

Bachelorarbeit

# Numerische Analyse des TrueSkill Verfahrens

Johannes Loevenich

Datum der Abgabe

Betreuung: Prof. Dr. Jochen Garcke

Fakultät für Mathematik  
Rheinische Friedrich-Wilhelms- Universität Bonn

# Inhaltsverzeichnis

# 1 Einleitung

Das Problem  $n$  Spieler zu bewerten ist ein wichtiger Forschungsbereich des Maschinellen Lernens.

Aus der Kombinatorik lässt sich schnell herleiten, dass es  $n!$  verschiedene Möglichkeiten gibt  $n$  Spieler in einem Ranking zu bewerten. Ziel ist es das eine Ranking zu finden, welches das tatsächliche Können der einzelnen Spieler am besten widerspiegelt. Nimmt man an, dass alle Rankings gleich wahrscheinlich sind, so würde dies bedeuten, dass  $\log_2(n!) \approx n \log_2(n)$  Spielausgänge nötig wären, um die korrekte Bewertung zu ermitteln. Diese Schranke ist jedoch nur dann aussagekräftig, falls die Ausgänge eines jeden Spiels gleichverteilt sind. Bei vielen Spielen wird durch das matchen von nahezu gleich starken Gegnern versucht diese Chancengleichheit zu erzwingen. In dieser Arbeit wird eine Wahrscheinlichkeitstheoretische Interpretation dieses Problems betrachtet, weshalb es im weiteren Verlauf von Bedeutung ist gewisse Unsicherheiten in Rang eines Spielers zu berücksichtigen. Interessanterweise reduziert sich die minimale Anzahl von aussagekräftigen Spielen auf  $n \log_2(m)$ , wenn wir annehmen, dass es  $m \ll n$  Äquivalenzklassen oder Level gibt.

Betrachten wir Spiele bei denen  $k$  Teams gegeneinander antreten und bewertet werden, dann erfüllt jedes Spiel  $\log_2(k)$  Ausgänge und es werden nur  $\frac{n \log_2(n)}{\log_2(k!)}$  aussagekräftige Spielausgänge benötigt. Man wird sehen, dass das hier vorgestellte Verfahren für das betrachtete Problem nahezu optimal ist, also bis auf geringe Abweichungen gegen die oben genannten Schranken konvergiert.

## 2 Das Ranking Problem

In dieser Arbeit werden Spiele betrachtet, bei denen mehrere Teilnehmende Teams bewertet werden sollen. Falls zwei Spieler oder Teams dieselbe Bewertung erhalten, sagen wir das das Spiel unentschieden ausgegangen ist. Im Spezialfall von zwei Teams gibt es genau drei mögliche Spielausgänge: Sieg, Niederlage oder unentschieden.

Nummerieren wir alle teilnehmenden Spieler eines Spiels von Eins bis  $n$ , so kann ein Spiel zwischen  $k$  Teams vollständig durch die  $k$  Indizes  $i_j \in \{1, \dots, n\}$  der  $n_j$  Spieler im  $j$ -ten Team beschrieben werden. Der von jedem Team erreichte Rang sei dann definiert als  $\mathbf{r} := (r_1, \dots, r_k)^T \in \{1, \dots, k\}^k$ . Es wird davon ausgegangen, dass der Gewinner eines Spiels den Rang 1 erhält.

Gesucht wird nach dem Skill  $s_i$  eines jeden Spielers,  $\mathbf{s} \in \mathbb{R}^n$ ; wobei der Skill des  $j$ -ten Teams eine Funktion  $S(s_{i_j})$  in Abhängigkeit aller Skills der Spieler des Teams ist. Im trivialen Fall, dass jedes Team nur einen Spieler enthält beschreibt  $S(s_i) = s_i$  also die Identität. Für Skills wird die folgende Eigenschaft gefordert.

**Definition 2.1. (Stochastische Transitivität)** Falls ein Team  $u$  vor einem Team  $v$  platziert wurde, so ist es wahrscheinlicher, dass Team  $u$  gegen Team  $v$  gewinnt als umgekehrt.

$$S(s_{i_u}) \geq S(s_{i_v}) \Rightarrow P(\text{Team } u \text{ gewinnt}) > P(\text{Team } v \text{ gewinnt}) \quad (2.1)$$

## 3 Wahrscheinlichkeitstheorie

Dieses Kapitel gibt einen Überblick über einige wichtige Konzepte der Wahrscheinlichkeitstheorie. Im Folgenden werden Mengen mit Großbuchstaben, z.B.  $X$ , und deren Elemente mit Kleinbuchstaben, z.B.  $x$ , bezeichnet. Für Mengen definiere die Indikatorfunktion  $I_X$  durch

$$I_X(x) := \begin{cases} 0 & \text{falls } x \notin X \\ 1 & \text{falls } x \in X \end{cases}$$

**Definition 3.1. ( $\sigma$ -Algebra)** Sei eine Menge  $\chi$  gegeben. Ein Mengensystem  $\mathcal{Y}$  von Mengen  $X \subseteq \chi$  wird genau dann  $\sigma$ -Algebra über  $\chi$  genannt, wenn

1. Ist eine Menge  $X \in \mathcal{Y}$  enthalten, so auch ihr Komplement  $X^c = \chi \setminus X$ .
2. Falls  $X_i \in \mathcal{Y}$ ,  $i = 1, \dots, \infty$  abzählbares Mengensystem in  $\mathcal{Y}$ , dann sind auch  $\bigcup_{i=1}^{\infty} X_i \in \mathcal{Y}$  und  $\bigcap_{i=1}^{\infty} X_i \in \mathcal{Y}$  in  $\mathcal{Y}$  enthalten.

Kurz, jede  $\sigma$ -Algebra ist abgeschlossen unter Komplementbildung und abzählbaren Vereinigungen oder Schnitten.

**Definition 3.2. (Borelmengen)** Sei  $\chi = \mathbb{R}^n$ , die Borelmengen  $B_n$  sind die kleinsten  $\sigma$ -Algebren, die alle offenen Intervalle

$$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \forall i \in \{1, \dots, n\} : x_i \in (a_i, b_i)\}$$

für alle  $a_i, b_i \in \mathbb{R}$ . Bemerke, dass  $B_n$  überabzählbar ist.

**Definition 3.3. (Maß- und Wahrscheinlichkeitsraum)** Ein messbarer Raum ist ein Tupel  $(\chi, \mathcal{Y})$ . Dabei ist  $\chi$  das Universum und  $\mathcal{Y}$   $\sigma$ -Algebra über  $\chi$ . Ein Wahrscheinlichkeitsraum ist ein Tripel  $(\chi, \mathcal{Y}, P)$ , wobei  $P$  ein Wahrscheinlichkeitsmaß auf  $\chi$  ist. D.h.  $P : \mathcal{Y} \rightarrow [0, 1]$ , sodass  $P(\chi) = 1$  und für disjunkte abzählbare Vereinigungen  $X_i \in \mathcal{Y}$ ,  $i = 1, \dots, \infty$  gilt

$$P(\cup_{i=1}^{\infty} X_i) = \sum_{i=1}^{\infty} P(X_i).$$

**Definition 3.4. (Messbarkeit)** Sei  $(\chi, \mathcal{Y})$  messbarer Raum. Eine reellwertige Funktion  $g : \chi \rightarrow \mathbb{R}$  heißt  $\mathcal{Y}$ -messbar (oder messbar) genau dann, wenn

$$\forall z \in \mathbb{R} : \{x \in \chi | g(x) \leq z\} \in \mathcal{Y}.$$

**Definition 3.5. (Zufallsvariable)** Sei  $(\chi, \mathcal{Y})$  messbarer Raum. Eine Zufallsvariable ist eine  $\chi$ -messbare reellwertige Funktion  $f : \chi \rightarrow \mathbb{R}$ .

Eine Zufallsvariable  $Y = f(X)$  induziert also ein Maß  $P_Y$  auf  $\mathbb{R}$ , für das die  $\sigma$ -Algebra  $\mathcal{A}$  die Intervalle der Form  $(-\infty, z] | z \in \mathbb{R}$  enthält. Das Maß  $P_Y$  ist vom Maß  $P_X$  und  $f$  induziert. Das bedeutet

$$\forall Y \in \mathcal{B}_1 : P_Y(Y) := P_X(\{x \in \chi | f(x) \in Y\}).$$

**Definition 3.6. (Verteilungsfunktion und Dichte)** Für eine Zufallsvariable  $X$  heißt die durch

$$F_X(x) := P_X(X \leq x)$$

definierte Funktion  $F_X : \mathbb{R} \rightarrow [0, 1]$  Verteilungsfunktion von  $X$ . Die Funktion  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  wird Dichte genannt, falls

$$\forall z \in \mathbb{R} : F_X(z) = \int_{x \leq z} f_X(x) dx.$$

Für weitere Betrachtungen ist Erwartung einer Zufallsvariablen von essentieller Bedeutung.

**Definition 3.7. (Erwartungswert)** Sei  $f : \chi \rightarrow \mathbb{R}$  messbare Funktion. Der Erwartungswert  $E_X[f(X)]$  von  $f$  über die Wahrscheinlichkeit von  $x$  wird durch

$$E_X[f(X)] := \int_{\mathbb{R}} f(x) dF_X(x)$$

definiert. Der Erwartungswert ist nur dann definiert, wenn  $\int_{\mathbb{R}} |f(x)| dF_X(x) < \infty$ .

**Definition 3.8. (Varianz)** Die Varianz  $\text{Var}(X)$  einer Zufallsvariable  $X$  ist definiert durch

$$\text{Var}(X) := E_X[(X - \mu)^2] = E_X[X^2] - \mu^2,$$

wobei  $\mu = E_X[X]$  der Erwartungswert der Zufallsvariable  $X$  ist.

**Definition 3.9. (Produkttraum)** Seien zwei Maßräume  $(\mathcal{X}, \mathcal{Y})$  und  $(\varphi, \Phi)$  gegeben. Definiere den Produkttraum durch  $(\mathcal{X} \times \varphi, \mathcal{Y} \times \Phi)$ . Hierbei bezeichnet  $\mathcal{Y} \times \Phi$  die kleinste  $\sigma$ -Algebra welche die Mengen  $\{X \times Y | X \in \mathcal{Y}, Y \in \Phi\}$  enthält.

**Definition 3.10. (Marginal und bedingte Wahrscheinlichkeiten)** Sei  $(\mathcal{X} \times \varphi, \mathcal{Y} \times \Phi, P_{XY})$  der Produkttraum von  $X$  und  $Y$ . Das Marginalwahrscheinlichkeitsmaß  $P_X$  ist dann durch

$$\forall X \in \mathcal{Y} : P_X(X) := P_{XY}(X \times \varphi)$$

definiert. Sei  $Y \in \Phi$  und  $P_Y(Y) > 0$ , dann ist das bedingte Wahrscheinlichkeitsmaß  $P_{X|Y \in \mathcal{Y}}$  durch

$$\forall Y \in \mathcal{Y} : P_{X|Y}(X) := \frac{P_{XY}(X \times Y)}{P_Y(Y)}$$

gegeben.

**Definition 3.11. Unabhängigkeit** Zwei Zufallsvariablen  $X$  und  $Y$  werden genau dann unabhängig genannt, wenn

$$\forall X \in \mathcal{Y} : \forall Y \in \Phi : P_{XY}(X \times Y) = P_X(X)P_Y(Y).$$

In diesem Fall genügen die Marginalverteilungen um den gesamten Produkttraum zu definieren.

Im Folgenden schreibe  $\mathbf{X}$ , falls  $\mathbf{X}$  eine Folge  $(X_1, \dots, X_n)$  von Zufallsvariablen definiert. Eine solche Folge kann je nach Kontext entweder als Zeilen- oder als Spaltenvektor interpretiert werden. Ein Element des Universums  $\mathcal{X}^n$  wird dann durch ein  $n$ -Tupel  $\mathbf{x}$  beschrieben. Sei  $\mathbf{x} = (x_1, \dots, x_n)$  ein solches  $n$ -Tupel, dann verstehe  $x \in \mathbf{x}$  als  $\exists i \in \{1, \dots, n\} : x_i = x$ .

**Definition 3.12. (Erwartungswert einer  $n$ -dimensionalen Zufallsvariable)** Seien  $\mathbf{X} = (X_1, \dots, X_n)$   $n$  Zufallsvariablen mit gemeinsamem Wahrscheinlichkeitsmaß  $P_X$ , dann ist die Erwartung  $E_{\mathbf{X}}[\mathbf{X}]$  durch das  $n$ -Tupel

$$E_{\mathbf{X}}[\mathbf{X}] = (E_{X_1}[X_1], \dots, E_{X_n}[X_n])$$

gegeben.

**Definition 3.13. (Kovarianz und Kovarianzmatrix)** Seien  $X$  und  $Y$  zwei Zufallsvariablen mit gemeinsamem Wahrscheinlichkeitsmaß  $P_{XY}$ , dann ist die Kovarianz  $Cov(X, Y)$  durch

$$Cov(X, Y) := E_{XY}[(X - \mu)(Y - \nu)]$$

definiert, wobei  $\mu = E_X[X]$  und  $\nu = E_Y[Y]$ . Es ist leicht einzusehen, dass  $Cov(X, X) = Var(X)$  Sei  $\mathbf{X} = (X_1, \dots, X_n)$  eine Folge von  $n$  Zufallsvariablen und  $\mathbf{Y} = (Y_1, \dots, Y_n)$  eine

weitere solche Folge mit gemeinsamer Dichte  $P_{XY}$ , dann ist die  $n \times m$  Kovarianzmatrix  $\text{Cov}(\mathbf{X}, \mathbf{Y})$  definiert durch

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) := \begin{pmatrix} \text{Cov}(X_1, Y_1) & \dots & \text{Cov}(X_1, Y_m) \\ \vdots & \dots & \vdots \\ \text{Cov}(X_n, Y_1) & \dots & \text{Cov}(X_n, Y_m) \end{pmatrix}$$

Falls  $\mathbf{X} = \mathbf{Y}$   $\text{Cov}(\mathbf{X}, \mathbf{X}) = \text{Cov}(\mathbf{X})$

### 3.1 Eigenschaften von Zufallsvariablen

**Satz 3.14. (Erwartungswert von Summen und Produkten)** Seien  $X$  und  $Y$  zwei unabhängige Zufallsvariablen, so gilt

$$E_{XY}[X \cdot Y] = E_X[X] \cdot E_Y[Y], \quad (3.1)$$

$$E_{XY}[X + Y] = E_X[X] + E_Y[Y], \quad (3.2)$$

immer dann, wenn die Terme auf der rechten Seite existieren. Die zweite Aussage gilt sogar, falls  $X$  und  $Y$  nicht unabhängig sind.

**Satz 3.15. (Linearität des Erwartungswertes)** Für jede  $n$ -dimensionale Zufallsvariable  $\mathbf{X}$ , jede Matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  und jeden stationären Vektor  $\mathbf{b} \in \mathbb{R}^m$  gilt

$$E_X[\mathbf{A}\mathbf{X} + \mathbf{b}] = \mathbf{A}E_X[\mathbf{X}] + \mathbf{b}.$$

**Satz 3.16. Varianz Zerlegung** Seien  $X$  und  $Y$  zwei unabhängige Zufallsvariablen, dann gilt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

**Beweis 1.** Setze  $\mu = E_X[X]$  und  $\nu = E_Y[Y]$ . Benutze die Definition der Varianz

$$\begin{aligned} E_{XY}[(X + Y - E_{XY}[X + Y])^2] &= E_{XY}[((X - \mu)(Y - \nu))^2] \\ &= E_{XY}[(X - \mu)^2 + 2 \cdot (X - \mu)(Y - \nu) + (Y - \nu)^2] \\ &= E_X[(X - \mu)^2] + 2E_{XY}[(X - \mu)(Y - \nu)] + E_Y[(Y - \nu)^2] \\ &= E_X[(X - \mu)^2] + 2E_X[(X - \mu)]E_Y[(Y - \nu)] + E_Y[(Y - \nu)^2] \\ &= \text{Var}(X) + \text{Var}(Y) \end{aligned}$$

**Lemma 3.17.** Für jede Zufallsvariable  $X$  und jede Konstant  $c \in \mathbb{R}$  gilt  $\text{Var}(cX) = c^2 \cdot \text{Var}(X)$ .

## 3.2 Mehrdimensionale Gauß Verteilung

In diesem Abschnitt werden einige der wichtigsten Eigenschaften der Gaussverteilung zusammengefasst.

**Definition 3.18.** Sei  $\mu \in \mathbb{R}^n$  ein Vektor und  $A \in \mathbb{R}^{n \times m}$  eine deterministische Matrix. Sei weiter  $Y = (Y_1, \dots, Y_m)$  eine Folge von  $m$  unabhängigen, normalverteilten Zufallsvariablen  $Y_i$  mit Mittelwert Null und Einheitsvarianz ( $Y_i \sim \text{Normal}(0, 1)$ ). Dann wird  $X = AY + \mu$  normal- oder gaußverteilt mit Erwartungswert  $E_X[X] = \mu$  und Kovarianzmatrix  $\text{Cov}(X) = \Sigma = AA'$  genannt. Da das Maß  $P_X$  eindeutig durch diese beiden Größen beschrieben wird schreiben wir im Weiteren auch  $Y \sim \text{Normal}(\mu, \Sigma)$ .

**Satz 3.19.** Falls  $X \sim \text{Normal}(\mu, \Sigma)$ , dann besitzt  $X$  genau dann eine Dichte  $f_X$ , wenn  $\Sigma$  positiv definit ist. Die Dichte  $f_X$  ist gegeben durch

$$f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (3.3)$$

**Satz 3.20.** Sei  $X \sim \text{Normal}(\mu, \Sigma)$  eine  $n$ -dimensional normalverteilte Zufallsvariable,  $A \in \mathbb{R}^{m \times n}$  stationäre Matrix und  $b \in \mathbb{R}^m$  stationärer Vektor. Dann ist auch die Zufallsvariable  $Y = AX + b$  normalverteilt mit  $Y \sim \text{Normal}(A\mu + b, A\Sigma A^T)$ .

**Satz 3.21.** Angenommen  $P_{X|Y=y} = \text{Normal}(Xy, \Gamma)$  ist normalverteiltes Wahrscheinlichkeitsmaß, wobei  $X \in \mathbb{R}^{m \times n}$  und  $\Gamma \in \mathbb{R}^{m \times m}$  stationäre Matrizen für alle Werte  $y \in \mathbb{R}^n$  sind. Falls  $P_Y = \text{Normal}(\mu, \Sigma)$  normalverteiltes Wahrscheinlichkeitsmaß ist, so gilt

$$P_{Y|X=x} = \text{Normal}(\Psi(X^T \Gamma^{-1} x + \Sigma^{-1} \mu), \Psi). \quad (3.4)$$

$$P_X = \text{Normal}(X\mu, \Gamma + X\Sigma X^T). \quad (3.5)$$

,wobei  $\Psi = (X^T \Gamma^{-1} X + \Sigma^{-1})^{-1}$

**Beweis 2.** Nach Satz ( ) wissen wir, dass

$$f_{Y|X=x(y)} = \frac{f_{X|Y=y} f_Y(y)}{\int_{\mathbb{R}^n} f_{X|Y=y'(x)} f_Y(y') dy'} = \frac{f_{X|Y=y}(x) f_Y(y)}{f_X(x)}.$$

Es fällt auf, dass der Nenner unabhängig von  $y$  ist. Betrachte nun deshalb zunächst den Zähler. Mithilfe von Definition ist Letzteres gegeben durch

$$c \cdot \exp\left(-\frac{1}{2}((x - Xy)^T \Gamma^{-1}(x - Xy) + (y - \mu)^T \Sigma^{-1}(y - \mu))\right),$$

wobei  $c = (2\pi)^{-\frac{m+n}{2}} |\Gamma|^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}}$  unabhängig von  $x$  und  $y$  ist. Diesen Ausdruck wiederum können wir umschreiben als

$$c \cdot \exp\left(-\frac{1}{2}((y - x)^T C(y - c) + d(x))\right),$$

mit



$$C = X^T \Gamma^{-1} X + \Sigma^{-1},$$

$$Cc = X^T \Gamma^{-1} X + \Sigma^{-1} \mu,$$

$$d(x) = (x - X\mu)^T (\Gamma + X\Sigma X^T)^{-1} (x - X\mu).$$

Da  $d(x)$  als Funktion nicht von  $y$  abhängt, kann der Term  $\exp(-\frac{1}{2}d(x))$  mit in die Konstante  $c$  gezogen werden und somit folgt die erste Gleichung des Satzes indem  $\Psi := C^{-1}$  gesetzt wird.

Um die Zweite Aussage zu zeigen kann die Definition von  $f_X(x)$  benutzt werden, d.h.,

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}^n} c \cdot \exp(-\frac{1}{2}((y' - c)^T C(y' - c) + d(x))) dy' \\ &= c \cdot \exp(-\frac{1}{2}d(x)) \cdot \int_{\mathbb{R}^n} \exp(-\frac{1}{2}((y' - c)^T C(y' - c))) dy' \\ &= c \cdot \exp(-\frac{1}{2}d(x)) \cdot (2\pi)^{\frac{n}{2}} |C|^{-\frac{1}{2}} = c' \cdot \exp(-\frac{1}{2}d(x)) \\ &= c' \cdot \exp(-\frac{1}{2}(x - X\mu)^T (\Gamma + X\Sigma X^T)^{-1} (x - X\mu)), \end{aligned}$$

wobei die dritte Zeile aus der Definition und der Tatsache folgt, dass sich Wahrscheinlichkeitsdichten zu Einer zusammenfassen lassen. Dies zeigt die zweite Aussage.

## 4 Faktorgraphen

Ein Faktorgraph ist ein bipartiter Graph, der beschreibt wie eine globale Funktion in Abhängigkeit von verschiedenen Variablen und Faktoren durch ein Produkt von lokalen Funktionen ausgedrückt werden kann.

### 4.1 Allgemeines

In Faktorgraphen wird zwischen zwei Typen von Knoten unterschieden: Die einen Knoten, welche mit Variablen identifiziert werden (nicht ausgefüllte Knoten) und die Anderen, welche mit lokalen Funktionen identifiziert werden (ausgefüllte Knoten). Kanten verbinden Variablenknoten  $x_i$  und Funktionenknoten  $f$  genau dann, wenn  $x_i$  Argument von  $f$  ist.

Sei  $X = \{x_i\}_{i \in \mathbb{N}}$  eine Menge von Variablen bzgl. der Indexmenge  $N = \{1, 2, 3, \dots, n\}$ . Falls  $E$  triviale Teilmenge von  $N$  ist, so bezeichne mit  $X_E$  die Teilmenge von  $X$ , welche durch  $E$  induziert wird. Für jedes  $i \in N$  nehme die Variable  $x_i$  Werte aus dem Alphabet  $A_i$  an. Desweiteren wird angenommen, dass  $A_i$  für alle  $i \in N$  stets endlich ist. Bezeichne eine bestimmte Belegung der Variablen aus  $X$  als Konfiguration der Variablen. Diese Konfigurationen können als Kartesisches Produkt  $W = \prod_{i \in N} A_i$ , den sogenannten Konfigurationsraum verstanden werden.

Ein Element  $w = (w_1, \dots, w_n) \in W$ , mit  $w_i \in A_i$  ist entspricht der Variablenbelegung  $x_1 = w_1, \dots, x_n = w_n$ .

Im weiteren sind Funktionen mit Urbild  $W$  von besonderem Interesse. Sei  $g : W \rightarrow R$  eine solche Funktion, auch globale Funktion genannt. Im Moment beziehe sich der Wertebereich auf die Reellen Zahlen. Im Allgemeinen sei jedoch jeder beliebige Semiring erlaubt.

### 4.2 Bayssche Netwerke

Bayssche Netze sind gerichtete, azyklische graphische Modelle auf einer Menge von Zufallsvariablen. Jeder Knoten  $v$  eines Baysschen Netzwerkes wird mit einer Zufallsvariable assoziiert. Bezeichnet  $a(v)$ , die Menge der Vorgänger des Knotens  $v$ , so hat die Wahrscheinlichkeitsverteilung des Baysschen Netzwerkes die Form

$$p(v_1, v_2, \dots, v_n) = \prod_{i=1}^n p(v_i | a(v_i)) \quad (4.1)$$

Falls  $a(v_i) = \emptyset$  (d.h.  $v_i$  hat keine Vorgänger), so setze  $p(v_i | \emptyset) = p(v_i)$ . Figur **FIGUR EINFÜGEN** zeigt beispielsweise eine Bayssches Netzwerk, welches die Faktorisierung

$$p(v_1, v_2, v_3, v_4, v_5) = p(v_1 | v_2) p(v_2) p(v_3 | v_2, v_4) p(v_4) p(v_5 | v_4)$$

beschreibt.

### 4.3 Sum-Product Algorithmus

Sei  $g(X)$  globale Funktion über die Variablen der Menge  $X = \{x_i : i \in N\}$ , wobei Variable  $x_i$  Werte der endlichen Menge  $A_i$  annimmt. In diesem Abschnitt wird eine Algorithmus beschrieben, um die marginal Funktionen

$$G_i(x_i) = \sum_{x_1 \in A_1, \dots, x_{i-1} \in A_{i-1}, x_{i+1} \in A_{i+1}, \dots, x_n \in A_n} g(x_1, \dots, x_n) \quad (4.2)$$

für Variablen  $x_i, i \in N$  zu berechnen. Im Weiteren sei  $\sum_{x_i} f(x_i) = \sum_{x_i \in A_i} f(x_i)$  und genauso sei für eine Teilmenge  $J \subset N$  mit  $\sum_{x_i; i \in J} f(X)$  die Summe über alle möglichen Konfigurationen der Variablen  $x_i$  über  $J$  gemeint. Damit gilt  $G_i(x_i) = \sum_{x_j; j \in N \setminus \{i\}} g(X)$ . Die Definition der marginalen Funktion  $G$  kann nun auf eine Teilmenge  $J$  von  $N$  ausgeweitet werden.

$$G_i(x_i) = \sum_{x_j; j \in N \setminus J} g(X). \quad (4.3)$$

Falls  $g(x_1, \dots, x_n)$  eine Wahrscheinlichkeitsverteilung beschreibt, so ist  $G_i(x_i)$  die Marginalverteilung und  $G_J(X_J)$  die gemeinsame Wahrscheinlichkeitsverteilung der Variablen über die Indexmenge  $J$ .

Ist die Anzahl  $n$  der Argumente von  $g$  klein, so nutze eine alternative Kurzschreibweise für die Marginalfunktionen. Schreibe statt einem Argument  $x_i$  von  $g$  ein  $+$  um anzudeuten, dass über diese Variable summiert wird.

### 4.4 Funktionsweise des Algorithmus

Der Sum-product Algorithmus operiert mithilfe einer "message passing"Prozedur, die Produkte von lokalen Funktionen entlang der Pfade des Faktorgraphen aufammelt. Es wird angenommen, dass dieser Graph ein Baum ist, d.h. dass dieser Graph keine Kreise enthält. Die Beschreibung des Algorithmus kann durch die Annahme, dass jeder Knoten wie ein Prozessor Nachrichten über Kanten übermittelt und empfängt.

Für diese vereinfachte Betrachtung arbeitet der Algorithmus wie folgt. Die Basisoperation an jedem Knoten ermittelt das Produkt aller eingehenden Nachrichten an diesem Knoten. Für Knoten, die eine Menge von Variablen darstellen, wird dieses Produkt um die zugehörigen lokalen Funktionen erweitert. Die so ermittelten Produkte werden dann mit dem Vorbehalt, dass Nachrichten über ausgehende Knoten keine Faktoren enthalten, über die ausgehenden Kanten übermittelt. Da vorausgesetzt wurde, dass der Faktorgraph keine Kreise enthält, enthält das Produkt der ausgehenden Nachrichten einer Kante und der empfangenen Nachrichten dieser Kante alle Faktoren der globalen Funktion.

Diese sogenannte "message-passing"Prozedur wird von den Blättern des Faktorgraphs aus gestartet und iteriert über alle Knoten des Graphen. An Blättern, die Variablenmengen darstellen entspricht die ausgehende Nachricht eine Representation der lokalen

Funktion dieses Knotens. An Blättern, die eine Variable darstellen entspricht sie hingegen der Indikatorfunktion. Alle anderen Knoten des Graphen warten zunächst bis sie genügend Nachrichten gesammelt haben um eine ausgehende Nachricht zu produzieren. Genauer soll das heißen, sie warten solange, bis an jeder bis auf einer eingehenden Kante Nachrichten empfangen wurden. Tritt dieser Fall ein, so wird das Produkt aller eingehenden Nachrichten mit der lokalen Funktion gebildet und über die freie Kante übermittelt. Wird an dieser übrig gebliebenen Kante eine Nachricht empfangen, so werden die Produkte der zugehörigen lokalen Funktion über alle anderen ausgehenden Kanten versendet. Diese Prozedur wird in Figur **FIGUR EINFÜGEN!!!** illustriert.

Dieser Algorithmus wird dann effektiv, wenn man beachtet, dass von lokalen Funktionen über Pfade gesammelte Produkte marginalisiert werden können. D.h., dass nicht alle Variablen entlang einer Kante beachtet werden müssen. Im Allgemeinen muss eine Variable beachtet werden, wenn sie Argument einer nachfolgenden lokalen Funktion ist. Andernfalls kann sie vernachlässigt werden.

Figur **FIGUR EINFÜGEN!!!** zeigt das Fragment eines Faktorgraphen. Die Update Regeln für dieses Fragment ergeben sich dann also

$$\mu_{x \rightarrow A}(x) = \mu_{B \rightarrow x}(x) \cdot \mu_{C \rightarrow x}(x) \quad (4.4)$$

$$\mu_{A \rightarrow x}(x) = \sum_{y,z} f_A(x, y, z) \cdot \mu_{y \rightarrow A}(x) \cdot \mu_{z \rightarrow A}(x) \quad (4.5)$$

$$F_x(x) = \mu_{x \rightarrow A}(x) \cdot \mu_{A \rightarrow x}(x) \quad (4.6)$$

## 4.5 Belief Propagation in Baysschen Netzwerken

FIGUR EINFÜGEN!!!

Die Verteilungsfunktion () eines Baysschen Netzwerkes erlaubt eine intuitive Umformulierung zur Representation eines Faktorgraphen. Eine zu einem einzigen Faktor gehörende lokale Funktion in (), hat die Form  $f(x|a(x))$ , wobei  $a(x)$  die Menge der Nachfolger von  $x$  im zugehörigen Baysschen Netzwerk ist. In Faktorgraphen wurden Nachfolgerknoten durch Pfeile gekennzeichnet. Diese Pfeile erlauben es uns einen Faktorgraphen ebenso als Bayssches Netzwerk ansehen zu können.

## 5 Expectation Propagation

In diesem Kapitel werden rekursive Approximationstechniken beschrieben, um die KL-Divergenz zwischen Posterior und Approximation zu minimieren. Das sogenannte Assumed-density Filtering ist eine schnelle Methode auf diesem Gebiet. Expectation Propagation oder kurz EP ist eine Erweiterung des Assumed-density Filtering um Situationen stapelweise zu verarbeiten. Es hat höhere Genauigkeit als das Assumed-density Filtering und andere vergleichbare Methoden um Schlussfolgerungen zu approximieren.

### 5.1 Assumed-density Filtering

Dieser Abschnitt fasst die Idee des Assumed-density Filtering (ADF) zusammen um die Grundlagen für die Methode des Expectation Propagation zu schaffen. Assumed-density Filtering ist eine der grundlegenden Techniken, um Posterior in Baysschen Netzwerken und anderen statistischen Modellen zu approximieren.

## 6 Das Trueskill Verfahren

Das Trueskill Verfahren ist ein Bayssches Ranking Verfahren, dass in seinen Grundzügen eine Erweiterung des Elo Ranking Systems aus dem Schach ist. Dieses neue System beachtet Unsicherheiten in der Bewertung eines Spielers, stellt neue Modelle für ausgeglichene Spiele auf und kann die Skills eines einzelnen Spielers aus dem Rang eines Teams ermitteln. Unsicherheiten werden durch eine Approximierung des message passing in faktorgraphen simuliert.

### 6.1 Der Trueskill Faktorgraph

### 6.2 Approximieren des Message Passing

## Literatur

## Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen, als die angegebenen Quellen und Hilfsmittel benutzt, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet habe.

Ort, den Datum