

Studien zur Performance des internistischen Konsultationssystems CADIAG-II/COLON

A. Rappelsberger¹, C. Lagor¹, K.-P. Adlassnig¹, W. Scheithauer², G.-V. Kornek²

¹Institut für Medizinische Computerwissenschaften, Spitalgasse 23, A-1090 Wien

²Univ.-Klinik für Innere Medizin I, Währinger Gürtel 18-20, A-1090 Wien

email: a.r@trulli.imc.akh-wien.ac.at

CADIAG-II (Computer-Assisted DIAGnosis) ist ein computergestütztes Konsultationssystem zur Assistenz der Differentialdiagnose in der Inneren Medizin. Es ist in der Lage, auf Basis von vorliegenden Krankheitszeichen in Frage kommende Diagnosen vorzuschlagen, diese gegebenenfalls zu beweisen oder auszuschließen. Falls notwendig werden weiterführende medizinische Untersuchungen vorgeschlagen. Jeder Schritt dieses Konsultationssystems wird detailliert begründet. CADIAG-II/Colon bearbeitet den Teilbereich der Colonerkrankungen; es enthält 37 Krankheiten unter Berücksichtigung von 436 Symptomen. Zur Erstellung der Wissensbasis wurden 2200 Beziehungen zwischen Symptomen und Krankheiten dokumentiert, davon 1152 für die Charakterisierung des Auftretens eines Symptoms bei einer Krankheit und 1048 als Wert für die Hinweiskraft des Symptoms für eine Krankheit. Interne Schranken für das Auftreten (A) und die Hinweiskraft (B) regulieren die Zahl der für die Hypothesengenerierung verwendeten Beziehungen. Ein weiterer Schrankenwert, der den Umfang der Hypothesengenerierung selbst bestimmt, ist H. Er gibt die Hinweiskraft vor, die erreicht werden muß, damit eine Diagnose als Hypothese generiert werden kann. Das Ziel der vorliegenden Studie [1] ist, die Hypothesengenerierung von CADIAG-II/Colon mit verschiedenen Werten für A, B und H zu evaluieren und zu bestimmen, ob dem Benutzer aufgrund der Ergebnisse eine optimale Einstellung empfohlen werden kann.

Für die Studie wurden 103 klinische Fälle mit insgesamt 119 Diagnosen von Colonerkrankungen verwendet. Pro Fall wurden von CADIAG-II/Colon unter Verwendung von sechs verschiedenen Einstellungen für A und B sowie fünf für H insgesamt 30 Hypothesenlisten generiert. Der grundlegende Evaluierungsschritt bestand im Vergleich dieser mit den entsprechenden Goldstandarddiagnosen; dabei handelte es sich sowohl um klinische Entlassungsdiagnosen als auch um histopathologisch oder chirurgisch bestätigte Diagnosen. Für die Interpretation wurden vier verschiedene Entscheidungskriterien zur Definition eines wahr-positiven Ergebnisses verwendet. Um Information zur Gesamtleistung des Systems zu erhalten, wurden ROC-Kurven (Receiver Operating Characteristic) generiert und Kappaindizes berechnet.

Die Ergebnisse bestätigen grundlegende Trends, die bereits in einer ähnlichen Studie zu CADIAG-II/Pancreas [2] beschrieben wurden. Anhand der vorliegenden Studie konnte als Folge der Verminderung von H die Erhöhung der wahr-positiven und falsch-positiven Raten verifiziert werden. Erneut konnte auch bei Verwendung von zunehmend streng gewählten Entscheidungskriterien die Tendenz zur Verschlechterung der Systemleistung beobachtet werden. Die Evaluierung ergab, daß die Schrankenwerte voneinander abhängig sind, indem A und B die untere Schranke für H vorgeben (und vice versa). Bei Verwendung der liberalsten Einstellungen für die Schrankenwerte generierte CADIAG-II/Colon 94% wahr-positiv, wobei die korrekte Hypothese in 53% der Fälle an erste Stelle gereiht war.

Gemäß den ROC-Kurven ist eine optimale Performance mit der liberalsten Definition eines wahr-positiven Wertes und einer Einstellung von $H=0,25$, $A=0,01$ und $B=0,01$ zu erreichen. Kappastatistiken zeigten eine angemessene Übereinstimmung in diesem Bereich. Nichtsdestotrotz scheint es jedoch ratsam, die Wahl der Schrankenwerte sowohl der jeweiligen klinischen Situation und dem daraus resultierenden Bedarf nach Sensitivität und Spezifität anzupassen, als auch die individuellen Schrankeneinstellungen der möglichen Krankheiten zu berücksichtigen.

Literatur:

[1] Lagor C., Adlassnig K.-P., Kornek G.-V., Scheithauer, W. Optimal Threshold Settings in CADIAG-II/COLON. *Cybernetics and Systems '98, Proceedings of the 14th European Meeting on Cybernetics and Systems Research*, Vol.1: 163-168, 1998

[2] Adlassnig K.-P. & Scheithauer, W. Evaluation of Medical Expert Systems Using ROC Curves. *Computers and Biomedical Research*, 22, 297-313, 1989