



Foto: Continental Automotive

AUTOMATISIERTE KLASSIFIKATION VON BILDDATEN

Man kann nur bewerten, was man erkennt

Udo Kreißig, Limbach-Oberfrohna

Die Güte von Lötstellen bei Halbleiterelementen wird gewöhnlich durch eine manuelle, visuelle Beurteilung von Schweiß- und Lötverbindungen eingeschätzt. Dies ist eine personal- und zeitintensive Prüfmethode, die Fachwissen und Erfahrung erfordert. Continental Automotive entwickelte einen automatisierten Prüfmodus, der sich auf ein mathematisches Verfahren der Klassifikation von Bilddaten stützt.

Continental Automotive produziert Diesel- und Benzineinspritzinjektoren, in welchen aktive Halbleiterelemente, sogenannte Piezoaktoren, als Antriebe zur Steuerung der Einspritzung zum Einsatz kommen. Gegenstand der Prüfung sind Drähte, welche mit Halblei-

terelementen mittels Lötung verbunden werden müssen. Pro Jahr sind etwa 700 Millionen Lötungen zu prüfen. In der Halbleiterfertigung ist es erforderlich, bei einigen Fertigungsschritten eine 100-Prozent-Prüfung durchzuführen. Bekannte Verfahren wie Strom-, Spannungs- und

Widerstandsmessungen sind an der Tagesordnung. Einige qualitätsrelevante Merkmale lassen sich aber nicht durch die Messung elektrischer oder geometrischer Kenngrößen hinreichend beurteilen. Speziell im Bereich der Kontaktbildung zwischen zwei Bauelementen ist das mensch-

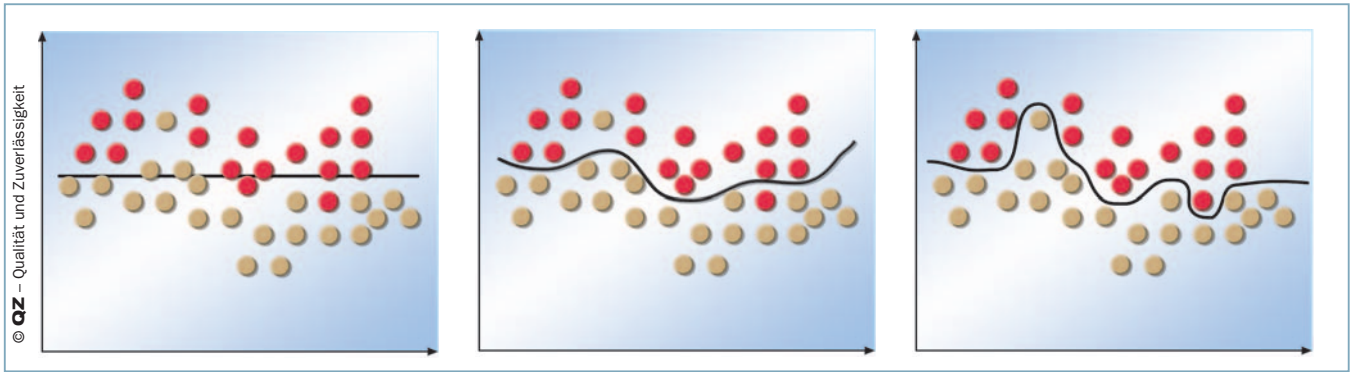


Bild 1. Am Beispiel eines zweidimensionalen Problems sieht man, wie mit verschiedenen Parametervariationen unterschiedliche Trennschärfen erreicht werden können.

liche Auge noch immer unverzichtbar. Durch eine rein mechanische Bindung zwischen zwei Bauelementen werden alle elektrisch messbaren Qualitätskriterien hinreichend positiv bewertet. Gerade im Bereich der Automobilindustrie ist eine Dauerhaltbarkeit der jeweiligen Verbindungen zwingend erforderlich, um die funktionale Sicherheit der mit Halbleitern bestückten Bauteile zu garantieren. Deshalb muss die Qualitätsprüfung über die Funktionskenngrößen hinaus erweitert werden. Durch die manuelle, visuelle Beurteilung von Schweiß- und Lötverbindungen (Lunkerbildung, Breite der Schweißnaht, Flussbild etc.) war es möglich, deren Güte einzuschätzen. Bei Großserienanwendungen wird diese manuelle Sichtkontrolle jedoch sehr personalintensiv und ist mit einer Fehlerhäufigkeit von durchschnittlich zwei Prozent verbunden. Die Technologieentwicklung von Continental Automotive entwickelte deshalb ein automatisiertes Prüfverfahren, das auf Erfahrungen aus der manuellen Sichtkontrolle basiert.

Variationen der Parameter C und Sigma

Um das Know-how der manuellen Sichtkontrolleure nutzen zu können, war es notwendig, ein mathematisches Verfahren der Klassifikation einzusetzen. Die Software Statistica Data Miner des Softwareanbieters Statsoft bietet für diese Aufgabe eine Palette an Methoden. Ausgewählt wurde das Verfahren Support Vector Machines (SVM), welches Parameter in einem n-dimensionalen Raum mittels einer Trennfunktion klassifiziert (Infobox und Bild 1). Um eine industriell einsetzbare Lösung zu generieren, muss eine zusätzliche Funktionalität verwendet wer-

den. Interessant ist nicht nur die Bewertung „gut“ bzw. „schlecht“, sondern auch, wie gut oder schlecht das jeweilige Bauteil ist und wie viel Vertrauen man in die jeweilige Bewertung des Systems haben kann. Es gibt Szenarien, in denen bereits

ein einziges als gut klassifiziertes Schlechteil den Worst Case darstellen kann. In diesem Fall wäre eine Verschiebung der Trennfunktion dahingehend sinnvoll, sich zum Preis von schlecht klassifizierten Gutteilen Sicherheit zu erkaufen. ▷

► **MUSTERERKENNUNG**

Support Vector Machines

Es handelt sich um ein rein mathematisches Verfahren aus dem Gebiet des maschinellen Lernens zur Erkennung von Mustern, mit deren Hilfe eine Klassifikation von Daten durchgeführt werden kann. Ziel dieses Klassifikationsverfahrens ist es, Mengen von Objekten so zu unterteilen, dass ein möglichst breiter Korridor zwischen den unterschiedlichen Klassen entsteht. Basis der Klassifikation sind

immer Trainingsdaten, welche hinsichtlich ihrer Ausprägung bekannt sind (im hier beschriebenen Fall wurden Bilddaten von Experten in gute und schlechte Lösungen unterteilt). An diesen sogenannten Trainingsdatensätzen wird der mathematische Algorithmus angeleitet, um ihn später auf unbekannte Daten anzuwenden und deren Ausprägung zu prognostizieren.

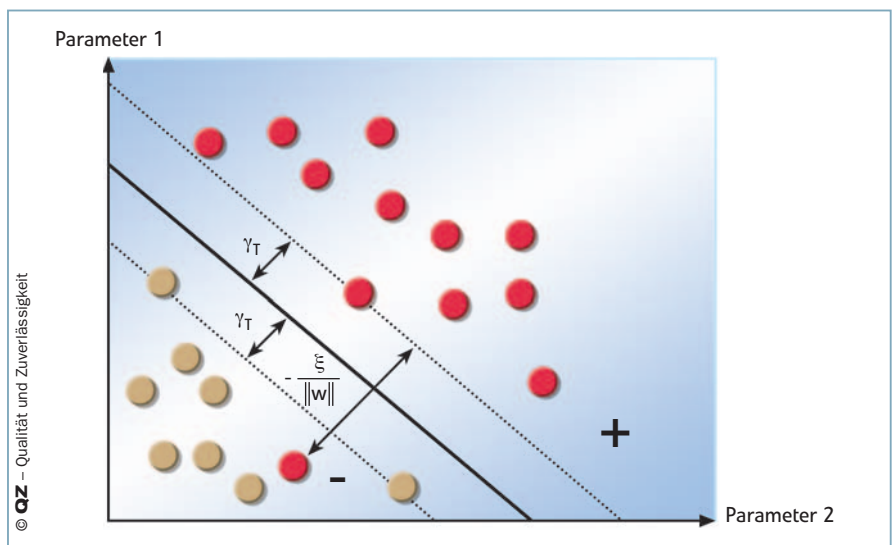


Bild 2. Linear getrennter, zweidimensionaler Datensatz, der eine Fehlklassifikation sichtbar macht

Bild 3. Um die Anzahl der Bilder zu minimieren, müssen diese referenziell ausgerichtet und um unbedeutende Dimensionen reduziert werden.

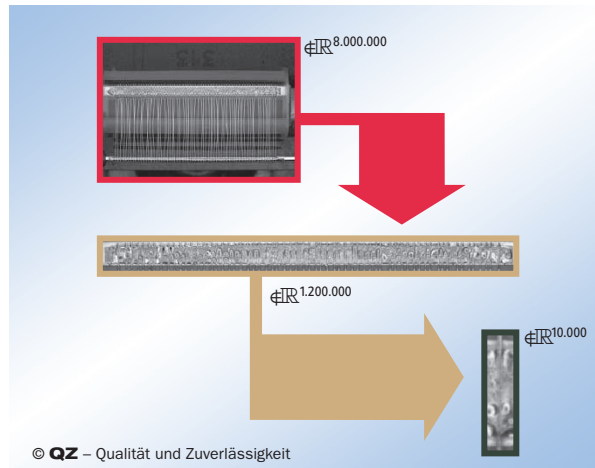
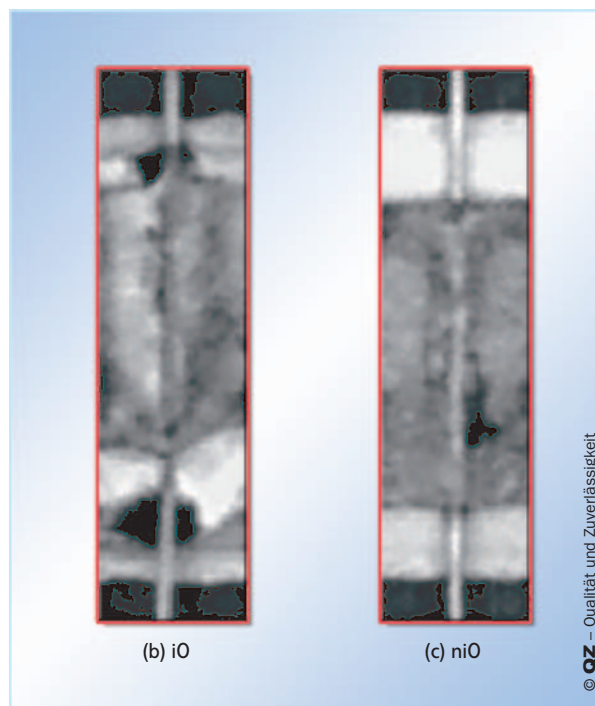


Bild 4. Der zu bewertende Bereich ist 10 000 Pixel groß und zeigt zwei gelötete Drähte (b = Gutteil, c = Schlechtteil). Für den Prüfer stellt sich die Frage: Ist das Lot wirklich geflossen oder sind die beiden Bauteile ausschließlich mechanisch verbunden?



Zur Steuerung dieser Grenzen wird ein Parameter C und ein dazugehöriges Vertrauensintervall eingeführt. Der Parameter C ist ein Maß dafür, wie stark eine Verletzung des jeweiligen Vertrauensintervalls bestraft wird. Wenn dieser Parameter so gewählt wird, dass jede Verletzung maximal bestraft wird, kommt es im Rahmen der Modellierung zur Überanpassung, dem sogenannten Overfitting. Die Trennfunktion ist in diesem Fall maximal unsteig und bildet quasi um jeden einzelnen Wert eine Insel. Die für die Modellbildung verwendeten Trainingsdaten wurden zwar optimal durch das Modell separiert, doch nachfolgende Testdatensätze konnten mit überangepassten Trennfunktionen nahezu gar nicht klassifiziert werden. An dieser Stelle wird deutlich, dass man einen

Kompromiss zwischen hinreichend gut getrennten Trainingsdatensätzen einerseits und dem Vermeiden von Overfitting andererseits finden muss. Dies geschieht über das gezielte und sinnvolle Variieren der Parameter C und Sigma.

Als Ergebnis der SVM-Analyse erhält man eine Distanz zur Trennfunktion mit dem jeweiligen Vorzeichen, welches für gut bzw. schlecht steht (Bild 2). Der Betrag der Distanz gibt an, wie gut oder schlecht, also wie sicher man sich bei dem jeweiligen Klassifikationsergebnis des speziellen Bauteils sein kann.

Vom Modell zur Bilddatenanalyse

Basis für die Berechnungen sind einzelne Bilder von Lötungen, die in eine definier-

te Zahlenmatrix umgewandelt werden müssen. Um eine numerische Verarbeitung in akzeptablen Rechenzeiten zu ermöglichen und den nötigen Lernaufwand zu minimieren, sollten die Bilder sinnvoll reduziert werden (Bild 3). Bildbereiche, die für die Beurteilung unwichtig sind, müssen entfernt werden. Hat man geeignete Bildausschnitte erstellt, sind die Einzeldrahtlötungen für die Modellbildung hinsichtlich ihrer Qualität zu bewerten. Bei der Umwandlung der einzelnen Bildausschnitte in Zahlen wurden die Graustufen der einzelnen Pixel in den jeweiligen Zahlenwert übersetzt und alle Graustufenwerte der Pixel eines Bildausschnitts einzeilig aufgereiht, und zwar in der Reihenfolge von links oben nach rechts unten. So entsteht eine Zahlenmatrix, die pro Bildausschnitt eine Zeile aufweist. Die resultierende Spaltenanzahl ist definiert über die Anzahl der auszuwertenden Bildpunkte (Bild 4). Um diesen hochdimensionierten Datenbestand einer Analyse zugänglich zu machen, wurde angenommen, dass explizit die Verteilung der Graustufen auf dem auszuwertenden Bild Aufschluss über die Ausprägung des qualitätsrelevanten Merkmals des geflossenen Lots geben muss.

Auch die Umwandlung der Bilddaten in mathematisch verarbeitbare Daten wurde mittels Statistica realisiert. Im Vorfeld wurden von Sichtkontrolleuren hinreichend viele Bilder hinsichtlich des qualitätsrelevanten Merkmals bewertet und klassifiziert. Den eingelesenen Werten wurde in einer separaten Spalte dieses Urteil (gut = 1, schlecht = 0) mitgegeben. Beim erstellten Datensatz für das Modelltraining wurde darauf geachtet, dass das Verhältnis von guten zu schlechten Teilen ausgewogen ist. Damit schließt man Effekte aus, welche maßgeblich auf die aktuelle Qualitätssituation zurückzuführen sind: Bei 95 Prozent Gutteilen ist die Wahrscheinlichkeit richtig zu urteilen ebenfalls 95 Prozent, wenn man immer „gut“ schätzt – für statistische Verfahren ist dies hinreichend. Mittels der in der Software enthaltenen SVM-Funktionalität wurde der bereitgestellte Datensatz angelehrt. Dabei wurden Funktionalitäten wie die Kreuzvalidierung weitestgehend ausgeschöpft. Nach dem Anlernen wurde ein neuer Datensatz mit Testdaten eingelesen und beurteilt. Da das Urteilsvermögen des Verfahrens sehr stark von den einzustellenden Parametern Sigma und C abhängig ist und die-

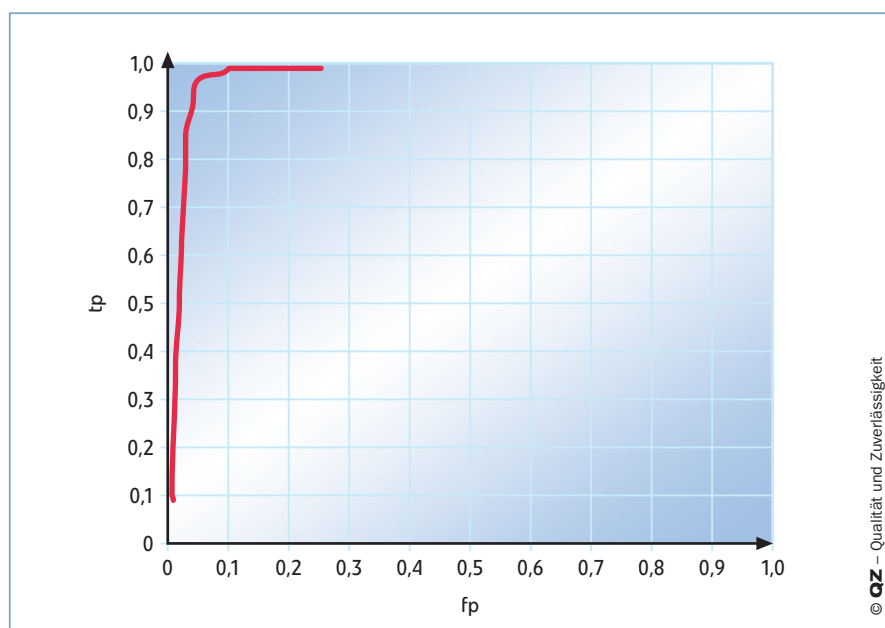


Bild 5. Im vorliegenden Objektbereich ist der Testfehler 4,01 Prozent (ROC-Kurve der SVM-Trennfunktion bei $\Sigma = 2$ und $C = 1$).

se wiederum an den jeweiligen Einsatzbereich anzupassen sind, ist eine Parametervariation sehr sinnvoll. Diese nimmt allerdings sehr viel Zeit und Rechenleistung in Anspruch. Die Erfahrung zeigt, dass der hohe Aufwand bei der Modellbildung und der Bestimmung der idealen Parameter für die zu ermittelnde Trennfunktion sich in der Trennschärfe widerspiegelt.

Erster Prüfdurchlauf automatisch, zweiter manuell

Zur Beurteilung der Güte des seitens SVM vorgeschlagenen Beurteilungsalgorithmus werden Receiver-Operating-Characteristic (ROC) -Kurven herangezogen

(Bild 5). Es handelt sich um eine Methode der Grenzwertoptimierung, mit der auch Analysestrategien bewertet und verbessert werden können. Der Verlauf einer ROC-Kurve veranschaulicht in diesem Fall die Güte des Prüfverfahrens. Je weiter entfernt die Kurve von der von links unten nach rechts oben verlaufenden Diagonalen liegt, desto besser ist die Trennschärfe des zugrunde liegenden Algorithmus.

Im vorliegenden Fall werden infolge der geforderten Verfahrenssicherheit in Bezug auf Fehlklassifikation 10 Prozent der guten Teile als schlecht klassifiziert, um einen hinreichend großen Sicherheitskorridor in Bezug auf eventuellen Schlupf von schlechten Bauteilen sicher-

zustellen. Alle schlechten Bauteile werden einer zweiten, manuellen Sichtkontrolle unterzogen. Da Support Vector Machines die Möglichkeit bietet, ständig neue Fehler anzulernen, wird die Trennfunktion wöchentlich mit neuen Fehlerbildern weiter trainiert. Das Modell ist direkt in die Fertigungslinie integriert und bewertet dort automatisch die Lötstellen. Dies führt zu einer erheblichen Einsparung an Ressourcen zur manuellen Überprüfung. Langzeituntersuchungen zur Überprüfung der Trennschärfe des eingesetzten Systems haben ergeben, dass der Fehlklassifikationsanteil aktuell bei nur 79 ppm liegt. Ein Wert, der für ein statistisches Modell exzellent ist. □

Autor

Dr. Udo Kreißig, geb. 1975, ist Leiter Technologieentwicklung für den Bereich Piezoaktor Klassifikation von Bilddaten zur Qualitätsüberwachung bei der Continental Automotive GmbH.

Kontakt zum Autor

Udo Kreißig
T 0175 7215316
udo.kreissig@continental-corporation.com

Kontakt zum Anbieter

Statsoft (Europe) GmbH
T 040 468866-0
info@statsoft.de

www.qm-infocenter.de

Diesen Beitrag finden Sie online unter der Dokumentennummer: QZ110070

QM-Infocenter.de
Das Portal für Qualitätsmanagement

QM-Foren:
Die Community im QM-Bereich
Allgemeines Forum und moderierte Spezialforen zu GPM, QM & Recht, DIN EN 10204 & QM im Gesundheitswesen

www.qm-infocenter.de/foren