Richtlinie zum „Software-Sprint“

Visual-Bayesic

Schlussbericht

Zuwendungsempfänger: Maximilian Hallenberger & Miguel de la Varga GbR

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS23S45 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Kurze Darstellung der Aufgabenstellung und Motivation

Was war Deine Motivation? Welches Problem wolltest Du mit Deinem Projekt lösen? Wie war die geplante Vorgehensweise zur Problemlösung (auch Angabe der wichtigsten Meilensteine)?

Unser Projekt „Visual Bayesic“ wurde mit der Motivation ins Leben gerufen, ein häufiges und grundlegendes Problem in unserer Gesellschaft anzugehen: die Schwierigkeit vieler Menschen, ungewisse Daten korrekt zu interpretieren.

Die Aufgabenstellung unseres Projekts war es, einen Beitrag zur Lösung dieses Problems zu leisten, indem wir ein Werkzeug entwickeln, das probabilistisches Programmieren und Bayes'sche Verfahren für ein breiteres Publikum zugänglich macht. Durch die Vereinfachung und Visualisierung komplexer statistischer Konzepte wollen wir Menschen befähigen, bessere Entscheidungen zu treffen und ihre Fähigkeit zur kritischen Analyse von Daten zu steigern.

Unsere geplante Vorgehensweise bestand darin, eine Plattform zu entwickeln, die bestehende „visual scripting“ Lösungen mit dem probabilistischen Framework pyMC3 verknüpft. Der Schwerpunkt lag dabei auf der Benutzerfreundlichkeit und der Präsentation einer transparenten und verständlichen Nutzererfahrung.

Die wichtigsten Meilensteine unseres Projekts umfassten:

- Funktionsplanung: Identifikation der benötigten Ziel-Funktionalitäten und Dokumentation in einem Arbeitsplan, einschließlich der Definition konkreter Beispielanwendungen.
- Evaluation des Tech-Stacks: Prüfung des geplanten Tech-Stacks auf Kompatibilität und Vollständigkeit in Bezug auf die definierten Ziele.
- Implementierung der visuellen Bausteine: Einbindung der Funktionalität in eine Visual-Scripting-Umgebung.
- Implementierung von Beispielen und Dokumentation: Umsetzung von Beispielanwendungen und nutzerseitige Dokumentation.

Beitrag des Projektes zu den Zielen der Förderinitative „Software-Sprint“

Welche Bezüge gibt es zu den Themenfeldern „Civic Tech“ und „Data Literacy“ des Software Sprints oder zu weiteren gesellschaftlich relevanten Zielen bzw. Lösungsansätzen?

Unser Projekt "Visual Bayesic" leistet einen signifikanten Beitrag zu den Zielen der Förderinitiative „Software-Sprint“, insbesondere in den Bereichen „Civic Tech“ und „Data Literacy“.

Civic Tech: "Visual Bayesic" trägt zum Feld des Civic Tech bei, indem es eine Plattform bietet, die es Bürgern ermöglicht, komplexe Daten besser zu verstehen und zu nutzen. Durch die Visualisierung und Vereinfachung bayes´scher statistischer Methoden unterstützt unser Tool aktive Bürgerbeteiligung und informierte Entscheidungsfindung.

Data Literacy: Unser Projekt fördert Data Literacy, indem es Anwendern die Werkzeuge und das Wissen an die Hand gibt, um Daten kritisch zu analysieren und zu interpretieren. "Visual Bayesic" erleichtert das Verständnis probabilistischer Konzepte und fördert somit eine fundierte Datenkompetenz, die für das Verständnis und die Analyse von Informationen in verschiedenen Bereichen wie Medizin, Klimawandel und Wirtschaft unerlässlich ist.

Durch die Bereitstellung von Visual Bayesic als Open-Source-Paket tragen wir zudem zur Demokratisierung von fortschrittlichen Datenanalyse-Tools bei. Dies ermöglicht einen breiten Zugang zu wertvollen Ressourcen und fördert die allgemeine Datenkompetenz, die für die Bewältigung der Herausforderungen des 21. Jahrhunderts entscheidend ist.

Ausführliche Darstellung der Ergebnisse

Welche konkreten Ergebnisse hast Du erzielt? Konnten alle Meilensteine erreicht werden? Welche zusätzlichen Erkenntnisse hast Du aus der Projektarbeit gewonnen, auch im Hinblick auf die Begleitung durch die Open Knowledge Foundation?

Im Rahmen unseres Projekts "Visual Bayesic" haben wir eine Reihe konkreter Ergebnisse erzielt, die einen bedeutenden Fortschritt in der Nutzung von Visual Scripting zur Definition Bayesianischer Netzwerke darstellen.

Zunächst wurden drei Visual Scripting-Frameworks in Python – Xircuits, Ryven und Pyflow – intensiv getestet und deren Vor- und Nachteile analysiert. Diese Analyse war entscheidend für unsere Entscheidung, welches Framework am besten für die Umsetzung unserer Ziele geeignet ist. Jedes Framework bot unterschiedliche Stärken und Schwächen, was uns wertvolle Einblicke in die Komplexität und die Möglichkeiten der Visual Scripting-Technologie gab.

Ein zentraler Erfolg unseres Projekts war die Validierung der Idee, Visual Scripting als graphische Schnittstelle zur Definition von Bayesianischen Netzwerken zu nutzen. Durch die Entwicklung von Lehrbeispielen konnten wir die Grundlagen der Bayesianischen Inferenz für Anfänger verständlich machen. Diese Beispiele waren sowohl für das Verständnis des Konzepts als auch für die praktische Anwendung von Visual Bayesic von großer Bedeutung.

Darüber hinaus haben wir ein neues GitHub-Repository erstellt und mit „Sphinx“ verknüpft um eine automatische Dokumentationen zu garantieren. Dies erleichtert anderen Entwicklern und Nutzern die Arbeit mit Visual Bayesic, da sie leicht auf umfassende und gut strukturierte Informationen zugreifen können.

Die Implementierung eines Pytest-Frameworks war ein weiterer wichtiger Schritt, um die Qualität und Zuverlässigkeit unserer Software zu gewährleisten. Es ermöglichte uns, unsere Entwicklungen kontinuierlich zu testen und zu validieren, um eine hohe Benutzerzufriedenheit sicherzustellen.

Zu guter Letzt konnten wir erfolgreich Beispiele, die ursprünglich mit der programmiertechnischen API Bayesianischer Frameworks erstellt wurden, vollständig mit der Visual Scripting-Schnittstelle reproduzieren (siehe Abbildung 1). Dies war ein bedeutender Meilenstein, der die Machbarkeit und Effektivität unseres Ansatzes unter Beweis stellte.

Bei der Feinabstimmung der Benutzeroberfläche haben wir mit verschiedenen Granularitätsstufen der Knoten experimentiert, um die Benutzerfreundlichkeit weiter zu verbessern. Dieser iterative Prozess half uns, ein ausgewogenes Verhältnis zwischen Funktionalität und Benutzererfahrung zu finden.

Insgesamt haben diese Ergebnisse gezeigt, dass unser Ansatz, komplexe statistische Konzepte durch Visual Scripting zugänglicher zu machen, erfolgreich ist und das Potenzial hat, das Verständnis und die Anwendung Bayesianischer Methoden in verschiedenen Bereichen zu verbessern.

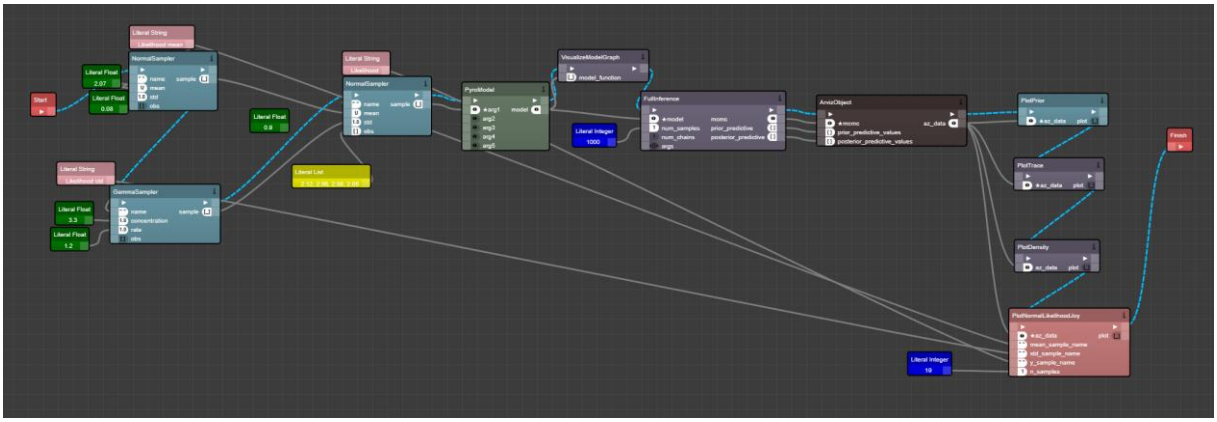


Abbildung 1. Bayesianisches Netzwerk erstellt durch die Verknüpfung von “Visual Scripting“ Knoten.

Unser Projekt hat damit abschließend die vier wichtigsten Meilensteine (Funktionsplanung, Evaluation des Tech-stacks, Implementierung der visuellen Bausteine, Implementierung von Beispielen und Dokumentation), die im Projektantrag umschrieben wurden, erreicht.

Zielgruppe, Nutzen und mögliche Weiterentwicklungen

Welcher Nutzen ergibt sich für die Zielgruppe aus den Ergebnissen Deines Projekts? Welche weitergehenden Effekte ergeben sich aus der Open-Source-Stellung der Ergebnisse? Gibt es Ideen für die Weiterentwicklung Deiner Lösung und Pläne zu deren Umsetzung?

Hat die Arbeit in dem Projekt Dich in Deiner persönlichen, fachlichen Weiterentwicklung unterstützt?

Zielgruppe und Nutzen: "Visual Bayesic" zielt darauf ab, eine breite Palette von Nutzern zu unterstützen, insbesondere Wissenschaftler, Studierende, Lehrkräfte und Fachleute aus verschiedenen Bereichen wie Medizin, Klimawissenschaft und Wirtschaft. Durch die Vereinfachung und Visualisierung komplexer bayes‘scher Statistikmethoden ermöglicht unser Tool diesen Nutzern, die Konzepte probabilistischen Denkens zu erlernen und praktisch anzuwenden. Dies fördert nicht nur das Verständnis für Unsicherheiten in empirischen Daten, sondern unterstützt auch bei der fundierten Entscheidungsfindung. Darüber hinaus profitieren Wissenschaftskommunikatoren von "Visual Bayesic", indem sie komplexe Sachverhalte einem breiteren Publikum zugänglich machen können.

Weitergehende Effekte der Open-Source-Stellung: Als Open-Source-Projekt fördert "Visual Bayesic" die Kollaboration und den freien Austausch von Wissen und Technologie. Dies ermöglicht es anderen Entwicklern und Forschern, das Tool weiterzuentwickeln, anzupassen und in neuen Kontexten zu verwenden. Die Open-Source-Natur von "Visual Bayesic" sorgt für Transparenz und Vertrauensbildung in der wissenschaftlichen Gemeinschaft und darüber hinaus.

Weiterentwicklungsideen und Pläne: Wir planen, "Visual Bayesic" kontinuierlich weiterzuentwickeln, um seine Funktionalität und Benutzerfreundlichkeit zu verbessern. Zukünftige Updates könnten erweiterte Visualisierungsoptionen, verbesserte Algorithmen zur Datenanalyse und eine noch intuitivere Benutzeroberfläche umfassen. Außerdem planen wir, die Plattform, um spezifische Module für verschiedene Anwendungsbereiche zu erweitern, um ihre Anwendbarkeit und Zugänglichkeit zu erhöhen.

Persönliche und fachliche Weiterentwicklung: Die Arbeit an "Visual Bayesic" hat uns in unserer persönlichen und fachlichen Entwicklung erheblich unterstützt. Die Herausforderungen bei der Entwicklung des Tools haben unser Verständnis für probabilistische Modellierung und Softwareentwicklung vertieft. Zudem hat die Zusammenarbeit mit einem breiten Spektrum von Nutzern und Stakeholdern unsere Fähigkeiten in den Bereichen Kommunikation, Projektmanagement und interdisziplinäre Zusammenarbeit verbessert.

Kurze Darstellung der Arbeiten, die zu keiner Lösung geführt haben

Gab es Arbeiten bzw. Lösungsansätze, die nicht weiterverfolgt wurden? Was waren die Hintergründe, und wie bist Du alternativ vorgegangen?

Während der Entwicklung von „Visual Bayesic“ gab es Arbeitsbereiche und Lösungsansätze, die schlussendlich nicht weiterverfolgt wurden. Diese Herausforderungen waren Teil unseres Lernprozesses und halfen uns, unseren Fokus zu schärfen und effektivere Lösungen zu entwickeln.

Einer der Hauptaspekte, bei dem wir Herausforderungen erlebten, betraf die Auswahl von Frameworks sowohl für das visuelle Scripting als auch für die Bayesianische Inferenz. Es gab keine Lösung, die alle unsere Anforderungen erfüllte, was zu Kompromissen auf beiden Seiten führte. Die endgültige Auswahl des Frameworks war besonders geeignet, um Python-Code zu generieren, bot jedoch nur eingeschränkte Möglichkeiten, Plots oder andere Datentypen einzubetten. Unser Idealziel ist es, „Visual Bayesic“ so weiterzuentwickeln, dass es unabhängig von jeglichen Backend-Frameworks ist, um unseren Nutzern eine flexible Lösung für verschiedene Problemtypen anbieten zu können.

Ein weiterer Bereich, der Schwierigkeiten bereitete, war die Erstellung komplexer synthetischer Netzwerke. Jedes Bayesianische Modell weist verschiedenste Besonderheiten auf, welche die Konstruktion komplexer Netzwerke schwieriger machten als erwartet. Daher konzentrierten wir uns darauf, einfachere und bildungsorientierte Modelle zu entwickeln. Diese Vorgehensweise erwies sich als effektiver, um die Grundprinzipien der Bayesianischen Inferenz verständlich zu machen.

Kurze Angabe von Präsentationsmöglichkeiten für mögliche Nutzer

Wo können sich Interessenten detailliert über Deine Projektergebnisse informieren (z.B. Webseite, GHitHub, Veröffentlichungen)?

Für Interessenten, die sich detaillierter über die Ergebnisse und Anwendungen unseres Projekts "Visual Bayesic" informieren möchten, bieten wir folgende Möglichkeiten:

Projekt-Webseite: Wir haben eine umfassende Webseite eingerichtet, die ausführliche Informationen über das Projekt, seine Ziele, Anwendungsfälle und Tutorials enthält. Diese Webseite ist eine zentrale Anlaufstelle für alle, die sich mit den Funktionen und dem Nutzen von "Visual Bayesic" vertraut machen möchten.

https://gempy-project.github.io/visual_bayesic/

GitHub-Repository: Für Nutzer und Entwickler bieten wir ein GitHub-Repository, das den vollständigen Quellcode des Projekts sowie Dokumentationen zur Installation, Nutzung und Beitragserbringung enthält. Dies ermöglicht es der Community, aktiv an der Weiterentwicklung des Tools teilzunehmen und eigene Verbesserungen oder Erweiterungen vorzuschlagen. Unser Repository ist zugänglich unter:

https://github.com/gempy-project/visual_bayesic

Durch diese verschiedenen Präsentationsmöglichkeiten streben wir danach, eine breite Palette von Nutzern zu erreichen und eine aktive, engagierte Gemeinschaft um "Visual Bayesic" aufzubauen.

Kurze Erläuterung zur Einhaltung der Arbeits- und Kostenplanung

Gab es im Projektverlauf Ereignisse, die eine Anpassung der Planung erforderlich machten – z.B. Mehr- oder Minderaufwand bei der Bearbeitung von Teilaufgaben?

Im Rahmen unseres Projekts wurden die meisten unserer Arbeitsprozesse gemäß unserem ursprünglichen Plan absolviert. Allerdings führten einige unserer Experimente, die nicht zu den gewünschten Lösungen führten, zu notwendigen Anpassungen in unserer Planung.

Diese Anpassungen waren hauptsächlich darauf zurückzuführen, dass wir bei der Auswahl der Frameworks für Visual Scripting und Bayesianische Inferenz verschiedene Ansätze ausprobieren mussten, bevor wir eine geeignete Lösung fanden. Diese Phase des Ausprobierens und Evaluierens, obwohl sie nicht direkt zu einer endgültigen Lösung führte, war für unser Verständnis und die Weiterentwicklung des Projekts entscheidend.

Obwohl diese Anpassungen erforderlich waren, führten sie nicht zu bedeutenden Überschreitungen unserer geplanten Kosten oder Zeitrahmen.

Zusätzlich möchten wir anmerken, dass der Zeitaufwand für Verwaltungsaufgaben, wie die Gründung einer GbR und die Abwicklung der Steuererklärungen usw., etwas höher waren als ursprünglich antizipiert. Diese Aufgaben waren zwar nicht direkt mit der technischen Entwicklung von "Visual Bayesic" verbunden, jedoch für die ordnungsgemäße Durchführung und Abwicklung des Projekts unerlässlich.

Zusammenfassend lässt sich sagen, dass unser Projekt trotz einiger notwendiger Anpassungen im Laufe der Entwicklung effizient und im Einklang mit den geplanten Ressourcen durchgeführt wurde. Diese Anpassungen waren Teil eines dynamischen Entwicklungsprozesses und trugen dazu bei, das Endprodukt zu verbessern und unser Verständnis für die zugrundeliegenden technischen Herausforderungen zu vertiefen.

Kurze Darstellung von etwaigen Ergebnissen bei anderen Stellen

Gab es Entwicklungen anderer Personen oder Institutionen, die Einfluss auf Deine Arbeiten und die Zielsetzung hatten? Wenn ja, worin bestand dieser und wie bist Du damit umgegangen?

Im Verlauf unseres Projekts „Visual Bayesic“ gab es keine nennenswerten Entwicklungen bei anderen Personen oder Institutionen, die einen direkten Einfluss auf unsere Arbeit oder unsere Zielsetzung gehabt hätten