

IL DATA MINING NELLO SVILUPPO DI MODELLI PREVISIONALI PER LA GESTIONE INTEGRATA DELL'AGRICOLTURA SOSTENIBILE

Stefania D'Arpa¹, Vito Felice Uricchio^{1*}, Emanuele Barca¹

¹ Consiglio Nazionale delle Ricerche - Istituto di Ricerca Sulle Acque CNR-IRSA, Via F. De Blasio, 5 70132 Bari

* stefania.darpa@ba.irsra.cnr.it

** vitofelice.uricchio@cnr.it

Riassunto

Il lavoro traendo spunto dall' articolata letteratura scientifica in tema di applicazioni del *Data Mining* (DM) per la gestione integrata dell' informazione ambientale, traccia un percorso applicativo riferito alla valorizzazione dei dati agrometeorologici per favorire l'adozione di pratiche gestionali rispettose dell'ambiente.

In particolare si focalizza l'attenzione sulla classe di algoritmi di DM detti "supervisionati", con riferimento al *Time-Series Data Mining*, illustrandone le potenzialità di utilizzo per la costruzione e lo sviluppo di modelli previsionali utili alla definizione di processi decisionali in fase di gestione e pianificazione di interventi. Tale approccio è particolarmente significativo per la gestione integrata dei dati agrometeorologici con quelli di carattere ambientale riferiti in particolare alle caratteristiche pedologiche dei suoli ed alle loro caratteristiche chimico-fisiche, alla disponibilità e qualità delle acque, nonché ad aspetti socioeconomici e congiunturali.

Parole chiave: Sistemi Agrometeorologici, Agricoltura di Precisione, Modelli Previsionali, Data Mining, Time Series.

Introduzione

È sempre più condivisa l'opinione secondo cui nella gestione delle problematiche complesse, quali quelle ambientali, agrometeorologiche e agroforestali si debbano prediligere approcci integrati piuttosto che settoriali, in cui primaria importanza venga attribuita all'analisi delle relazioni tra le molte parti che contribuiscono alla definizione integrale dei fenomeni.

L'agrometeorologia con la complessità delle problematiche affrontate è una delle scienze che più si presta ad approcci di studio di tipo integrato ed i sistemi agrometeorologici, studiando i parametri ambientali che direttamente influenzano l'attività e la produzione agricola, sono ottimi strumenti che permettono la gestione di una delle priorità maggiori da conseguire in agricoltura ed in ambiente in generale: la sostenibilità.

Un tipico esempio di sostenibilità nei sistemi agricoli ci perviene dall'agricoltura di precisione in cui la gestione agricola del terreno tiene conto della variabilità intrinseca e indotta del suolo e delle specifiche esigenze delle colture al fine di migliorare la produzione, minimizzare i danni ambientali ed elevare gli standard qualitativi dei prodotti agricoli (Pierce e Nowak, 1990; National Research Council, 1997; Robert, 2002).

Strumenti indispensabili dell'Agricoltura di Precisione, sono i modelli matematici di simulazione che permettono di effettuare valutazioni agro ambientali ed economiche e previsioni di resa in funzione di cambiamenti climatici (Matthews, 2002).

Lo sviluppo di modelli previsionali potrebbe permettere, ad esempio, la previsione dei tempi di resa delle colture in condizioni di stagionalità meteorologica particolari per pianificare in modo ottimale l'attività agricola. L'accuratezza del modello predittivo fornirebbe il necessario supporto nell'evitare il più possibile sia sprechi nell'utilizzo delle risorse che costi associati alla pianificazione economica di attività e produzione.

Obiettivo del presente lavoro è illustrare come gli algoritmi di DM implementati nell'analisi di serie di dati agrometeorologici possono, attraverso il riconoscimento di patterns significativi, facilitare e valorizzare il processo di sviluppo dei modelli previsionali utili alla gestione sostenibile dell'attività agricola.

Di seguito saranno esposti sinteticamente strumenti e metodologie di DM più idonei e utili alla costruzione dei modelli previsionali più efficaci e meglio performanti con i dati futuri.

Materiali e metodi

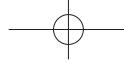
Le strategie di DM più indicate ai fini dell'efficacia del carattere predittivo nelle modellizzazioni agrometeorologiche sono quelle di DM supervisionato dove, partendo dall'obiettivo da prevedere (analisi top-down) generano previsioni, stime e caratterizzazioni del comportamento di alcune variabili target, individuate in funzione di variabili di input.

In tali ambiti l'analisi delle serie storiche (*Time series*) è da ritenersi di particolare importanza sia per le idonee caratteristiche dei dataset che per gli interessanti output conseguibili (Crone, 2009), che consentono di analizzare i fenomeni ponendoli in relazione con gli effetti ambientali ed agronomici, ottenendo così importanti stimoli di carattere gestionale.

I dataset considerati contengono dati sul fenomeno monitorato e informazioni sull'attività condotta, sulla collocazione geografica, sulle tecniche impiegate, sul tipo e sul numero di risorse impiegate ecc.

Una volta individuata la variabile dipendente monitorata da analizzare, le metodologie di DM valorizzando ed integrando il prezioso ruolo della fase di preparazione preliminare dei dati in cui la presenza di errori comporterebbe perdita d'informazioni e conseguentemente accuratezza, possono prevederne precisamente, i valori grazie all'ap-





plicazione ai dati di più potenti e più numerosi strumenti e algoritmi disponibili ed alla capacità di tali metodologie di giudicare il valore dei modelli risultanti basati sui rispettivi valori predittivi e descrittivi.

Discussione e conclusioni

Nell'ambito della gestione integrata ed in particolare della predittività, entrambe perfettamente convergenti nelle tematiche agrometeorologiche, il valore aggiunto apportato dall'uso delle tecniche di DM deriva dall'efficienza e dall'accuratezza con cui le sue metodologie ed i suoi algoritmi riescono a gestire i dataset agrometeorologici facilitando l'identificazione di input, predittori o indicatori che influenzano maggiormente le variabili dipendenti analizzate e di cui vogliamo prevedere il comportamento.

Gli svariati strumenti di screening di cui si avvalgono le metodologie di DM, infatti, consentono l'individuazione automatica dei predittori più importanti che influenzano la domanda di risorse, valorizzando e conferendo maggiore grado di precisione alla fase di pre-elaborazione dei dati. In tali ambiti possono essere compiute, ad esempio, segmentazioni dello spazio di input in diverse regioni, ognuna delle quali caratterizzata dal proprio valore statistico di previsione.

Ovviamente analisi di DM richiedono nella fase di modellizzazione dei dati l'indispensabile sperimentazione per testare l'efficienza dei diversi metodi.

Differenti tecniche potrebbero dare luce a nuove problematiche o confermare precedenti conclusioni.

Riassumendo potremmo asserire che le metodologie applicate di DM, contribuendo ad una migliore valorizzazione delle variabili di input, ad eccellenti correlazioni statistiche e alla superiore accuratezza nella modellizzazione, giocano un ruolo fondamentale nello sviluppo di modelli previsionali con le migliori prestazioni adeguate a fronteggiare le attuali sfide previsionali che la gestione integrata di sistemi complessi, come quelli agrometeorologici, richiede.

Bibliografia

- Crone S. F., 2009. Mining the past to determine the future: Comments. *International Journal of Forecasting*, 25, 456-460.
- Matthews R., 2002. *Crop-soil simulation models: Applications in developing countries*. CABI Publishing, Matthews, R. and Stephens, W, editors, Wallingford, 280 pp.
- National Research Council, 1997. *Precision Agriculture in the 21st Century*, Geospatial and Information technologies in Crop Management. National Research Academy Press, 16-19.
- Pierce F.S., Nowak P., 1990. Aspects of precision agriculture. *Advances in Agronomy*, 67, 1-86.
- Robert, P.C., 2002. Precision Agriculture: a challenge for crop nutrition management. *Plant and Soil*, 247, 1, 143-149.

