

Weinanalyse durch Künstliche Intelligenz am Beispiel von burgenländischen Rotweinen

WALTER FLAK¹, ERICH WALLNER¹, MARTIN TIEFENBRUNNER² und WOLFGANG TIEFENBRUNNER¹

¹Bundesamt für Weinbau
A-7000 Eisenstadt, Gölbeszeile 1

²Logistic Management Service
D-80331 München, Rosenstraße 7
E-mail: W.Flak@bawb.at

Künstliche Neuronale Netze sind lern- und abstrahierfähig und werden daher oft zur Gestalt- und Musterwahrnehmung eingesetzt. Im Rahmen dieser Arbeit wurden Neuronale Netze benutzt, um die Fähigkeit menschlicher Weinkoster nachzuahmen, höchst komplexe olfaktorische Eindrücke zu verarbeiten. Verwendet wurden 107 für den Handel bestimmte Weine aus den vier Weinbaugebieten des Burgenlandes. Die Weine der Sorten 'Blaufränkisch' und 'Zweigelt' entstammten den Jahrgängen 2003 bis 2006. Von allen wurde ein Aromaspektrum von 64 Komponenten aus unterschiedlichen chemischen Gruppen bestimmt. Es sollte überprüft werden, ob es durch Anwendung eines gut geschulten Neuronalen Netzes möglich ist, aus der Aromakomposition die Sorte, den Jahrgang und das Anbaugebiet einer unbekannt Probe festzustellen. Die Zuordnung zur Weinsorte erfolgte in 98 % aller Fälle korrekt (Wahrscheinlichkeit einer zufällig korrekten Zuordnung 50 %), die Jahrgangszuordnung in 93 % (25 %) und die Identifikation des Weinbaugebietes in 92 % (25 %). Für die hohe Erkennungsrate ist im Fall der regionalen Zuordnung allerdings Kenntnis über Sorte und Jahrgang Voraussetzung. Die Region wirkt sich daher auf die Aromakomposition verschiedener Sorten und Jahrgänge unterschiedlicher aus als die anderen beiden Kategorien, aber grundsätzlich enthält das Aromaspektrum die erforderliche Information, um Weine verschiedener Weinbaugebiete zu unterscheiden. Aus diesen Resultaten lässt sich schließen, dass eine umfassende Charakterisierung von Weinen mittels Neuronalen Netz möglich und sinnvoll ist.

Schlagwörter: Künstliches Neuronales Netz, Wein, Weinverkostung, Rebsorte, Aromaprofil

Analysis of red wines from the Burgenland by means of artificial intelligence. Artificial neural networks have the ability to learn and to abstract and are therefore often used for pattern recognition and shape perception. In this investigation neural networks were used to simulate the abilities of human wine tasters to process highly complex olfactory sensations. 107 marketable wines (varieties 'Blaufränkisch' and 'Zweigelt'; vintages 2003 to 2006) from the four winegrowing regions of the Burgenland were used. For all samples an aroma spectrum consisting of 64 components from different chemical groups was determined. It should be investigated, if a well-calibrated neural network can determine vintage year and geographic origin of an unknown sample through the aroma composition of the variety. The identification of the variety was correct in 98 % of all cases (probability of a randomly correct identification: 50 %), the identification of the vintage was correct in 93 % (25 %) and that of the winegrowing region in 92 % (25 %). For the high percentage in the case of the regional allocation, however, knowledge on variety and vintage is a precondition. Therefore the region affects the aroma composition of different varieties and vintages more than the other two categories, but basically the aroma spectrum contains the information required to differentiate wines from different regions. From these results it can be concluded that a comprehensive characterization of wines by means of neural networks is possible and reasonable.

Keywords: artificial neural network, wine, wine tasting, grape variety, aroma profile

Analyse du vin à l'aide de l'intelligence artificielle à l'exemple des vins rouges du Burgenland. Les réseaux neuronaux artificiels sont capables d'apprendre et d'abstraire et, de ce fait, sont souvent utilisés aux fins de perception de

formes et de dessins. Dans le cadre du présent travail, les réseaux neuronaux ont été utilisés pour imiter l'aptitude de dégustateurs humains à traiter des impressions olfactives extrêmement complexes. On a utilisé 107 vins destinés au commerce provenant des quatre régions viticoles du Burgenland. Les vins des cépages 'Blaufränkisch' et 'Zweigelt' sont issus des millésimes 2003 à 2006. Un spectre aromatique constitué de 64 composants de différents groupes chimiques a été fixé pour l'ensemble de ces vins. Le but était de vérifier s'il est possible de déterminer le cépage, le millésime et la région de provenance d'un échantillon inconnu à l'aide d'un réseau neuronal après apprentissage. Les échantillons ont été affectés au cépage correct dans 98 % de tous les cas (vraisemblance d'une affectation correcte accidentelle: 50 %); l'affectation au millésime a été correcte dans 93 % des cas (25 %) et l'identification de la région viticole dans 92 % des cas (25 %). Pour ce qui est de l'attribution régionale, le taux de reconnaissance élevé est toutefois soumis à la connaissance du cépage et du millésime. La région a donc une influence plus différenciée sur la composition aromatique des différents cépages et millésimes que les deux autres catégories, mais en principe, le spectre aromatique contient l'information nécessaire pour distinguer les vins provenant de différentes régions viticoles. On peut conclure de ces résultats qu'une caractérisation globale de vins à l'aide de réseaux neuronaux est possible et judicieuse.

Mots clés : Réseau neuronal artificiel, vin, dégustation des vins, cépage, profil aromatique

Eine der beeindruckendsten Fähigkeiten des menschlichen Gehirns ist die Gestaltwahrnehmung und -interpretation. Sie ist auch nur sehr schwer durch informationsverarbeitende Maschinen nachzuahmen, geschweige denn zu übertreffen. Elektronische Netzwerke aus künstlichen Neuronen oder deren Simulation in einem herkömmlichen Computer wurden ursprünglich dazu entwickelt, die Fähigkeiten des Gehirns besser zu verstehen, haben aber mittlerweile in bestimmten Bereichen, in denen klassische Methoden wenig erfolgreich sind, eine große Anzahl nützlicher Applikationen gefunden. Häufige Anwendung finden Feedforward Backpropagation Neuronale Netze (FBNN) gerade in Zusammenhang mit Gestaltwahrnehmung. So werden sie zur Fingerabdruckidentifikation, Gesicht- und Unterschrifterkennung eingesetzt (MARINAI et al., 2005). Im Weinbereich wurden Neuronale Netze zur Erkennung von Rebsorten anhand der Analyse von Blattsilhouetten (MANCUSO, 1999 und 2002) sowie zur Herkunftszuordnung von Weinproben der Sorte 'Grüner Veltliner' anhand der Aromastoffprofile eingesetzt (SCHREINER et al., 2007).

Auch in anderen Bereichen sind FBNNs von Nutzen. So werden sie im Rahmen der Statistik als nichtlineare Alternativen zur Hauptkomponenten-, Cluster- oder Diskriminanzanalyse verwendet, weiters zur Diagnose und Prognose in der Medizin (STEPHAN et al., 2005) und Ökonomie (PINO et al., 2008) sowie zur Vorhersage von menschlichem Verhalten (MARCHIORI und WARGLIEN, 2008).

Das Gehirn ist nicht nur in der Erkennung von visuellen Eindrücken hocheffizient, auch komplexe olfaktorische Muster werden erfolgreich analysiert. Der typische Geruch und Geschmack, z. B. von Wein, besteht aus

hundertenden Komponenten sehr unterschiedlicher Konzentration, trotzdem kann ein erfahrener Koster wesentliche Informationen, z. B. über Sorte oder auch Jahrgang, aus der Gesamtkomposition erschließen. Ob ein synthetisches Gegenstück zum Gehirn ähnliche Fähigkeiten aufweist, wurde von TIEFENBRUNNER et al. (2009) untersucht. Insbesondere wandten sich die Autoren der Frage zu, ob ein gut trainiertes FBNN aus dem Aromaprofil des Weins auf den verwendeten Traubenmost und die zur Vergärung benützte Hefesorte rückschließen kann und weiters, ob, wenn man Hefesorte und Trauben vorgibt, das Aromaspektrum des Weins prognostiziert werden kann. Verwendet wurden Weine, die aus verschiedenen Traubenmosten mit Hilfe von unterschiedlichen Hefen unter Laborbedingungen hergestellt wurden. Das Ergebnis des Versuchs war sehr zufriedenstellend. So wurde in fast allen Fällen der Traubenmost und in 2/3 aller Fälle sogar die Hefe richtig erkannt. Die prognostizierten Aromaspektren der Weine wichen nur geringfügig von den tatsächlichen ab. Wie bereits erwähnt, wurden alle Weine für diesen Versuch unter genau kontrollierten, vergleichbaren Laborbedingungen hergestellt. Das ist natürlich nicht sehr praxisnahe und relativiert Aussagen über die Verwendbarkeit einer „künstlichen Nase“ unter realitätsnäheren Bedingungen. Das Aromaprofil eines Weines wird nicht nur vom verwendeten Traubenmost und der fermentierenden Hefe geprägt, auch önologische Maßnahmen, Kellertechnologie und die Lagerung des Weines spielen eine wichtige Rolle. Weiters ist der von verschiedenen Rebstöcken in oft unterschiedlicher Lage gewonnene Traubenmost auch dann nicht identisch, wenn alle Stöcke der gleichen Rebsorte angehören. Um die Fähigkeiten eines Neuronalen Netzes un-

ter realitätsnahen Bedingungen zu testen, verwendeten wir für den Handel bestimmte Weine verschiedener Sorten und Jahrgänge der Weinbauregion Burgenland. Von Interesse waren nicht nur die Effizienz des Neuronalen Netzes beim Erkennen der Sorte und des Jahrgangs, sondern auch wie gut die Zuordnung zu einem der vier Weinbaugebiete des Burgenlands erfolgt.

Material und Methoden

Weine

107 Weine aus der Weinbauregion Burgenland standen für diese Untersuchung zur Verfügung, 32 aus dem Weinbaugebiet Neusiedlersee, ebenso viele aus dem WBG Neusiedlersee-Hügelland, 29 aus dem Mittelburgenland und 14 aus dem Südburgenland. 58 dieser Weine waren von der Sorte 'Blaufränkisch', 49 von der Sorte 'Zweigelt'. Die meisten stammten von den Jahrgängen 2005 und 2006 (Tab. 1). Die Verteilung der Häufigkeit der Weine in dieser Stichprobe war durch die Kultivierungshäufigkeit in den Weinbaugebieten beeinflusst.

Analyse der Aromakomponenten

Für das Aromaprofil wurden 64 Aromakomponenten von allen Weinen untersucht, insbesondere Acetate, Ethylester, freie Fettsäuren, Alkohole, Terpene, Aromen, die den Alterungston beeinflussen, und auch typische Barrique-Aromen.

5 ml der Weinprobe mit 2 g Natriumchlorid und 50 µl 3-Decanol (48,8 mg/l) als interner Standard wurden in ein 10 ml Headspace-Vial eingebracht, ausgestattet mit einem Magnetrührer und verschlossen mit einem PTFE-beschichteten Siliconseptum. Die Anreicherung der Aromakomponenten erfolgte mittels einer 2 cm Car/PDMS/DVB Faser (Fa. Supelco) im Dampfraum für 30 min bei 30 °C. Unmittelbar nach der Extraktion wurde die Faser in den GC Injektor Port eingefügt und dort für 2 min bei 250 °C zur thermischen Desorption belassen. Zur Auftrennung der verschiedenen Aromakomponenten wurde eine CP-WAX 52 CB Kapillarsäule (50 m x 0,32 mm, 0,4 µm df; Fa. Varian Inc., USA) verwendet. Die Ofentemperatur wurde für 3 min bei 50 °C belassen, bevor sie pro Minute um 4 °C bis auf 180 °C gesteigert wurde und danach mit einer Rate von 25 °C/min auf 230 °C. Diese Temperatur wurde für 7,5 min gehalten. Der komplette Zyklus währte 45 min. Als Trägergas wurde Helium mit einer konstanten Flussrate von 1,0 ml/min verwendet.

Tab. 1: Verteilung der Versuchsweine ('Blaufränkisch' und 'Zweigelt') auf die Weinbaugebiete nach Jahrgang gegliedert

	Blaufränkisch		Zweigelt		
Neusiedlersee	2003	-	2003	1	
	2004	-	2004	3	
	2005	2	2005	4	
	2006	5	2006	17	
		7		25	32
Neusiedlersee-Hügelland	2003	-	2003	-	
	2004	-	2004	-	
	2005	12	2005	5	
	2006	6	2006	9	
		18		14	32
Mittelburgenland	2003	-	2003	-	
	2004	1	2004	-	
	2005	17	2005	2	
	2006	9	2006		
		27		2	29
Südburgenland	2003	-	2003	-	
	2004	-	2004	-	
	2005	3	2005	4	
	2006	3	2006	4	
		6		8	14
Gesamt		58		49	107

Neuronales Netz

Mit Hilfe der Entwicklungsumgebung Borland Developer Studio 2006 und unter Verwendung der Sprache Delphi Pascal (Fa. Borland International, USA) wurden mehrere dreischichtige Feedforward Backpropagation Neuronale Netze entwickelt. Die Eingabeschicht bestand aus mindestens 64 Inputneuronen, für jede Aromakomponente eines. In einigen Fällen wurde dem Neuronalen Netz Zusatzinformation zur Verfügung gestellt. Wenn z. B. die Zielvorstellung das Erlernen des Weinbaugebiets war, wurde dem entsprechenden Netz z. T. auch noch die Information über die Weinsorte und/oder den Jahrgang angeboten. Entsprechend existierten ein oder zwei zusätzliche Eingänge. Je nach Fragestellung umfasste die Ausgabeschicht 4 (Weinbaugebiet und Jahrgang) bzw. 2 (Weinsorte) Neuronen, die Zwischenschicht enthielt so viele „versteckte“ Neuronen, dass ein Auswendiglernen der Weine nicht mehr erfolgte. Als Aktivierungsfunktion der Neuronen wurde eine Sigmoidfunktion, $1/(1+e^x)$, gewählt. Zur Fehlerückführung (Backpropagation) wurde die Summe der quadrierten Differenzen zwischen der gemessenen und der erwarteten SIM-Fläche gebildet (RUMELHART et al., 1986). Das Lernen wurde solange durchgeführt, bis ein minimaler Fehler erreicht war. Als Lernmenge dienten die Aromaprofile von 106 Weinen, die Testmenge umfasste den Wein, der

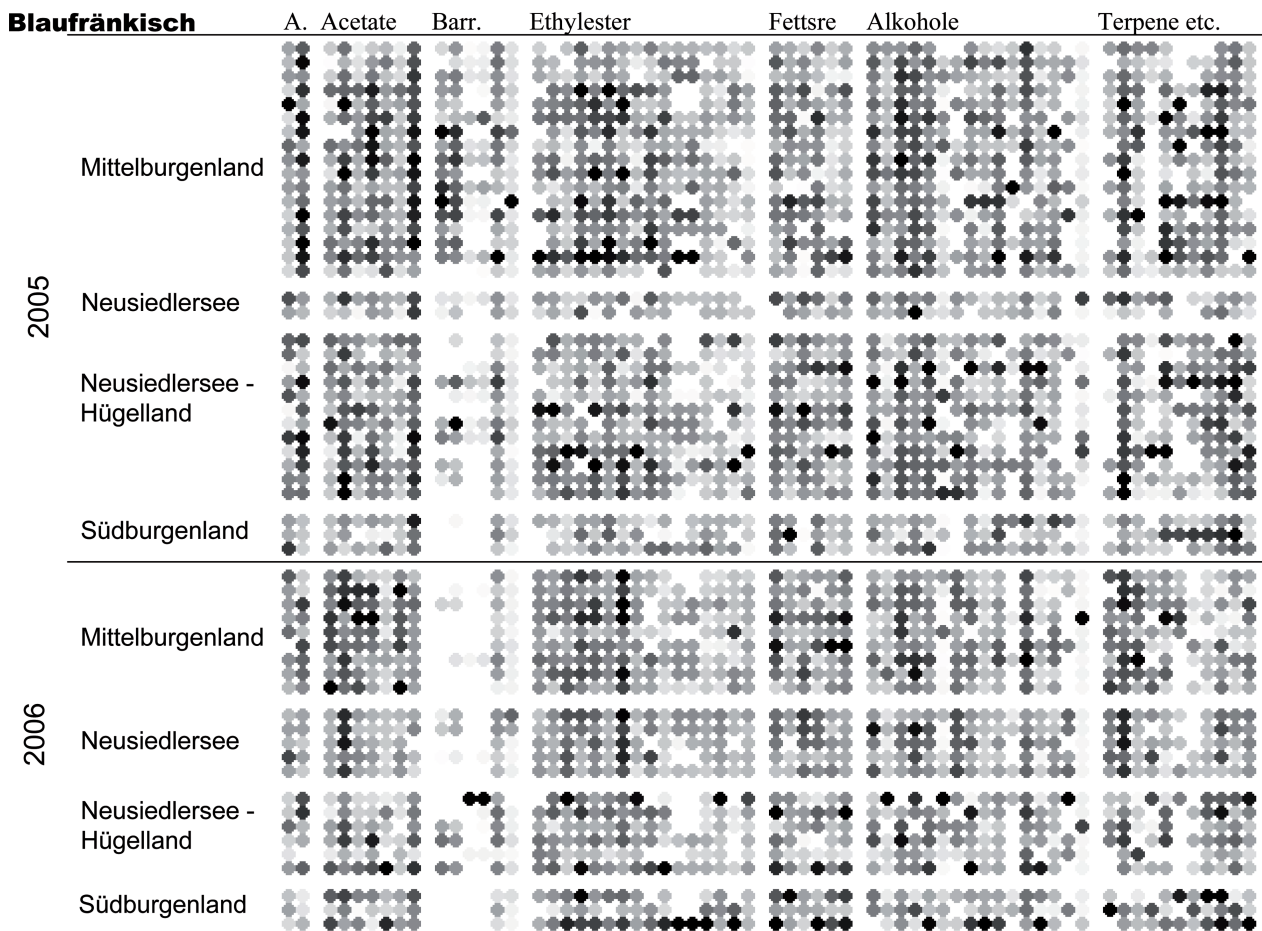


Abb. 1: Aromaspektrum der Blaifränkisch-Weine nach Jahrgang und Weinbaugebiet. Je dunkler der Kreis, desto höher die Konzentration (bzw. desto größer das SIM-Areal). Alle Daten sind innerhalb einer Spalte Minimum-Maximum-skaliert (gleicher Grauwert bei unterschiedlichen Spalten bedeutet daher nicht notwendigerweise gleiche Konzentration).

nicht der Lernmenge angehörte. Diese Prozedur wurde für alle 107 Weine durchgeführt (Jackknife-Methode), wobei natürlich jeweils mit ungeschulten Netzen begonnen wurde. Jeder Wein wurde daher einmal verwendet, um festzustellen, ob eine korrekte Zuordnung erfolgte, und letztlich lagen 107 Zuordnungsdaten vor.

Analyse der für die Unterscheidbarkeit wichtigen Aromakomponenten

Um zu erkennen, welche Aromakomponenten die Unterscheidbarkeit in bedeutender Weise beeinflussen, wurde ein Ausschlussverfahren verwirklicht, d.h., der oben beschriebene Prozess wurde mit weniger Aromakomponenten als Eingangsdaten durchgeführt und die Differenzialqualität mit jener verglichen, die bei Verwendung aller Aromasubstanzen erreicht wurde. Für das Ausschlussverfahren wurde ein genetischer Algo-

rithmus entwickelt, d.h., es wurde ein quasi-evolutionärer Selektionsprozess realisiert, bei dem bessere Aromakombinationen weniger gute verdrängen. Das war erforderlich, weil die Anzahl möglicher Aromakombinationen, die durchgetestet werden müsste, astronomisch hoch ist (etwa $3,7 \times 10^{19}$). In das Selektionskriterium ging einerseits die Korrektheit der Zuordnung ein, andererseits wurde eine minimale Anzahl von Aromakomponenten angestrebt.

Datenanalyse

Als Eingabewerte dienten die SIM (single ion mode)-Flächen für jede Aromakomponente, wobei diese Daten Minimum-Maximum-skaliert wurden, sodass die Eingabe zwischen 0 und 1 liegen konnte. Die Ergebnisse wurden zu einer Konfusions- (Erkennungs-)matrix zusammengefasst.

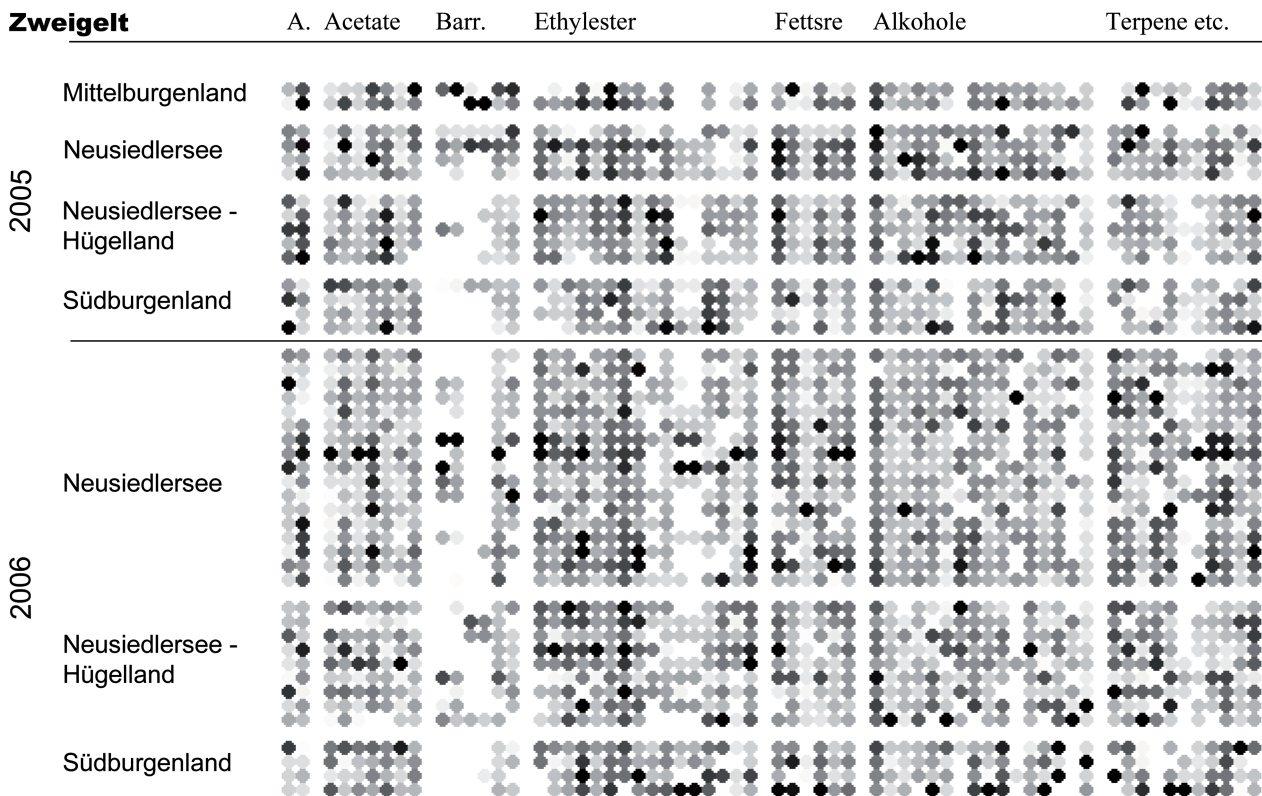


Abb. 2: Aromaspektrum der Zweigelt-Weine nach Jahrgang und Weinbaugebiet. Je dunkler der Kreis, desto höher die Konzentration bzw. desto größer das SIM-Areal. Alle Daten sind innerhalb einer Spalte Minimum-Maximum-skaliert.

Ergebnisse und Diskussion

Beschreibung der Weine und ihrer kategorischen Unterscheidbarkeit

Beim statistischen Vergleich der vorliegenden Weine der Sorten 'Blafränkisch' und 'Zweigelt' fällt auf, dass erstere im Mittel höhere Ethylester- und Terpenkonzentrationen aufweisen, auch einzelne Acetate (z. B. I-Octyl-) sind höher konzentriert. Der Unterschied kann aber zum Teil darauf beruhen, dass diese Komponenten gerade im Mittelburgenland höhere Werte aufweisen, wo erheblich mehr Blafränkisch- als Zweigelt-Weine (27 versus 2) untersucht wurden. Auch einige für Barriquetöne verantwortliche Substanzen (z. B. Holzlactone) sind im 'Blafränkisch' in größerer Dichte vorhanden, wenigstens im Mittelburgenland. Die Alkohol- und Fettsäurekonzentration war hingegen 2005 in 'Zweigelt' höher als in den Blafränkisch-Weinen.

Innerhalb der Weine der Sorte 'Blafränkisch' gibt es auffällige Jahrgangsunterschiede (Abb. 1).

Im Mittelburgenland dominieren 2005 im Vergleich zu 2006 die Holzlactone und Furfural deutlich, auch andere barriquetrelevante Komponenten. Im Südburgenland sind hingegen 2006 viele Ethylester und einige Alkohole höher konzentriert als im Vergleichsjahr. Auch viele Acetate überwiegen 2006 in diesem Weinbaugebiet. In beiden Jahrgängen nimmt die Konzentration der Acetate im Mittel (Median) in der Reihenfolge Mittelburgenland, Neusiedlersee-Hügelland, Neusiedlersee, und Südburgenland ab. Auch die Ethylesterkonzentration ist im Mittelburgenland am höchsten, Fettsäuren und Alkohole werden überwiegend homogener ausgebildet. Bei vielen Substanzen ist die Streuung der Konzentration auch innerhalb eines Jahrgangs und Weinbaugebiets so hoch, dass eine unmittelbare Zuordnung eines Weines nach dem Aromaspektrum mit freiem Auge (aus Abb. 1) wenig Erfolg versprechend erscheint. Bei 'Zweigelt' sind die Gebiets- und Jahrgangsunterschiede nicht so auffällig bzw. werden sie von der Varianz innerhalb der Gebiete und Jahrgänge überlagert. Im Jahr 2006 sind in den Weinbaugebieten Neusiedlersee-Hügelland und

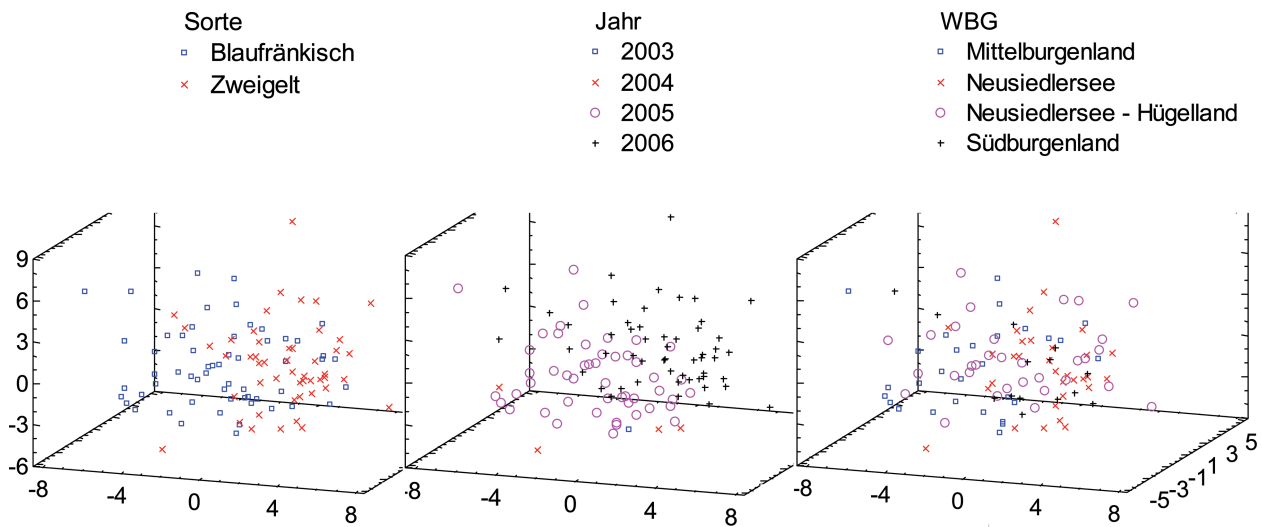


Abb. 3a-c: Hauptkomponentenanalyse der 107 Weine.

Südburgenland einige Terpene höher konzentriert. Innerhalb eines Jahrgangs werden im WBG Neusiedlersee einige Fettsäuren, Alkohole und Terpene am stärksten synthetisiert. Die Ethylesterproduktion ist im WBG Neusiedlersee-Hügelland aber annähernd ebenso groß wie im WBG Neusiedlersee. Auch bei diesen Weinen erscheint eine Zuordnung zu den Kategorien Jahrgang und Weinbaugebiet nicht einfach zu sein. Um die Unterscheidbarkeit der Weine bezüglich der Kategorien Sorte, Jahrgang und Weinbaugebiet zu untersuchen, wurde eine Hauptkomponentenanalyse durchgeführt. Dabei wurde natürlich auch untersucht, ob die Datenvariabilität auf wenige Hauptkomponenten zurückgeführt werden kann, die eine Linearkombination der Aromakomponenten darstellen. Dies ist jedoch nur in geringem Umfang der Fall. Um 95 % der Variabilität erklären zu können, sind immer noch 35 Komponenten erforderlich, und 16 Komponenten (77,8 % der Variabilität) haben Eigenwerte ≥ 1 . Benützt man die ersten drei Hauptkomponenten (34,4 % der Variabilität) zur graphischen Repräsentation (Abb. 3) ergibt sich Folgendes:

Nach der Sorte lassen sich die Weine zwar nicht klar trennen, aber es ergibt sich ein deutlicher schwerpunktmäßiger Unterschied. Die Auftrennung nach Jahrgang ist zumindest für 2005 und 2006 sogar noch deutlicher. Bei den Weinbaugebieten lassen sich die Weine vom Neusiedlersee und aus dem Mittelburgenland noch am besten mit diesem Verfahren trennen, aber generell ist die Überschneidung der den verschiedenen Weinbaug-

bieten zugeordneten Weine und ihrer Aromaspektren sehr hoch.

Sortenerkennung

Die Identifikation der Rebsorte aus dem Weingeschmack ist eine Fähigkeit, die man von einem gut trainierten Weinkoster bis zu einem gewissen Grad erwartet. Um einem Neuronalen Netz diese Fähigkeit beizubringen, wurde ein Lernset mit der für diesen Zweck maximal zur Verfügung stehenden Anzahl von Weinen - 106 - verwendet, der letzte Wein musste aus dem Aromaspektrum heraus nach den aus dem Lernset gewonnenen Abstraktionsregeln der Sorte zugeordnet werden. Jeweils mit einem ungeschulten Neuronalen Netz beginnend, wurde dieses Verfahren für alle 107 Weine wiederholt. Jeder Wein wurde also einmal verwendet, um zu erkennen, ob eine korrekte Zuordnung erfolgte. Die Zuordnung zur Weinsorte aus dem Aromaspektrum erfolgte in 105 von 107 Fällen korrekt, d.h. in 98 % aller Fälle, bei einer erwarteten zufälligen Erfolgsrate von 50 %. Gibt man dem Netz noch zusätzlich die Information, um welchen Jahrgang es sich bei dem Wein handelt und aus welchem Weinbaugebiet er stammt, lässt sich der Prozentsatz korrekter Zuordnungen noch geringfügig auf über 99 % (106 Weine) steigern (ein 'Zweigelt' wurde falsch als 'Blaufränkisch' identifiziert).

Unter einigermaßen realistischen Bedingungen ist das Neuronale Netz daher offenbar in der Lage, die allgemeinen Abstraktionsregeln zu erlernen, die für die Sor-

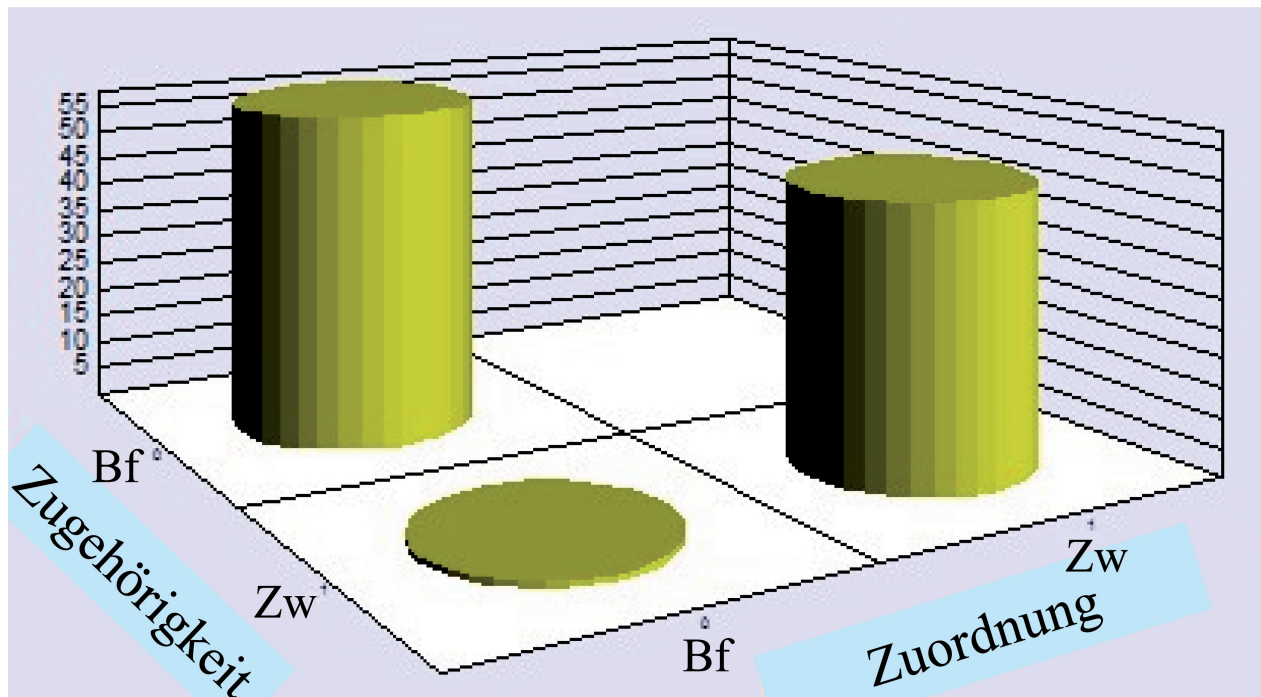


Abb. 4: Erkennungsmatrix für die Weinsorte. Jene Weine, in denen tatsächliche Zugehörigkeit und Zuordnung durch das Netz übereinstimmen, sind korrekt identifiziert.

tenzuordnung aus der Aromakomposition erforderlich sind.

Zuordnung zum Jahrgang

Die Zuordnung zum Jahrgang erscheint wesentlich schwieriger als die zur Sorte, schon deshalb, weil man vermuten könnte, dass sich Jahresunterschiede in verschiedenen Weinbaugebieten unterschiedlich auswirken können und natürlich auch in verschiedenen Weinsorten. Der Lernprozess für diese Fragestellung erfolgte analog jenem, der für die Identifikation der Sorte beschrieben wurde.

Wir erwarten in diesem Fall eine zufällige Übereinstimmung von Zugehörigkeit zu einem bestimmten Jahrgang und der Zuordnung durch das Netz von 25 %. Tatsächlich sind 92 % aller Weine korrekt zugeordnet worden, auch dann, wenn keine Zusatzinformation über Sorte und Weinbaugebiet zur Verfügung gestellt wurde. Überraschend ist, dass sich durch Zusatzinformation die Zuordnung kaum verbessern lässt - sie steigt auf 93 % (Abb. 5). Demnach wäre anzunehmen, dass sich Jahrgangsunterschiede in den verschiedenen Weinbaugebieten und den unterschiedlichen Sorten auf ähnliche Weise manifestierten. Zu erwarten war hingegen die falsche Zuordnung des einsamen Weines aus dem

Jahr 2003. Hier fehlten die Lernbeispiele. Die wenigen Lernbeispiele, die für das Jahr 2004 vorlagen, genügten hingegen für eine überwiegend korrekte Zuordnung (3 von 4). Vier Weine aus 2005 wurden irrtümlich als dem Jahrgang 2006 zugehörig identifiziert und nur einer der 2006er Weine als Jahrgang 2005.

Zuordnung zum Weinbaugebiet

Wie Abbildung 3 zu entnehmen ist, haben die Weinbaugebiete von den drei hier untersuchten Kategorien die am wenigsten ausgeprägten Eigenheiten, weniger als Sorte und Jahrgang. Man sollte auch annehmen, dass hier zusätzliche Kenntnis über den Wein die Zuordnungsqualität deutlich erhöht.

Wir erwarten eine zufällig korrekte Zuordnung von 25 %. Die korrekte Identifizierung des Weinbaugebietes durch das Neuronale Netz aus dem Aromaprofil des Weines beträgt 85 %. Teilt man dem Netz zusätzlich mit, welche Weinsorte und welcher Jahrgang vorliegen, erhöht sich die Qualität der Zuordnung sehr deutlich. Das Neuronale Netz erkennt nun die Weinbauregion in 92 % aller Fälle, eine Leistung, die der der Jahrgangszuordnung kaum nachsteht.

Am charakteristischsten sind offenbar die Weine des Mittelburgenlandes; kein einziger wurde einem falschen

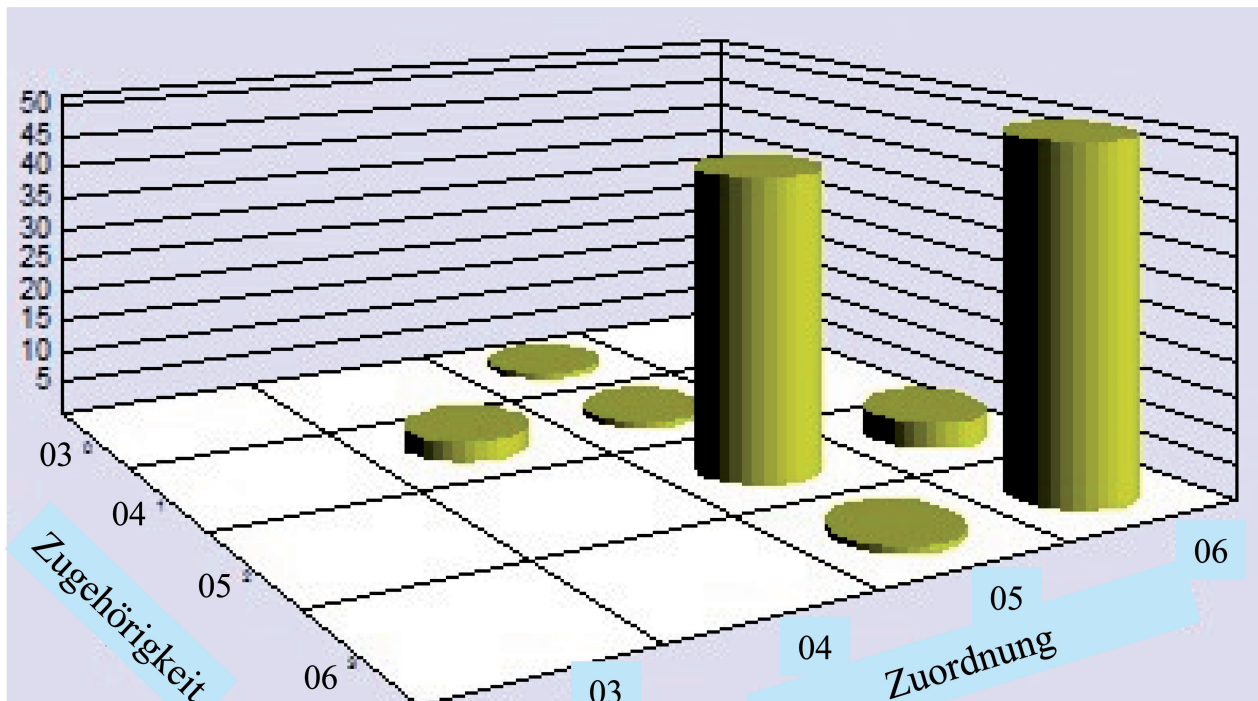


Abb. 5: Erkennungsmatrix für den Jahrgang. Jene Weine, in denen tatsächliche Zugehörigkeit und Zuordnung durch das Netz übereinstimmen, sind korrekt identifiziert (Diagonale von links nach rechts).

Weinbaugebiet zugeordnet und lediglich ein Wein wurde irrtümlich als „mittelburgenländisch“ identifiziert. Am häufigsten wurden die Weine aus dem Neusiedlersee-Hügelland inkorrekt beigeordnet (5 Weine), wobei die Verwechslung primär mit dem Weinbaugebiet Neusiedlersee erfolgte (4 Weine). Analog resultierte die inkorrekte Zuordnung von Weinen aus dem Gebiet Neusiedlersee ausschließlich mit dem benachbarten Neusiedlersee-Hügelland (2 Weine). Südburgenländische Weine wurden in je einem Fall dem Neusiedlerseegebiet und dem Neusiedlersee-Hügelland zugeordnet. Die gute Unterscheidbarkeit von Weinen der verschiedenen Weinbaugebiete ist auf den ersten Blick überraschend, beträgt doch die Nordsüd-Erstreckung des Burgenlandes nur etwa 140 km. Unterschiede in der Jahrestemperatur zwischen den Weinbaugebieten bedingt durch Änderung der geographischen Position in Nordsüd-Richtung sind daher nicht allzu hoch. Andererseits ist das Klima im Burgenland durch seine nach Osten offene Lage und daher durch kontinentale Einflüsse bestimmt. Im nördlichen Landesteil herrscht pannonisches Klima vor und es ergeben sich die höchsten Durchschnittstemperaturen und die längste Sonnenscheindauer Österreichs. Im Hügelland des mittleren und südlichen Burgenlandes sind hingegen die Som-

mer weniger heiß und die Winter milder. Auch die Bodenunterschiede zwischen den Weinbaugebieten des Burgenlandes können beträchtlich sein. So dominieren im Weinbaugebiet Neusiedlersee 10 bis 15 m mächtige eiszeitliche Donauschotter, was einen bedeutenden Einfluss auf den Wasserhaushalt des Bodens hat. Für das Mittelburgenland hingegen sind tiefgründige, schwere Böden mit gutem Wasserspeichervermögen charakteristisch. Insgesamt ergeben sich dadurch selbst für die sehr kleinräumigen Weinbaugebiete des Burgenlandes signifikante Unterschiede des Klimas und der Bodenverhältnisse, die sich offenbar auf die Weincharakteristik hinreichend auswirken, um eine hohe Unterscheidbarkeit zu gewährleisten. Am wenigsten ist das natürlich für die benachbarten Weinbaugebiete Neusiedlersee und Neusiedlersee-Hügelland der Fall, die zudem die gleiche Nordsüderstreckung aufweisen.

Für die Fragestellung nach der Zuordnung der Aromaspektren zu den Weinbaugebieten wurde untersucht, welche Aromakomponenten für eine korrekte Identifikation besonders bedeutend sind. Der genetische Algorithmus erkannte schließlich 20 Aromasubstanzen als signifikant, wobei alle sieben untersuchten Aromagruppen bei den für die Zuordnung wichtigen Aromasubstanzen inkludiert sind (Tab. 2). Es bedarf also eines

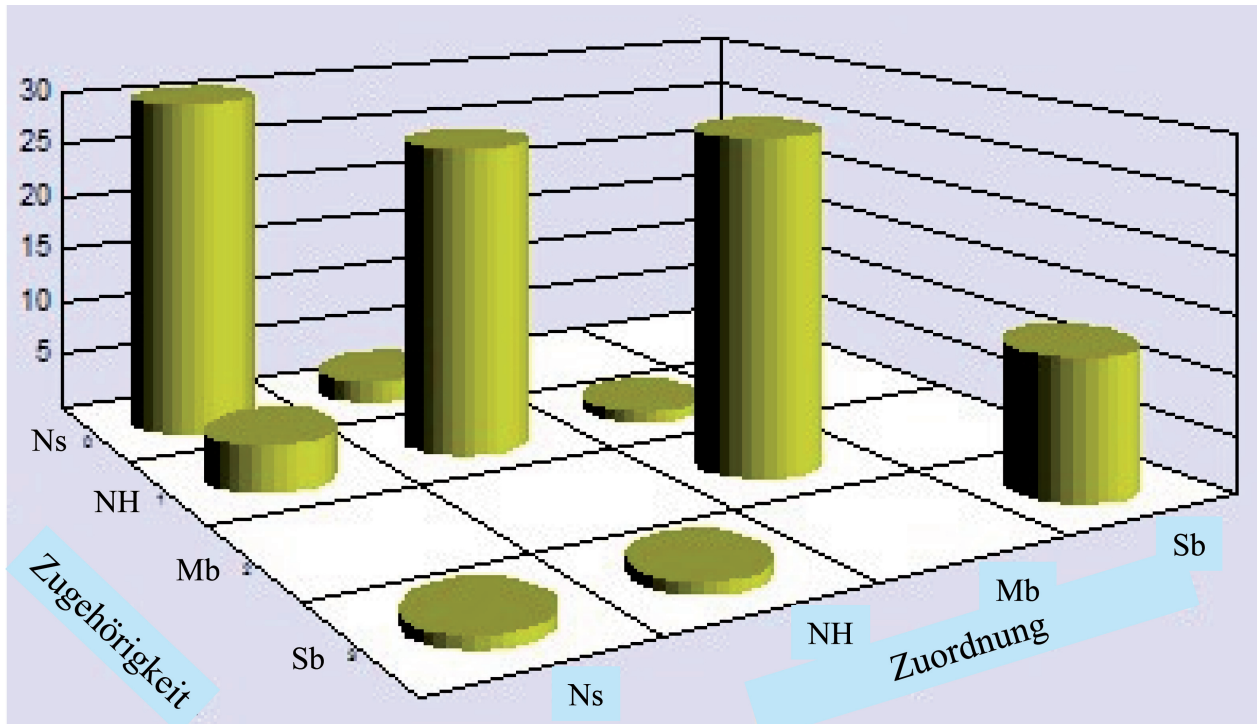


Abb. 6: Erkennungsmatrix für die Weinbauregion. Jene Weine, in denen tatsächliche Zugehörigkeit und Zuordnung durch das Netz übereinstimmen, sind korrekt identifiziert (Diagonale von links nach rechts).

breiten Spektrums unterschiedlicher Substanzen, um eine Zuordnung mit hoher Sicherheit zu bewerkstelligen. Die geringste Bedeutung hat dabei offenbar die Gruppe der Alkohole, bei der nur zwei von insgesamt 16 Vertretern als wichtig eingestuft wurden.

Die vorliegende Analyse zeigt, dass eine Zuordnung unbekannter Weine mittels ihrer Aromaspektren zu den Kategorien Sorte, Jahrgang und sogar Weinbauregion mit hoher Qualität möglich ist. Es ist daher auch sinnvoll, Weine nach diesen Kategorien zu beschreiben und herauszuarbeiten, welche die jeweils prägenden Merkmale sind. Wünschenswert wäre natürlich, die genauen Abstraktionsregeln zu kennen, die eine Zuordnung ermöglichen. Obwohl ein Neuronales Netz abstrahieren kann, bleiben dem Anwender des Netzes die Abstraktionsregeln, nach denen das Netz vorgeht, unbekannt. Gerade aber solange diese Regeln nicht bekannt sind und daher andere Verfahren nicht verwendet werden können, ist das Neuronale Netz die einzige Möglichkeit, eine objektive Zuordnung zu erlernen, da ein Koster oder eine Kostkommission nie vollständig objektiv sein kann bzw. sein können.

Das Neuronale Netz benötigt wenig Speicherplatz und ist - einmal etabliert - sehr schnell. Bei einer klassischen

Tab. 2: Wichtige Aromen für die Zuordnung zum Weinbauregion

"Alter des Weines"	Spiroverbindung	2 von 2
	Diethylsuccinat	
Acetate	Ethylacetat	1 von 7
"Typische Barrique-Aromen"	Holzlacton 1	1 von 6
Ethylester	Ethylbutyrat	7 von 16
	Ethyllactat	
	Ethyloctanoat	
	Ethylfuroat	
	3-Hexensäureethylester	
	Ethylheptanoat	
Säuren	Butansäure	2 von 6
	3-Methylbutansäure	
Alkohole	n-Butanol	2 von 16
	2-Phenylethanol	
Terpene und sonstige	2-Nonanon	5 von 11
	Benzaldehyd	
	β-Damascenon	
	D-Limonen	
	Linalool	

Analyse mittels Computer müssen die Abstraktionsregeln bereits bekannt sein. Klassische Analyse und Untersuchungen mittels Neuronalem Netz sind so fundamental verschieden, dass hier kurz auf die Unterschiede eingegangen werden soll. Bei ersterer sind Speicher- und Recheneinheit strikt getrennt. Man müsste also, wenn wir das vorliegende Beispiel nehmen, zunächst die Information bezüglich der Weine speichern, pro Aromakomponente und Wein ist eine Gleitkommazahl mit dem Speicherumfang von üblichen 8 Byte (64 Bit) erforderlich, insgesamt bei 64 Aromakomponenten und 107 Weinen also 54.784 Byte. Die Recheneinheit würde diese Datenmenge mit einem vorgegebenen mathematischen Algorithmus verarbeiten, um entsprechend der Fragestellung zu einem erwünschten Resultat zu gelangen. Natürlich sollte das Verfahren für die Praxis etabliert werden und der Aufwand stiege dann entsprechend der Anzahl der Weine, die untersucht wurden und beurteilt werden sollen.

Das Neuronale Netz speichert Information in den Synapsen, aber es ist weniger Speicherplatz vorhanden, als zur Speicherung aller Daten erforderlich wäre. Wenn das Neuronale Netz die Zuordnung zum Weinbaugebiet erlernt, sind es beispielsweise in der einfachsten Variante nur 1.656 Byte. Zudem sind Speicher und Recheneinheit nicht getrennt. Beides zusammen ermöglicht bzw. zwingt das Netz, durch Lernen zu abstrahieren, im angeführten Beispiel also eine allgemeine „Vorstellung“ davon zu entwickeln, wie das Aroma eines bestimmten Jahrgangs „aussieht“. Diese „Vorstellung“ ist über die Gesamtheit der Synapsen verteilt. Je größer die Lernmenge ist, desto besser wird diese Vorstellung, ohne dass dabei der Speicherbedarf wächst. Deshalb eignet sich ein Neuronales Netz sehr gut dazu, z. B. die regionale Typizität eines Weines rasch und sicher festzustellen, ohne dass der Aufwand entsprechend steigt. Es kann dies, wie bereits erwähnt, auch objektiver, als z. B. eine Kostkommission. Allerdings besteht derzeit noch keine Gefahr, dass der Mensch in diesem Bereich durch Künstliche Intelligenz ersetzt wird, da die Produktion des Aromaspektrums noch zu teuer und zeitaufwändig ist. Eine bereits eingangs besprochene Fragestellung, die uns sehr interessiert hat, ist,

ob das Neuronale Netz überhaupt lernen kann, wenn die Weine nicht unter streng kontrollierten und möglichst homogenen Bedingungen hergestellt werden, sondern in verschiedenen Weinkellern unter unterschiedlichen keller-technischen und Lagerbedingungen. Tiefenbrunner et al. (2009) hatten die Eignung eines „Weinkosters mit Künstlicher Intelligenz“ an unter Laborbedingungen mittels einheitlicher Herstellungsbedingungen produzierten Weinen getestet und waren zu einem äußerst positiven Resultat gekommen. Mit der vorliegenden Arbeit ist nun gezeigt, dass auch praxisnahe Herstellungsbedingungen kein Hindernis für die sinnvolle Anwendung Künstlicher Intelligenz zur kategorialen Weinanalyse sind.

Literatur

- MANCUSO, S. 1999: Elliptic Fourier Analysis (EFA) and Artificial Neural Networks (ANNs) for the identification of grapevine (*Vitis vinifera* L.) genotypes. *Vitis* 38(2): 73-77
- MANCUSO, S. 2002: Discrimination of grapevine (*Vitis vinifera* L.) leaf shape by fractal spectrum. *Vitis* 41(3): 137-142
- MARCHIORI, D. and WARGLIEN, M. 2008: Predicting human interactive learning by regret driven neural networks. *Science* 319: 1111-1113
- MARINAI, S., GORI, M. and SODA, G. 2005: Artificial neural networks for document analysis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(1): 23-35
- PINO, R., PARRENO, J., GOMEZ, A. and PRIORE, P. 2008: Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. *Eng. Appl. Artif. Intell.* 21(1), 53-62
- RUMELHART, D.E., HINTON, G.E. and WILLIAMS, R. J. 1986: Learning representations by back-propagating errors. *Nature* (323): 533-536
- SCHREINER, A., BRANDES, W., VERSINI, G., BERGHOFER, E. and EDER, R. 2007: Herkunftsunterscheidung von Weinen der Sorte 'Grüner Veltliner' anhand ihrer Aromaprofile mit Hilfe von Data Mining-Techniken und Neuronalen Netzwerken. *Mitt. Klosterneuburg* 57(1): 3-22
- STEPHAN, C., CAMMANN, H. and JUNG, K. 2005: Artificial neural networks: has the time come for their use in prostate cancer patients? *Nat. Clin. Pract. Urol.* 2(6): 262-263
- TIEFENBRUNNER, M., GANGL, H., TSCHEIK, G. and TIEFENBRUNNER, W. 2009: Creating an artificial wine taster: Inferring the influence of must and yeast from the aroma profile of wines using Artificial Intelligence, *Vitis* 48(2): 97-100

Manuskript eingelangt am 28. April 2010