Argumentation-based Approach to Scientific Investigation: The Role of Cover Crops in Weed Suppression

Gheorghe Tecuci1, Dorin Marcu1, Muthukumar Bagavathiannan2, Steven Mirsky3, Alison Robertson4

1Learning Agents Center and Department of Computer Science, George Mason University, Fairfax, VA 22030, USA
2Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA
3Department of Plant Pathology and Microbiology, Iowa State University, 2213 Pammel Drive, Ames, IA 50011, USA
4Department of Computer Science, George Mason University, Fairfax, VA 22030, USA

tecuci@gmu.edu; dmarcu@gmu.edu; muthu@tamu.edu; steven.mirsky@usda.gov; alisonr@iastate.edu

Abstract

Weed management is a major challenge in conventional and organic production systems, amplified by the increasing resistance of weeds to herbicides. Cover crops offer a particularly promising strategy for management, as they minimize selection pressure by herbicides and synergistically improve weed control with existing herbicides, but the precise mechanisms through which cover crops provide weed suppression are not well-understood. Current approaches to scientific investigation use statistical machine learning to discover knowledge from data. This paper proposes a different artificial intelligence approach that synergistically integrates evidence-based reasoning, argumentation-based explanations, multi-strategy learning, and hypothesis-guided search in farm data to discover knowledge on how climate, soil, and weed seedbank size interact with cover crop biomass to drive weed suppression. The presented approach works with both large and small amounts of farm data, and is also applicable to other agricultural and food production domains.

Introduction

Weed management is a major challenge in field crop production systems. Overreliance on herbicides has led to the proliferation of herbicide-resistant weed biotypes at an alarming rate (Heap, 2020). Integrated weed management (IWM), which calls for the use of a combination of weed control tools, is touted as an effective strategy for improving weed control and reducing selection pressure exerted by individual weed control tools (Mortensen et al., 2012; Bagavathiannan and Davis, 2018).

Cover crops (CCs) are non-market crops that are planted between periods of cash crop production to provide a diverse array of ecosystem services including increased water and nutrient retention and availability, pest management, and greater soil health.

CCs offer a particularly promising strategy for IWM, as they minimize selection pressure by herbicides and synergistically improve weed control with existing herbicides (Mirskey et al., 2010; Norsworthy et al., 2012). For example, late season herbicide-resistant water hemp emergence was reduced up to 40% when CCs were used (Cornelius and Bradley, 2017). In another study, Wiggins et al. (2017) showed CCs with pre-plant residual herbicides significantly improved control of herbicide-resistant Palmer amaranth compared to no CC.

CCs outcompete weeds for resources while living, thus dominating the field and preventing weeds from growing (Osipitan et al., 2018) (Figure 1). Once terminated, they provide physical and chemical suppression which lowers weed germination, growth, and development, and reduces weed vigor and competition with cash crops (Wells et al., 2013; Palhano et al., 2018). Terminated CC mulches suppress weeds physically by impeding emergence or attenuating environmental cues that otherwise break weed seed dormancy (i.e., light and temperature) (Teasdale and...
Mohler, 1993; Teasdale and Mohler, 2000), by releasing phytotoxic compounds (i.e., allelopathy) (Creamer et al., 1996; Teasdale et al., 2012), and/or bio-geochemically by immobilizing soil nitrogen (another weed seed germination cue) in the case of high carbon/nitrogen (C:N) ratio grass CC mulches (Wells et al., 2013).

To provide sufficient season-long physical suppression of annual weeds, it has long been suggested that CC residues must be present in high amounts. In the mid-Atlantic region, Teasdale and Mohler (2000) showed that greater than 75% inhibition of weed emergence is consistently achieved only when CC mulch biomass exceeds 8,000 kg ha⁻¹ and mulch thickness exceeds 10 cm. In environments with considerably lower biomass production potential, however, Teasdale and Mohler (1993) showed intermediate residue levels can still be sufficient to limit light and temperature fluctuation cues that weed species often require for germination. More recent evidence suggests that even low (~2,500 kg ha⁻¹) to moderate (5,000 kg ha⁻¹) levels of CC biomass can have significant impacts on weed growth (Mirsy et al., unpublished) (Figure 2), through other mechanisms. These mixed results are speculated to be due to differences in climate, soil moisture and nutrient dynamics, and weed population densities (Mirsy et al., unpublished). CCs are clearly not a one-size-fits-all weed control tool because CC effects on weeds are highly variable across environments (Pittman et al., 2020). Farms differ in climate, soil, and management practices, all of which have been identified as primary factors influencing CC performance and subsequent impact on weed suppression. However, we have a severely limited understanding of how climate, soil, and weed density interact with CC performance (biomass production and C:N ratio) and their subsequent impact on weed suppression (Figure 3). This undermines our ability to integrate CCs into overall weed management programs in a meaningful way. Addressing this knowledge gap is critical for providing farmers with economically attractive and practically viable agronomic solutions that address the herbicide-resistant weed epidemic. Moreover, ecologically-based weed management strategies sought by stakeholder groups require site-specific knowledge to develop locally-adapted management strategies that consider the impacts of cropping system diversity, production practices, and environmental factors on the assembled weed community (Mortensen et al., 2012; Liebman et al., 2016). There is a critical need to consolidate findings and thereby derive a comprehensive understanding of factors controlling the effectiveness of CCs to suppress weeds and build this into decision support systems for growers.

Cereal rye (Secale cereale L.) is the most-used CC in the U.S. (CTIC and NCR SARE, 2016) due to low seed cost, broad geographic adaptability, and winter hardiness. In a recent farmer survey (n=1375), 59% of respondents reported

![Figure 2. Weed suppression by a terminated cereal rye CC even under low (~2,600 kg ha⁻¹) biomass production in a South Texas environment, suggesting potential interaction with other factors in driving weed suppression by the CC.](image)

![Figure 3. Causal relationships among different factors affecting weed germination, emergence and growth as affected by soil, climate, seedbank density, and CC biomass.](image)
having herbicide-resistant weeds (n=736), 25% reported that cereal rye ‘always improved’ weed control, and another 44% reported that cereal rye ‘improved’ control (CTIC, 2017). Cereal rye can produce biomass levels (without supplemental fertilization) of 5,000 to 7,000 kg ha⁻¹ at maturity, which may not completely eliminate weed emergence but can lower weed germination, emergence, growth rate, and biomass accumulation. Therefore, we will use cereal rye CC in field crop production systems (corn, soybean, and cotton) to illustrate our approach.

Current empirical and process-based models are inadequate at predicting weed suppression by CCs, therefore making reliable site-specific recommendations very challenging. On the other hand, there are large amounts of already-collected farm data on weeds and CCs. The question is: How can we discover new knowledge on weed suppression by CCs from existing data that are not uniform, are incomplete, and possibly partially incorrect?

This type of question is addressed by the emerging Science of Evidence (Schum, 2009). Evidence is any observable sign, datum, or item of information that is relevant in deciding whether a statement or hypothesis (e.g., a scientific claim) is true or false (Schum, 2001). Evidence is always incomplete, usually inconclusive (consistent with the truth of more than one hypothesis), frequently ambiguous (we cannot always determine exactly what the evidence is telling us), commonly dissonant (some evidence favors one hypothesis but other evidence favors other hypotheses), and has various degrees of credibility (Schum, 2001; Tecuci et al., 2016a).

Our previous research on intelligent knowledge-based agents has led to the development of the Disciple multistrategy apprenticeship learning approach to teaching agents rather than programming them (Tecuci, 1988; 1998; Boicu et al., 2001), and the demonstration of such agents in a variety of domains, including military planning (Tecuci and Hieb, 1996), course of action critiquing (Tecuci et al., 2000), military center of gravity determination (Tecuci et al., 2002; 2005), intelligence analysis (Tecuci et al., 2007; 2008), and cybersecurity (Huang et al., 2020).

Research on helping human analysts to perform evidence-based reasoning tasks has led to a computational theory of evidence-based reasoning (Tecuci et al., 2011; 2016a) and the development of cognitive assistants for analysis, such as Disciple-LTA (Tecuci et al., 2008; Schum et al., 2009), TIACRITIS (Tecuci et al., 2011), Disciple-CD (Tecuci et al., 2016a) and Cogent (Tecuci et al., 2018).

More recent work focuses on the ability of such agents to learn from their users how to assess the probability of hypotheses based on the available evidence by employing augmented Wigmorean argumentations (Tecuci et al., 2019; 2020). Wigmorean argumentations were initially introduced a century ago by Henry John Wigmore, a famous American jurist, as a graphical representation of how evidence supports or refutes claims in a court of law (Wigmore, 1913; 1937). They were resurrected by David Schum, who promoted their application both in law and in intelligence analysis (Schum, 1987; 2001). Their logical structure was augmented with Baconian probability (Cohen, 1977) and Fuzzy qualifiers (Zadeh, 1983), such as ‘likely’ or ‘almost certain’ (Tecuci et al., 2016a, pp. 159-172). These augmented Wigmorean networks use minimum/maximum probability combination rules common to the Baconian and Fuzzy probability views. These rules are much simpler than the Bayesian probability combination rule, which is important for the human understandability of the analysis.

The availability of data collected from previous farm experiences across wide regions in the U.S., together with the above developments in evidence-based reasoning and learning, provide an unprecedented opportunity to develop an AI approach to facilitate critical insights on the mechanisms and interactions of climate, soil, and weed density with CC biomass to suppress weeds in cropping systems. We present this approach in the next section.

**Illustration of the Discovery Approach**

Figure 4 presents the proposed sequence of steps to uncover the role of CCs for weed suppression as influenced by soil, climate, and cover crop performance. The Knowledge Base contains our incomplete understanding of the factors...
influencing weed growth.

The Case Data Base contains recorded data from a variety of past weed/CC experiences (called “cases”). Each case consists of evidence obtained from a specific farm in a specific year, such as CC biomass, weed biomass, and environmental factors (e.g., light, temperature, precipitation). The role of CCs in weed suppression appears to be complex and dependent on many factors, including geographical region, climate, and soil. On some farms CC resulted in small weed biomass, while on others CC had little effect and large weed biomass was observed. Even on the same farm, weed biomass following a CC can be different in different years.

**Step 1: Selection of a Reference Farm Case**

As shown in Figure 4, the first step of the investigation and knowledge discovery process is to select a reference farm case $F_{ref}$ that will guide the uncovering of knowledge applicable to the class of cases similar to it. In this illustration, $F_{ref}$ is the specific summer annual weed biomass experience on our reference farm case, Texas Farm1 in 2019. Low biomass of summer annual broadleaf weeds was attributed to a cereal rye CC.

The description of $F_{ref}$ consists of all the characteristics of $F_{ref}$ that may potentially relate to the resultant low biomass of summer annual broadleaf weeds, including light, temperature, precipitation, oxygen, soil moisture, soil N, C:N ratio of the residue, and allelopathic potential. Its ontological representation is illustrated in Figure 5. Notice that Texas Farm1 2019 had Plot1 planted with corn following a cereal rye CC with high CC biomass. It had low weed biomass of summer annual broadleaf weeds. As discussed later, the ontology plays a major role in our approach as the generalization hierarchy for learning.

**Step 2: Argumentation Explaining Weed Biomass**

In Step 2 (Figure 4) we use our current understanding of the factors influencing weed growth to explain the resulting low weed biomass on Texas Farm1 in 2019. We use Wigmorean argumentations to represent such explanations using the Cogent cognitive assistant. Figure 6, for example, shows a simple Wigmorean argumentation that explains the resultant low weed biomass on Texas Farm1 2019. It shows how the evidence $E_1$ of high CC biomass on Texas Farm1 in 2019 favors the hypothesis $H_1$ (The cover crop of cereal rye in Texas Farm1 2019 has high CC biomass), and how $H_1$ favors our main hypothesis $H$ (The summer annual broadleaf weeds in Texas Farm1 2019 with cover crop of cereal rye have low weed biomass).

First, one directly assesses the probability of hypothesis $H_1$ based on the item of evidence $E_1$ by assessing the three credentials of evidence: credibility, relevance, and inferential force, as shown in Figure 6.

The credibility of evidence answers the question: *What is the probability that the evidence is true?* As shown in the left-hand side of Figure 6, Cogent employs a system of symbolic probabilities with Fuzzy qualifiers, such as $BL$ (barely likely, 50 to 55% probability of being true), $VL$ (very likely, 80 to 95% true) or $C$ (certain, 100%). In this case the credibility of $E_1$ was assessed as $C$ (certain) because CC biomass was reliably measured as high (over 5,000 kg ha$^{-1}$). The relevance of evidence to a hypothesis answers the question: *What would be the probability of the hypothesis if the evidence were true?* In this case, if $E_1$ is true then $H_1$ is true, and therefore the relevance of $E_1$ is $C$ (certain). The inferential force or weight of the evidence on the hypothesis answers the question: *What is the probability of the hypothesis, based only on this evidence?* Obviously, an irrelevant item of evidence will have no inferential force, and will not convince us that the hypothesis is true. An item of evidence that is not credible will have no inferential force either. Only an item of evidence that is both relevant and credible may convince us that the hypothesis is true. Consistent with both the Baconian and the Fuzzy min/max...
probability combination rules, the inferential force of an item of evidence on a hypothesis is determined as the minimum between its credibility and its relevance which, in this illustration, is \( C \) (certain). Because in the situation from Figure 6 we have only one item of favoring evidence, its inferential force on the hypothesis is also the probability of the hypothesis. In general, however, the probability of the hypothesis would be the result of the balance of probabilities between the combined inferential force (maximum) of the favoring evidence items (under the left green square) and the combined inferential force of the disfavoring items (represented under the right pink square). The probability of the main hypothesis \( H \) is assessed in a similar way as \( VL \) (very likely), based on the probability of its sub-hypothesis \( H_1 \) (C) and the relevance (VL) of \( H_1 \) to \( H \).

We say that \( H_1 \) represents a favoring argument for the truthfulness of \( H \). Another favoring argument is represented by the direct evidence \( E_2 \) (the actual measurement of the weed biomass), as shown in Figure 7. Therefore, for our reference farm, Texas Farm1 in 2019, the explanation is consistent with the direct evidence.

**Step 3: Generalization of the Argumentation**

Next, we determine to what extent the developed argumentation shown in Figure 6 also explains weed biomass produced in other cases similar to that of \( F_{ref} \). This involves a process of knowledge-based learning and evidence-based reasoning where the specific argumentation is automatically generalized to an argumentation pattern and an associated applicability condition, shown in Figure 8. For example, the specific argumentation in Figure 7 will be generalized to the argumentation pattern from the right-hand side of Figure 8 by:

- Replacing each instance (e.g., Texas Farm1 2019) with a variable (i.e., ?O1);
- Replacing each evidence item (e.g., E1 High CC biomass in Texas Farm1 2019) with an evidence collection request. This evidence collection request will call a specialized collection agent that will automatically search the Case Data Base for the evidence specified in an instantiated request.

Additionally, the learning process will generate two bounds for the variables used in the pattern, indicating the possible values of these variables. These bounds will be obtained as minimal and maximal ontology-based generalizations, respectively, of the corresponding instances from the argumentation. For example, the minimal generalization of Texas Farm1 2019 would be the concept Texas Farm1 in Figure 5 (i.e., the generalized argumentation is expected to be applicable for any weed suppression experience on this farm), while the maximum generalization would be any farm (i.e., the generalized argumentation might also be applicable for any weed suppression experience on any farm including, for example, Maryland Farm4 2019 in Figure 5). The two bounds will converge toward one another based on additional argumentations developed during the discovery process.

**Step 4: Discovery of Favoring and Disfavoring Cases**

This step involves a process of knowledge-based search and classification where the generalized argumentation is automatically applied to cases similar to the reference case \( F_{ref} \), splitting these cases into a set of favoring cases and a set of disfavoring cases.

**Step 5: Selection of the Most Similar Disfavoring Case**

The existence of disfavoring cases shows that the argumentation from Figure 6 is incomplete or partially incorrect. We have to discover what factors were not taken into account, and improve this argumentation. To facilitate this complex knowledge discovery process, we select a new case \( F_s \) from the set of disfavoring ones that is most similar to our reference farm case \( F_{ref} \) (i.e., Texas Farm1 2019) because there will be very few factors that are different, some of which are responsible for the difference in weed biomass. This farm case \( F_s \) might be, for example, the CC experience on the same Texas farm in the previous year (i.e., Texas Farm1 2018). The corresponding instantiation of the generalized argumentation from Figure 8 is shown in Figure 9. Notice that, in this case, the direct evidence \( E_4 \) disfavors...
the top hypothesis because the actual weed biomass produced was high.

---

**Step 6: Hypothesis-Driven Explanation Discovery**

Now we have to hypothesize an explanation for the difference in weed biomass between the two very similar cases $F_{ref}$ and $F_s$. In this instance the hypothesized explanation is that Spring 2018 in Texas was wet, while Spring 2019 was dry. The wet Spring 2018 resulted in high soil moisture that led to high weed seedling density and high weed biomass. In contrast, the dry Spring 2019 led to low weed seedling density and low weed biomass.

**Step 7: Refinement of the Argumentation**

Based on the discovered explanation (Spring moisture) the argumentation in Figure 7 is refined as shown in Figure 10, generalized as discussed in Step 3, and automatically applied to all similar cases in Step 4. If the total number of favoring cases is not increased (or, equivalently, the total number of disfavoring cases not decreased), the explanation formulated in Step 6 is rejected and we have to hypothesize a new explanation. If, however, the total number of favoring cases is increased, the explanation is accepted and represents new discovered knowledge.

**Loop 5-6-7-3-4: Learning Explanation Models**

The 5-6-7-3-4 loop (Figure 4) is repeated, leading to new discovered knowledge, until all cases similar to $F_{ref}$ are correctly classified, with the possible exception of a few anomalous cases. The result of this process is the discovery of new knowledge and a generalized argumentation that correctly infers and explains weed biomass for the class of cases similar to $F_{ref}$.

The entire process is then repeated for other classes of cases from the Case Data Base, leading to the discovery of additional knowledge about weed biomass suppression and the development of additional argumentations.

**Discussion**

*How much data is needed?* The more data the better, but the proposed approach works with both large and small amounts of data. Assume, for example, the most extreme case when only two cases are available. In this situation, one can still follow the steps in Figure 4. In Step 1, one of these two cases will be selected as $F_{ref}$. In Step 2, the argumentation that explains the phenomenon on $F_{ref}$ is developed, and in Step 3 it is generalized, so that it can be applied to the second case. If this case is classified as disfavoring (Step 4), then it will be selected as the most similar case $F_s$ (Step 5), an explanation of the differences is discovered (Step 6), and the argumentation (prediction model) is refined (Step 7). The explanation of the differences between these two cases is the discovered new knowledge. Of course, the generality of this new knowledge is minimal because it is based on only two farm cases. However, it will increase in the future when new cases become available. If the second case is classified as favoring (Step 4), the two cases are consistent with respect to the prediction model and there is nothing to discover.

*Does the data need to be complete and uniform?* No. Our approach treats data as evidence on the considered hypotheses. It employs Wigmorean argumentations integrating Baconian and Fuzzy probabilities that cope with evidence that is incomplete, inconclusive, ambiguous, dissonant, and with various degrees of accuracy.

*Do we need to perform actual farm experiments?* No because we can test the formulated hypotheses on previously collected data. Scientists might wish, however, to perform additional experiments.
Current Status and Future Research

We are currently developing an intelligent system, called the Knowledge Discovery Assistant (KDA) that implements the presented approach. Many of the modules of the KDA are customizations of modules of Cogent, a cognitive assistant for intelligence analysis (Tecuci et al., 2018) and of Disciple-EBR, a tool set for developing agents for evidence-based reasoning tasks (Tecuci et al. 2016b). Once the prototype of the KDA is developed, it will be used by agricultural scientists to discover knowledge in three areas, one being the use of CCs for weed suppression (discussed above). The other two are briefly presented below.

Biomass Accumulation of Cover Crops. As previously mentioned, CCs are non-marketed crops that are planted between periods of cash crop production to provide a diverse array of ecosystem services including increased water and nutrient retention and availability, pest management, and greater soil health. As a result, CCs could also increase crop yield stability and resilience in a changing climate (Villamil et al., 2006; Kaspar et al., 2001; Meisinger et al., 1991; Ruffo et al., 2004; Shipley et al., 1992). To reap various ecosystem services (reduced soil erosion, improved nutrient cycling, etc.), substantial CC biomass production is required. Extrinsic and intrinsic factors affect biomass production. We need to discover knowledge of how abiotic and biotic factors, and management practices (planting and termination dates, and methods) influence biomass accumulation across corn-, soybean-, and cotton-growing U.S. regions.

Foliar Fungicide Applications to Field Crops. In the past two decades, fungicide use on field crops has increased considerably (Mueller et al., 2021). Several factors have been suggested for this growth in fungicide use including increased commodity prices, more fungicide products registered for use on field crops, greater disease prevalence, and marketing (Wise and Mueller, 2011). Fungicides have traditionally been applied to reduce disease, however, in more recent years they have been applied for their physiological plant effects that may contribute to yield increases. Research in soybean and other crops, however, indicates these physiological effects are inconsistent, or do not always result in a measurable yield increase (Bertelsen et al., 2001; Khan and Carlson, 2009; Robertson et al., 2020). We need to understand the abiotic and biotic, management, and environmental factors that contribute to greater yields in field crops to identify situations where an application of a fungicide could result in greater yields and a return on investment.

Acknowledgements

David Schum (1932 – 2018) has significantly influenced our work on evidence-based reasoning. Mihai Boicu contributed to our research on knowledge representation, reasoning and learning, and the development of cognitive agents. Victoria Akroyd provided useful comments on this research.

References


