



Telematik im Gesundheitswesen

Hinweis:

Auf den nachfolgenden Seiten sehen Sie den gewünschten Beitrag / Poster / Kurzpräsentation.

Wenn Sie dieses Dokument ausdrucken möchten, drücken Sie die Taste F8 und klicken die entsprechende Schaltfläche an

Um zur Ausgangsseite des Menüs „10 Jahre TELEMED“ zurückzukehren, klicken Sie auf diese Schaltfläche

[zurück zum Menü](#)

Automatische Korrektur fehlerhafter Patientenzuordnungen im Rahmen von medizinischen Forschungsnetzen

Dr. rer. nat. Andreas Faldum, Prof. Dr. rer. nat. Klaus Pommerening
Universitätsklinik Mainz Institut für Medizinische Biometrie, Epidemiologie und Informatik
der Universitätsklinik Mainz
Obere Zahlbacher Straße 69 - D-55101 Mainz
Tel. +49-6131-17-3938 - Fax +49-6131-1747-3938
EMail: faldum@imsd.uni-mainz.de

1. Einführung:

In medizinischen Forschungsnetzen sollen Patienten durch eine eindeutige pseudonyme Identifikationsnummer (PID) charakterisiert werden. Fehler bei der Eingabe oder Übermittlung dieser Identifikationsnummer führen zu einer fehlerhaften Zuordnung des jeweiligen Patienten. Daraus können schwerwiegende Fehleinschätzungen bei der Behandlung des Patienten oder der statistischen Analyse des Patientenkollektivs folgen.

Da Fehler bei der Eingabe häufig vorkommen, werden in der Regel die Patientendaten doppelt erfasst. Ein Vergleich der Doppeleingabe soll dann helfen, Eingabefehler zu entdecken. Dieses Standardverfahren bedeutet allerdings eine erhebliche Erhöhung des Aufwandes und garantiert auch keinen vollständigen Schutz vor Fehlern.

Wünschenswert wäre daher die Möglichkeit, schon anhand der Patientenummer Fehler erkennen und eventuell automatisch korrigieren zu können, um den Patienten richtig zuzuordnen. Um eine solche Fehlererkennung oder -korrektur ohne Doppeleingabe zu ermöglichen, müssen der eigentlichen Identifikationsnummer Zusatzzeichen hinzugefügt werden. Für die Praktikabilität eines solchen Verfahrens ist es wichtig, mit möglichst wenig Zusatzaufwand eine höchstmögliche Sicherheit bei Korrektur und Fehlererkennung zu erreichen. Mit Hilfe codierungstheoretischer Methoden lassen sich Verfahren finden, die diesem Anspruch gerecht werden und effektiver arbeiten als die Standarddoppeleingabe.

2. Projekt:

Im Rahmen eines Pilotprojektes für das Kompetenznetz der Pädiatrischen Onkologie wurde am Institut für Medizinische Statistik und Dokumentation der Universitätsklinik Mainz eine PID entwickelt, die jedem Patienten eine 8stellige, pseudozufällig erzeugte Nummer zuordnet. Geplant ist der Einsatz des Verfahrens im Rahmen des Pseudonymisierungsdienstes der TMF.

Das zugrundeliegende Alphabet besteht aus den Ziffern 0-9 und den Buchstaben A-Z mit Ausnahme von B, I, O und S, die leicht mit Ziffern verwechselt werden können. Jeder Patient wird durch die ersten 6 Zeichen seiner PID charakterisiert. Die letzten beiden Stellen der PID stellen Prüfzeichen dar, mit deren Hilfe eine Fehlererkennung und -korrektur möglich wird.

Um diese Aufgabe bestmöglich zu erfüllen, werden die beiden Prüfzeichen so gewählt, dass die gültigen PIDs einen Maximum Distance Separable (MDS) Code der Länge 8 und Dimension 6 bilden. Dabei wird das zugrundegelegte Alphabet mit dem endlichen Körper $GF(32)$ identifiziert. Bei einem solchen MDS Code unterscheiden sich zwei verschiedene PIDs an mindestens drei Positionen. Somit können auch nach einem Eingabefehler Nummern noch eindeutig einer PID zugeordnet werden.

3. Korrekturkapazität des Verfahrens:

Aufgrund dieser Eigenschaft führt eine automatische Korrektur einer PID zu einem korrekten Ergebnis, falls bei der Eingabe der Identifikationsnummer nicht mehr als ein Fehler gemacht wurde. Darüber hinaus kann das Verfahren falsche Nummern erkennen, die bis zu zwei Fehler aufweisen. Mit Hilfe codierungstheoretischer Argumente lässt sich zeigen, dass die hier vorgestellte Methode bei nur zwei Prüfzeichen pro Identifikationsnummer die bestmögliche Fehlererkennung bzw. Fehlerkorrektur besitzt [1]. Der hier verwendete Code ist damit optimal.

Das Vertauschen benachbarter Zeichen stellt bei der Eingabe eines Patientenkennzeichens eine häufige Fehlerquelle dar. Da dabei genau zwei Fehler entstehen, wird – wie oben erwähnt – diese PID als inkorrekt erkannt. Wichtig ist aber zudem, dass bei einer automatischen Fehlerkorrektur keine falsche Identifikationsnummer erzeugt wird. Daher wurden die Prüfzeichen so gewählt, dass ein solcher Fehler nicht auftreten kann. Außerdem wird neben der Fehlererkennung auch die Position der vertauschten Zeichen eindeutig bestimmt.

Sowohl die Generierung einer korrekten PID, als auch die Fehlererkennung und –korrektur können durch einfache binäre Operationen erreicht werden. Dies stellt eine schnelle Verarbeitung der PIDs ohne Wartezeiten sicher.

4. Vertrauenswürdigkeit des Verfahrens:

Selbstverständlich kann selbst ein optimaler Code nicht verhindern, dass die Patientenzuordnung fehlerhaft wird, wenn die Fehlerrate bei der Eingabe sehr hoch ist. Die Fehlerkorrektur ist dann entweder nicht möglich oder führt zu falschen Resultaten. Es ist daher wichtig, die Vertrauenswürdigkeit des Verfahrens zu überprüfen. Zudem stellt sich die Frage, ob eine alleinige Fehlerkontrolle einer automatischen Korrektur vorzuziehen sei. Dazu muss die Wahrscheinlichkeit einer übersehenen falschen Identifikationsnummer oder fehlerhaften Korrektur bestimmt werden.

Unter Annahme der Gleichverteilung von PIDs und Eingabefehlern lassen sich diese Wahrscheinlichkeiten in Abhängigkeit der Eingabefehlerrate angeben [2]. Dabei kann aufgrund der pseudozufälligen Wahl der Identifikatoren von einer Gleichverteilung der PIDs ausgegangen werden. Die Verteilung der Eingabefehler hängt dagegen vom Eingabemedium und der Schreibtechnik - 1/10-Fingersystem - ab. Zur Abschätzung der Vertrauenswürdigkeit kann aber auch hier vereinfachend die Gleichverteilung vorausgesetzt werden.

Sei p die Wahrscheinlichkeit für einen Fehler bei der Eingabe eines Einzelzeichens. Dann beträgt die Wahrscheinlichkeit für eine übersehene falsche Identifikationsnummer $0,06 p^3$ und die Wahrscheinlichkeit für eine fehlerhafte automatische Korrektur $5,42 p^2$. Die alleinige

Fehlerkontrolle führt also stets zu den vertrauenswürdigeren Resultaten. Eine automatische Korrektur empfiehlt sich daher nur bei einer niedrigen Eingabefehlerrate. Da die Eingabefehlerrate durch die Rate entdeckter, falscher PIDs geschätzt werden kann, kann auch ein adaptives Verfahren zur Entscheidung, ob nur kontrolliert oder auch automatisch korrigiert werden soll, Anwendung finden. Welche Fehlerraten bei der Wahl der Strategie als noch akzeptabel gelten, hängt dabei natürlich von der jeweiligen Anwendungssituation ab.

Geht man in der Praxis z.B. von einer Fehlerrate der Einzelzeichen von 0,003 aus, so ergibt sich bei Anwendung einer automatischen Korrektur mit $4,9 \cdot 10^{-5}$ eine für die statistische Analyse des Patientenkollektivs ausreichend niedrige Rate falscher PIDs. Spielen die Identifikatoren jedoch bei der Behandlung der Patienten eine Rolle, so wäre eine solche Fehlerrate inakzeptabel. Die Wahrscheinlichkeit für eine übersehene falsche Identifikationsnummer bei reiner Fehlerkontrolle ist jedoch mit $1,6 \cdot 10^{-9}$ in jedem Falle ausreichend niedrig.

5. Vergleich mit anderen Verfahren:

Mit Hilfe der Codierungstheorie lässt sich zeigen, dass der hier verwendete MDS Code bei einer Redundanz von nur zwei Zeichen die größtmöglich erreichbare Vertrauenswürdigkeit bei Kontrolle und Korrektur bietet [2].

Vergleicht man das Standardverfahren der Doppeleingabe von je 6 Informationszeichen mit der 8stelligen PID, so führt die Doppeleingabe zu einem höheren Mehraufwand von 6 statt nur 2 Zeichen. Außerdem liegt bei der Doppeleingabe die Wahrscheinlichkeit für eine übersehene falsche Identifikationsnummer mit $0,19 p^2$ deutlich über der entsprechenden Wahrscheinlichkeit von $0,06 p^3$ bei Verwendung der 8stelligen PID. Die Möglichkeit einer automatischen Korrektur fehlt bei der Doppeleingabe ohnehin. Allerdings ist in dem oben angegeben Beispiel die Rate falscher Patientenzuordnungen bei Doppeleingabe mit $1,7 \cdot 10^{-6}$ nicht überragend besser als die bequeme automatische Korrektur der 8stelligen PID ($4,9 \cdot 10^{-5}$) und deutlich schlechter als die reine Fehlererkennung der 8stelligen PID ($1,6 \cdot 10^{-9}$).

Somit ergeben sich bei der in diesem Projekt verwendeten PID auch gegenüber dem Standardverfahren der Doppeleingabe deutliche Vorteile bezüglich dem Zusatzaufwand und der Vertrauenswürdigkeit des Ergebnisses.

6. Fazit:

Das hier vorgestellte Verfahren einer 8stelligen pseudonymen Identifikationsnummer ermöglicht bei nur 2 redundanten Zeichen pro PID die bestmögliche Fehlererkennung und -korrektur. Zudem bietet das Verfahren größtmögliche Sicherheit zur Vermeidung falscher Patientenzuordnungen. Mehr Sicherheit oder höheres Korrekturpotential sind nur mit einer größeren Redundanz erreichbar.

Eine automatische Korrektur von PIDs mit einem Fehler ist möglich. Eine flexible Anpassung an die jeweiligen Sicherheitsanforderungen bei verschiedenen Eingabefehlerraten kann durch Wechsel zwischen reiner Fehlerkontrolle und automatischer Korrektur adaptiv vorgenommen werden.

Zudem ermöglicht ein geringer Rechenaufwand bei Generierung, Kontrolle und Korrektur der PIDS ein unbehindertes Arbeiten.

Literatur:

[1] Wolfgang Willems:
Codierungstheorie. De Gruyter,
Berlin, 1999.

[2] Andreas Faldum: Trustworthiness of
Error-Correcting Codes.
ISCB-GMDS-99 Conference.