem $\lim_{n\to\infty} x_n = \tilde{x}$. Die Abbildung φ besitzt also einen eindeutig bestimmten Fixpunkt und dieser stimmt für alle Startwerte der Iteration mit dem Grenzwert der Iteration überein.

Damit haben wir bewiesen, dass die Wahrscheinlichkeit von $Q(y)$ gegen den Fixpunkt $\tilde{x}(\in (0,1))$ konvergiert. Der Grenzwert lässt sich von einem konkreten Gewicht ω durch die Gleichung 48 bestimmen. Dieses Konvergenzverhalten wirkt auf allen Gewichten $\omega \in \mathbb{R}$.

7 Anwendungsfälle für die gewichtsabhängigen Wahrscheinlichkeiten von DA-MLN

Das Ziel dieser Arbeit ist es zu untersuchen, ob die gewichtsabhängigen asymptotischen Wahrscheinlichkeiten von DA-MLN uns helfen können, besser zu verallgemeinern, wenn Trainings- und Testdomänen unterschiedlich groß sind. Insbesondere sind wir an so einer Situation interessiert, in der die Trainingsdomäne im Vergleich zu den Testdomänen viel kleiner ist - ein Szenario, das aufgrund der hohen Kosten der Datenbeschaffung am häufigsten im echten Leben erwartet wird.

Um diese Fragen zu beantworten, sind drei Anwendungsfälle hier aufgelistet, bei denen der Tester Mittal et al. (2019) die Größe der Trainingsdaten festlegten und die Größe der Testdaten variierten. Dann verglichen sie die Leistung von drei konkurrierenden Algorithmen:
(a) DA-MLN (aktuelle Arbeit)
(b) (Standard) MLN [2]
(c) Adaptive MLN (AMLN) [DJ10, S.937-S.942].

Sie verglichen jeden Algorithmus in zwei verschiedenen Dimensionen:

(1) AUC: Fläche unter der Kurve / Area Under the Curve(AUC)

ROC-Kurven (Receiver-Operating-Characteristics) geben einen Überblick über die Güte eines diagnostischen Tests. In ihnen werden für verschiedene Cut-Off-Werte - in der Regel wird jeder Messpunkt verwendet - die Richtig-Positiv-Rate der Falsch-Positiv-Rate gegenübergestellt. Die Richtig-Positiv-Rate (TPF) entspricht der Sensitivität, die Falsch-Positiv-Rate (FPF) der Differenz 1-Spezifität (man beachte die umgekehrte Skalierung des x-Achse, wenn die Spezifität angegeben wird).

Die Fläche, die eine ROC-Kurve mit den Achsen des Koordinatensystems bildet, heißt AUC. Die AUC zu maximieren ist eines der wichtigsten Ziele der statischen relationalen Modelle, da die Größe der Fläche bedeutet, wie viel groß die True-Positive-Rate größer als die False-Positive-Rate ist. Wenn AUC gleich 0,5 ist, sind die beiden Raten gleich.

(2) Durchschnittliche CLL (Conditional Log-likelihood): Durchschnitt über logarithmische Grenzwahrscheinlichkeiten von Abfragen.

Die Likelihood-Funktion (oft einfach nur Likelihood), gelegentlich auch Plausibilitätsfunktion, oder Mutmaßlichkeitsfunktion genannt, ist eine spezielle reellwertige Funktion in der mathematischen Statistik, die aus einer Wahrscheinlichkeitsdichtefunktion oder einer Zähldichte gewonnen wird, indem man einen Parameter der Dichte als Variable behandelt. Zentrale Verwendung der Likelihood-Funktion ist die Konstruktion von Schätzfunktionen durch die Maximum-Likelihood-Methode. Zudem werden aus ihr weitere Funktionen wie die Log-Likelihood-Funktion und die Score-Funktion abgeleitet, die beispielsweise als Hilfsfunktionen bei der Maximum-Likelihood-Methode oder zur Konstruktion von Optimalitätskriterien in der Schätztheorie verwendet werden.

Die genaue Definition sieht wie folgendermaßen aus. Gegeben sei eine Wahrscheinlichkeitsdichtefunktion oder eine Zähldichte: $f\colon \mathbb{R}^n \to \mathbb{R}$, welche noch zusätzlich von einem oder mehreren Parametern ϑ aus einer Parametermenge Θ abhängt. Es ist also $f = f_\vartheta(x)$. Dann heißt die Funktion: $L\colon \Theta \to \mathbb{R}$, die durch:$L_x(\vartheta) = f_\vartheta(x)$ definiert wird, die Likelihood-Funktion. Die Dichtefunktion wird somit zur Likelihood-Funktion, indem man den Parameter ϑ als Variable auffasst und die Variable x als Parameter behandelt. Wird ein konkretes $\tilde{x} \in \mathbb{R}^n$ fixiert, so nennt man auch $L_{\tilde{x}}(\vartheta)$ die Likelihood-Funktion zum Beobachtungswert $\tilde{x}$. Im Falle einer Zähldichte gibt die $L_{\tilde{x}}(\vartheta)$ somit die Wahrscheinlichkeit von $\tilde{x}$ an bei gegebenem Parameter ϑ. [Rü14]

7.1 Verbreitung von Krebs und Rauchen im Freundeskreis

Der erste Anwendungsfall handelt sich um Verbreitung von Krebs und Rauchen im Freundeskreis. Wir nennen die Datenbank Freunde & Raucher (FS).

Diese Datenbank enthält Informationen über die Rauchgewohnheiten von Menschen, ihre Freundschaftsbeziehungen und ob sie an Krebs erkrankt sind oder nicht. Die Datensätze enthalten die (DA)-MLN-Theorie mit drei Prädikaten: Raucher(Person), Krebs(Person), Freunde(Person A, Person B). Die Tester haben echte Daten zur Modellierung dieser Verhältnisse in realer Welt gesammelt.

In diesem Rahmen könnten die Formeln beispielsweise so sein:
Raucher(Person) $\Longrightarrow$ Krebs(Person), ein Element aus dem ersten Fragment Abschnitt 5.1, oder
Raucher(Person A) $\wedge$ Freunde(Person A, Person B) $\Longrightarrow$ Raucher(Person B).

Die gesamte Population der Größe n wurde zunächst in $\sqrt{n}$ Gruppen geteilt. Jede Gruppe wurde nach dem Zufallsprinzip als "rauchend" oder "nicht rauchend" bezeichnet, wobei die Wahrscheinlichkeit der Rauchergruppe in den gesamten Gruppen 0,3 betrug.

Die Wahrscheinlichkeit, innerhalb der gleichen Gruppe Freunde zu sein, wurde auf 0,8 gesetzt, und die Wahrscheinlichkeit, außerhalb der Gruppe Freunde zu sein, wurde auf 0,1 gesetzt. Jede Person wurde nach dem Zufallsprinzip "Raucher" oder "Nichtraucher", abhängig von der Gruppenzugehörigkeit der Rauchgewohnheiten.

Für die Rauchergruppen wurde die Wahrscheinlichkeit des Rauchens auf 0,7 gesetzt. Für eine Nichtrauchergruppe wurde sie auf 0,1 gesetzt.

Bei einer rauchenden Person wurde die Wahrscheinlichkeit, an Krebs zu erkranken, auf 0,5, bei einer Nichtraucherin auf 0,1 gesetzt.

Wir lernten auf zufällig generierten Datensätzen mit den Domänengrößen 20, 40, 60, 80, 100 und leiteten statistische Vorhersagen für zufällig generierte Datensätze mit Domänengrößen zwischen 50 und 500 ab.

Die Testergebnisse sehen folgendermaßen aus.

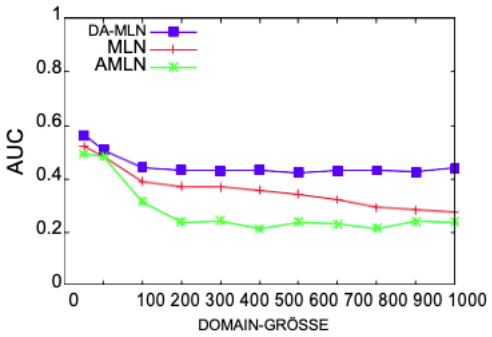

Abb. 8. AUC Leistungen von FS mit variierenden Domänengrößen

Anhand der beiden Graphen sehen wir, dass die ROC-Kurve von DA-MLN die größte Fläche mit beiden Achsen bildet. Das heißt, dass dieses AUC am größten ist, und dass die Wahrscheinlichkeiten von DA-MLN mit zunehmender Domänengröße am stärksten von den Formelgewichten abhängen, im Vergleich zu den anderen beiden statistischen Modellen.

Diese Tendenz weist darauf hin, dass dieses DA-MLN sehr wahrscheinlich gewichtsabhängige asymptotische Wahrscheinlichkeiten hat, wenn die Domänengröße gegen unendlich geht. Die Korrektheit der Vorhersagen ist auch näher zur Realität.

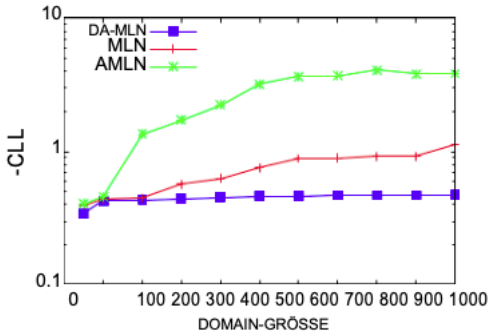

Abb. 9. CLL Leistungen von FS mit variierenden Domänengrößen

7.2 Hyperlinks und Klassifizierung von Uni-Webseiten

Der zweite Anwendungsfall handelt sich um Charakteristika von Uni-Webseiten, z.B. die Länge der Inhalte, die Hyperlinks und die Klassifizierung von den Webseiten. Die Datensätze kommen aus der Alchemy-Webseite [CSE19]. Wir nennen die Datenbank WebKB.

Die Datenbank WebKB enthält Informationen über öffentlichen Webseiten verschiedener US-Universitäten. Die Tester haben mit einer Datenmenge von etwa 1300 Webseiten von einer der Universitäten gearbeitet.

Ein Datensatz ist durch drei Prädikate definiert: Hat(Webseite, Länge der Webseite), Links(Webseite, Webseite) und Klasse(Webseite, Kategorie). Jede Webseite besteht aus Wörtern und Hyperlinks. Jede Webseite kann zu einer Untergruppe von Kategorien gehören: Person, Student, Fakultät, Professor, Abteilung, Forschungsprojekt und Kurs.

In diesem Rahmen könnte eine Formel beispielsweise so aussehen:
Klasse(Webseite A, Kategorie X) $\wedge$ Links(Webseite A, Webseite B) $\Longrightarrow$ Klasse(Webseite C, Kategorie B), ein Element aus dem zweiten Fragment in Abschnitt 5.2.

Um Datengruppen mit unterschiedlichen Größen zu erstellen wählten die Tester nach dem Zufallsprinzip eine Untergruppe von Webseiten und die zugehörigen Einrichtungen aus. Die Tester lernten auf zwei Untergruppen, die nach dem Zufallsprinzip 50 bis 100 ausgewählte Webseiten haben.

Die Tester testeten die Untergruppen von Daten, die 500 bis 800 zufällig ausgewählte Webseiten (anders als beim Training) haben. Außerdem betrachteten die Tester ein Prädikat beim Lernen und Ableiten der Daten.

Die Testergebnisse sehen folgendermaßen aus.

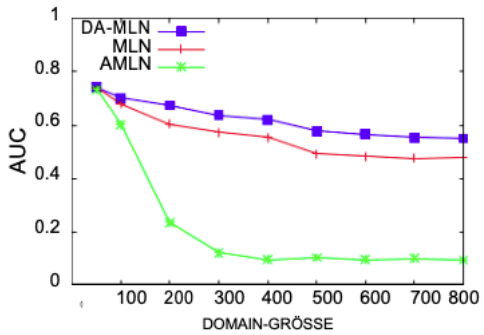

Abb. 10. AUC Leistungen von WebKB mit variierenden Domänengrößen

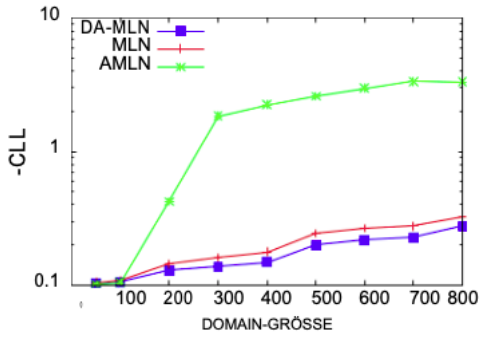

Abb. 11. CLL Leistungen von WebKB mit variierenden Domänengrößen

Anhand der beiden Graphen sehen wir, dass die ROC-Kurve von DA-MLN die größte Fläche mit beiden Achsen bildet. Das heißt, dass dieses AUC am größten ist, und dass die Wahrscheinlichkeiten von DA-MLN mit zunehmender Domänengröße am stärksten von den Formelgewichten abhängen, im Vergleich zu den anderen beiden statistischen Modellen.

Diese Tendenz weist darauf hin, dass dieses DA-MLN sehr wahrscheinlich gewichtsabhängige asymptotische Wahrscheinlichkeiten hat, wenn die Domänengröße gegen unendlich geht. Die Korrektheit der Vorhersagen ist auch näher zur Realität.

7.3 Arbeitsbeziehungen in Filmindustrie

Der zweite Anwendungsfall handelt sich um Zuordnung von Mitarbeitern in Filmindustrie. Die Datensätze kommen aus einem Online-Kaggle- Wettbewerb [Inc19] mit echten Daten zwischen 2006 und 2016. Wir nennen die Datenbank IMDB.

Die Datensätze enthalten Informationen über 1000 Filme und ihre Besetzungen. Diese Datenbank ist durch vier Prädikate definiert: Schauspieler(Person), Direktor(Person), Film(Titel, Person) und ArbeitenUnter(Person A, Person B).

In diesem Rahmen könnte eine Formel beispielsweise so aussehen:
ArbeitenUnter(Person A, Person B) $\wedge$ Director(Person A) $\wedge$ Film(Titel C, Person A) $\Longrightarrow$ Film(Titel C, Person B). Obwohl diese logische Formel nicht aus den vier Fragmenten kommt, ist das Ergebnis trotz für die zukünftige Forschung einleuchtend.

Um Datengruppen mit unterschiedlichen Größen zu schaffen, wählten die Tester nach dem Zufallsprinzip eine Untergruppe von Direktoren aus, und suchten die von ihnen gedrehten Filme und die Schauspieler, die in diesen Filmen mitgearbeitet haben.

Die Tester erfuhren von 4 zufällig generierten Untermengen mit 2, 4, 5 und 10 Direktoren und traffen Vorhersagen für zufällig generierte Untermengen mit Direktoren zwischen 10 und 50. Es wurde sichergestellt, dass es keine Überschneidungen zwischen den Trainings- und Testdaten von Direktoren gab.

Die Testergebnisse sehen folgendermaßen aus.

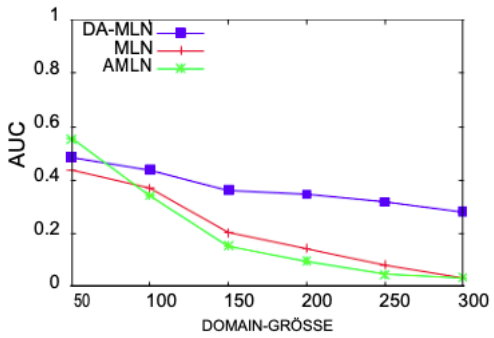

Abb. 12. AUC Leistungen von IMDB mit variierenden Domänengrößen

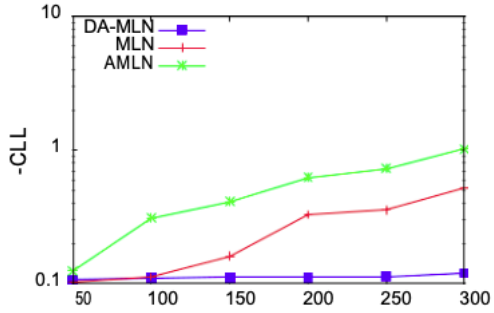

Abb. 13. CLL Leistungen von IMDB mit variierenden Domänengrößen

Anhand der beiden Graphen sehen wir, dass die ROC-Kurve von DA-MLN die größte Fläche mit beiden Achsen bildet. Das heißt, dass dieses AUC am größten ist, und dass die Wahrscheinlichkeiten von DA-MLN mit zunehmender Domänengröße am stärksten von den Formelgewichten abhängen, im Vergleich zu den anderen beiden statistischen Modellen.

Diese Tendenz weist darauf hin, dass dieses DA-MLN sehr wahrscheinlich gewichtsabhängige asymptotische Wahrscheinlichkeiten hat, wenn die Domänengröße gegen unendlich geht. Die Korrektheit der Vorhersagen ist auch näher zur Realität.

8 Zusammenfassung und Einblicke in die zukünftige Forschung

In dieser Arbeit haben wir das Problem mit der Aggregationsfunktion in (DA-)MLN mit zunehmenden Domänengröße behandelt. Wir haben das Problem, dass wenn die Domänengrößen sich dem Unendlichen annähern, die Wahrscheinlichkeiten für einige einfachen MLN-Formeln entweder zum Extrem tendieren oder zu einer Konstanten konvergieren (unabhängig vom gelernten Formelgewicht).

Als Lösung haben wir eine zusätzliche Beschränkung für den Verbindungsvektor eines DA-MLN genannt, die eine exakte Skalierung jedes Literals in einer Formel ermöglicht. Daraus sind einige Fragmente von DA-MLN entstanden, die durch Beweise garantiert gewichtsabhängige asymptotische Wahrscheinlichkeiten vorweisen, wenn die Domänengrößen sich dem Unendlichen annähern. Damit kann man in vielen Anwendungsfällen die gelernten Formelgewichte viel sinnvoller einsetzen und realitätsnahe Ergebnisse erzielen.

Außerdem gibt es noch einen offenen Punkt im Beweis für die Wahrscheinlichkeiten der Literale der Formel $P(X) \vee Q(y)$ in einem DA-MLN aus dem Kapitel 6. In der Anwendung des Fixpunktsatzes von Banach müsste man noch zeigen, dass $|\varphi(x) - \varphi(y)| \leq k \cdot |x - y|$ mit $k \in [0, 1)$ für alle $x, y \in \mathbb{R} \setminus [-2, 2]$ gilt. Die Klärung dieses offenen Punktes vervollständigt den Gesamtbeweis im Kapitel 6 und kann Gegenstand zukünftiger Forschung sein.

Allerdings ist das ursprüngliche Problem bei der Epidemie-Modellierung noch nicht gelöst, da das DA-MLN-Modell für die Epidemie-Modellierung über keinen Verbindungsvektor mit identischen Einträgen verfügt. Die Verbindungsanzahl zwischen den Literalen einer Formel ist leider wegen dem komplexen Sachverhalt sehr unterschiedlich. Diese Arbeit hat jedoch eine einleuchtende Richtung für die asymptotischen gewichtsabhängigen Wahrscheinlichkeiten von DA-MLN für die Epidemie-Modellierung gezeigt, nämlich alle Literale einer Formel auf einem geeigneten Niveau zu skalieren.

Für die zukünftige Forschung könnte man noch einen Blick in die Auswahl der Aggregationsfunktion werfen, da die aktuellen Aggregationsfunktionen viele Literale einer Formel entweder "überkompensieren" oder "unterkompensieren".

Zu potenziellen Themen für künftige Arbeiten gehören die Weiterentwicklung der Theorie um DA-MLN (z.B. Beweise für weitere Fragmente von DA-MLN), bessere Auswahl für Verbindungsvektor, die Durchführung zusätzlicher Experimente in neueren Bereichen oder mit anderen Ansätzen, die wir aus zeitlichen Gründen nicht untersucht haben.

Literaturverzeichnis

[Bau96] Bauer, Heinz: *Probability Theory*. De Gruyter, 1996 https://www.springer.com/de/book/9783540725367. – ISBN 978–3–11–013935–8

[CSE19] Computer Science, Paul G.Allen S. ; Engineering, University of W.: *Alchemy: Open Source AI*. http://alchemy.cs.washington.edu/. Version: 2019

[DJ10] Dominik Jain, Andreas Barthels und Michael B.: Adaptive markov logic networks: Learning statistical relational models with dynamic parameters. In: *Proceedings of the 10th European Conference on Artificial Intelligence* (2010), S. 937–942

[D.L09] D.Lowd, P. D.: Markov Logic: An Interface Layer for Artificial Intelligence. In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* (2009)

[HM19] Happy Mittal, Vibhav Gogate und Parag S. Ayush Bhardwaj B. Ayush Bhardwaj: *Domain-Size Aware Markov Logic Networks*. http://proceedings.mlr.press/v89/mittal19a/mittal19a.pdf. Version: 2019

[Inc19] Inc., Kaggle: *Internet Movie Database popular movies from 2006 to 2016*. https://www.kaggle.com/PromptCloudHQ/imdb-data/data. Version: 2019

[Mat10] Mathepedia: *Cauchysches Konvergenzkriterium*. https://mathepedia.de/Cauchysches_Konvergenzkriterium.html. Version: 2010

[Rü14] Rüschendorf, Ludger: *Mathematische Statistik*. Springer, 2014 https://link.springer.com/book/10.1007%2F978-3-642-41997-3#authorsandaffiliationsbook. – ISBN 978–3–642–41996–6

[Wei20] Weitkämper, Felix Q.: *Scaling the weight parameters in Markow logic networks and relational logistic regression models*. Lehr- und Forschungseinheit für Programmier- und Modellierungssprachen, Ludwig-Maximilians-Universität München. https://epub.ub.uni-muenchen.de/71690/1/IJCLR_submission_manuscript.pdf. Version: 2020

[Wer11] Werner, Dirk: *Funktionanalysis*. Springer, 2011 https://www.degruyter.com/view/title/3786. – ISBN 978–3–64–221016–7