

Riconoscimento vocale di formule

Matteo Amore⁴, Tiziana Armano¹, Cristian Bernareggi³, Anna Capietto¹, Sandro Coriasco¹, Roberta Crespan⁴, Mattia Ducci¹, Maria Luisa Gabrielli⁴, Maria Francesca Guadalupi⁴, Alessandro Mazzei², Antonio Mazzei⁴, Adriano Sofia¹, Francesco Tarasconi⁴

¹Dipartimento di Matematica – Università di Torino, ²Dipartimento di Informatica – Università di Torino, ³Università di Milano, ⁴Maize srl

Abstract. Le tecnologie in generale, ed in particolare le tecnologie assistive e gli strumenti compensativi, hanno attualmente un livello di sviluppo elevato ma tuttavia presentano criticità nella gestione di contenuti scientifici (formule e grafici). Queste problematiche disincentivano persone con disabilità e disturbi specifici di apprendimento (DSA) ad accedere a percorsi di formazione di tipo scientifico e precludono in seguito la possibilità di accedere a professioni STEM. Il Laboratorio per la Ricerca e la Sperimentazione di Nuove Tecnologie Assistive per le STEM "S. Polin" del Dipartimento di Matematica "G. Peano" dell'Università di Torino si occupa di sviluppare nuove soluzioni tecnologiche e testare, adattare e diffondere le tecnologie già esistenti allo scopo di risolvere i problemi relativi all'accesso dei contenuti STEM (Armano et al, 2020). In questo ambito il Laboratorio si occupa anche di ricerca e sviluppo in relazione al riconoscimento vocale di formule.

Keywords. Accessibilità, formule, riconoscimento vocale, disabilità, DSA.

Introduzione

Il progetto Matematica a voce, cofinanziato da Fondazione CRT, nasce dall'esigenza di avere sistemi per la dettatura e la trascrizione di formule che sarebbero molto utili a persone con disabilità motorie (permanenti o temporanee), uditive e a persone con DSA.

L'“oggetto” video in generale ed in particolare i video di lezioni (molto diffusi in questo periodo in cui prevale la didattica a distanza) presentano diversi problemi di accessibilità: servono sottotitoli adeguati per persone con disabilità uditive e la trascrizione per persone con disabilità motorie, visive e con DSA che hanno difficoltà a prendere appunti. Esistono attualmente numerosi sistemi per la trascrizione e la sottotitolazione automatica ma nessuno gestisce le formule in modo opportuno: per la fruizione di un video con contenuti scientifici serve la trascrizione di formule in linguaggio specifico (LaTeX o MathML).

I sistemi di interazione vocale, molto diffusi ed evoluti, hanno scarse funzionalità di gestione del linguaggio matematico mentre le persone con disabilità motorie agli arti superiori, permanenti o temporanee, avrebbero bisogno di strumenti per scrivere ed elaborare in autonomia le espressioni matematiche. Alcuni software promettenti in questo campo come MathTalk ed Equatio funzionano solo per la lingua inglese.

Il nostro progetto prevede due moduli relativi a trascrizione e sottotitolazione di video di lezioni con formule e dettatura e gestione di formule matematica:

1. VoiceMath: sviluppo di software per la trascrizione in testo + LaTeX (o MathML) di lezioni universitarie “con formule”, utile a persone con disabilità motorie, uditive e con DSA;
2. SpeechMatE: sviluppo di software per la dettatura e la modifica di formule matematiche con il riconoscimento della lingua italiana, utile a persone con disabilità motorie.

1. VoiceMath

La soluzione ha come obiettivo la produzione automatica di un flusso ordinato di contenuti elettronici (txt e LaTeX) a partire da una videoregistrazione di lezioni e corrispondenti immagini del testo scritto a mano dall'insegnante. Le materie in oggetto sono: algebra/logica, analisi, fisica matematica, geometria e probabilità/statistica. Per raggiungere questo scopo saranno sviluppate componenti basate su Machine Learning, in particolare di Speech-to-Text e Computer Vision. Partendo dalla videoregistrazione di una lezione reale bisognerà analizzare, con strumenti e modalità specifiche, sia la componente audio, cioè il parlato del docente, sia le immagini, costituite dal testo scritto dal docente durante la lezione.

L'audio, isolato dalla videoregistrazione, dovrà essere sottoposto a trascrizione automatica mediante tecnologia di Natural Language Processing Speech-to-Text (Jurafsky, Manning, 2012). Attualmente è in uso il trascrittore Nuance Transcription Engine (NTE3). Il progetto ha avuto una fase esplorativa, con obiettivo quello di verificare eventuali problematiche degli audio, valutare l'accuratezza della trascrizione automatica ed effettuare le dovute attività di post processing. Per effettuare la valutazione dell'accuratezza e gestire le miglitorie, siamo partiti da dalle trascrizioni manuali delle lezioni (Savy 2005), che sono valse come standard. Il focus di queste lezioni, ovvero le formule, è stata la parte più sfidante da dover trascrivere: i trascrittori automatici sono generalmente calibrati attorno a modelli linguistici generalistici, mentre i nostri audio trattano una terminologia specifica. In linguistica questa viene definita come terminologia di dominio. Per migliorare la trascrizione anche di formule ed espressioni specifiche del linguaggio matematico abbiamo prodotto un file specifico detto DLM (Domain Language Model), contenente tali termini, per un numero ridotto di audio; tale file ha come obiettivo quello di andare ad addestrare il trascrittore, per renderlo più tarato su quello che è il dominio applicativo. Il test effettuato ha evidenziato delle miglitorie per le trascrizioni automatiche, motivo per cui puntiamo ad ampliare tale DLM e ad utilizzarne uno specifico per insegnamento. Al momento, siamo ancora in una fase di sperimentazione e di analisi prevalentemente qualitativa degli errori, quindi non escludiamo di portare avanti nuovi test successivamente.

Il modulo “Image to Latex” si occuperà di identificare in modo automatico le formule contenute in un'immagine e di convertirle in una stringa LaTeX. Il dataset finora conta circa 50 immagini per ognuna delle materie analizzate. In una prima fase ci siamo occupati di analizzare le performance di Mathpix, tool che si occupa della “digitalizzazione” di testi stampati o handwritten con contenuti matematici e scientifici in generale. Le immagini originali, corrispondenti solitamente a intere pagine, generalmente contengono molte in-

formazioni: per migliorare le analisi delle componenti successive, le abbiamo segmentate isolando le parti significative contenenti le formule. È stata posta molta attenzione alla preparazione di un dataset rappresentativo, che contenesse un'ampia gamma di simboli e terminologia del dominio matematico. Terminata questa attività di preprocessing, le immagini sono state convertite in LaTeX mediante tecnologie di Computer Vision. Anche per questo modulo, stiamo effettuando delle analisi qualitative degli errori per capire in generale le performance del tool ed identificare eventuali problematiche sul riconoscimento di testi handwritten o di particolari simboli/caratteri, che permetteranno di decidere i nuovi test da portare avanti.

Al termine di queste attività di Speech-to-Text e Computer Vision, gli output ottenuti dovranno confluire in un unico flusso, che consentirà agli studenti di ripercorrere ordinatamente la lezione. Alle formule LaTeX viene associato un demarcatore temporale che consente di inserirle al punto opportuno della lezione.

2. SpeechMatE

Il riconoscimento vocale ha avuto negli ultimi anni un significativo progresso: sono disponibili applicazioni azionabili tramite comandi vocali, assistenti personali intelligenti basati su interfacce di dialogo, sistemi operativi con assistenti vocali. Tuttavia ci sono pochi tentativi di sviluppare soluzioni per il riconoscimento di formule matematiche, soprattutto in lingua italiana, anche se l'utilità di tali soluzioni sarebbe di grande impatto per un'alta percentuale della popolazione. Sicuramente questo è dovuto ad alcune difficoltà tra cui: la rappresentazione non lineare delle formule, l'ambiguità della lettura di formule matematiche e l'impatto di percentuali di errore nel riconoscimento.

In questo ambito il Laboratorio ha sviluppato il prototipo del software SpeechMatE, per la dettatura e la modifica di formule con riconoscimento della lingua italiana. Il software ha cinque principali componenti: speech recognizer, parser, LaTeX editor, LaTeX compiler, e PDF viewer. L'interazione con il software avviene con comandi vocali, dettatura di testo o formule. L'input vocale viene processato dallo speech recognizer che lo trascrive in lingua italiana. Il testo viene poi processato dal parser che converte la parte di matematica in un frammento di codice LaTeX, che viene successivamente passato all'editor LaTeX e trasformato in PDF. Il modello di interazione vocale utilizza tre ambienti attivati dalle relative parole chiave: testo, matematica e comandi (per i comandi di edit). Dopo una fase di confronto sono stati scelti, per il riconoscimento vocale, i servizi di Google Cloud e, come editor LaTeX, TeXStudio. Attualmente viene gestita una parte limitata di simboli e espressioni matematiche: l'estensione di queste funzionalità è attualmente in corso di sviluppo (Bernareggi et al, 2020).

Riferimenti bibliografici

Armano T., Capietto A., Ahmetovic D., Bernareggi C., Coriasco S., Ducci M., Magosso C., Mazzei A., Murru N., Sofia A. (2020), Accessibilità di contenuti digitali per le STEM: un problema aperto. Alcune soluzioni inclusive per l'accessibilità di formule e grafici, Mondo Digitale, vol. 89, pag. 1-12.

Bernareggi C., Ducci M., Mazzei A., Ahmetovic D., Armano T., Capietto A., Coriasco S., Murru N. (2020), SpeechMatE: A Speech-driven Maths Editor for Motor-Impaired People, ICCHP open access compendium "Future Perspectives of AT, eAccessibility and eInclusion".

Jurafsky, D., & Manning, C. (2012). Natural language processing. *Instructor*, 212(998), 3482.

Savy, R. (2005), "Specifiche per la trascrizione ortografica annotata dei testi" in *Italiano Parlato. Analisi di un dialogo*, Liguori, Napoli.

Autori

Amore Matteo matteo.amore@h-farm.com

Laureato in Linguistica presso l'Università di Pavia, tesi sulla semantica distribuzionale. È un linguista computazionale che si occupa di attività di estrazione di testo per progetti di Customer Experience, ossia Text Classification, Sentiment Analysis, Opinion Mining, Named Entities Recognition. Si occupa inoltre di: progettazione e valutazione di corpora di training; test di sistemi di Automatic Speech Recognition, valutazione di tecnologie vocali, annotazione di dati per progetti di ricerca.

Tiziana Armano tiziana.armano@unito.it

Tecnico della ricerca presso il Dipartimento di Matematica "G. Peano" dell'Università degli Studi di Torino. Si occupa di e-learning e tecnologie per la didattica. Ha fatto parte del gruppo di ricerca del progetto "Per una matematica accessibile e inclusiva" dal 2012 ed è membro del Laboratorio Polin. In questo ambito ha seguito in particolare lo sviluppo del pacchetto Accessibility e attualmente si occupa del progetto Matematica a voce.

Cristian Bernareggi cristian.bernareggi@unimi.it

Dottorato di ricerca in informatica nel 2006. Tra il 2004 e il 2009 ha condotto ricerca in ambito HCI nei progetti europei LAMBDA ed @Science. Dal 2007 è collaboratore tecnico scientifico in ambito informatico per l'Università di Milano. Tra il 2011 e il 2018 ha partecipato alla fondazione e alla crescita dello spin-off EveryWare Technologies per lo sviluppo di tecnologie assistive su dispositivi mobili. Dal 2018 collabora con il Laboratorio Polin sull'accessibilità alle STEM per persone con disabilità.

Anna Capietto anna.capietto@unito.it

Professore Ordinario di Analisi Matematica. Da 15 anni è Referente per le tematiche degli studenti con disabilità presso il Dipartimento di Matematica. Dal 2013 è responsabile del progetto "Per una matematica accessibile e inclusiva", le cui attività hanno condotto nel novembre 2018 all'istituzione del Laboratorio per la ricerca e la sperimentazione di nuove tecnologie assistive per le STEM "S.Polin". È presidente del Comitato Public Engagement dell'Università di Torino.

Sandro Coriasco sandro.coriasco@unito.it

Professore Associato di Analisi Matematica presso il Dipartimento di Matematica dell'Università di Torino. Si occupa di analisi microlocale e di analisi globale su varietà non compatte e su varietà con

singularità. È un componente del gruppo di ricerca sulle tecnologie assistive presso il Laboratorio "S. Polin". In tale ambito si occupa, in particolare, dello sviluppo del pacchetto LaTeX accessibility e, più in generale, dei temi collegati all'accessibilità dei contenuti matematici.

Roberta Crespan roberta.crespan@h-farm.com

Laureata in Mediazione Linguistica e Culturale presso l'Università di Padova. Ha approfondito le conoscenze nell'ambito della Traduzione audiovisiva e Accessibilità Audiovisiva. Dal 2019 lavora come linguista presso CELI Language Technology dove ha avuto modo di avvicinarsi al mondo della linguistica computazionale legata all'NLP. Svolge attività di test sui chatbot, di verifica dei flussi. Ha partecipato ad attività legate a progetti di ASR e si occupa di localizzazione e UX Writing di contenuti.

Mattia Ducci mattia.ducci@unito.it

Laurea magistrale in informatica all'Università degli Studi di Milano nel 2019. Durante e dopo l'università ha collaborato a vari progetti incentrati sul tema dell'accessibilità: da soluzioni su dispositivi mobili per persone con disabilità visive (Università degli Studi di Milano) a soluzioni per rendere più accessibili le materie scientifiche (Università degli Studi di Torino). Dal 2018 lavora come consulente informatico in un'azienda per sviluppo mobile iOS e cross-platform e full-stack.

Maria Luisa Gabrielli marialuisa.gabrielli@h-farm.com

Lavora come Data Scientist presso Celi Language Technology. Laureata in Matematica presso l'Università di Torino, si è specializzata in Analisi numerica ed in particolare modo si è avvicinata al mondo dell'elaborazione e del riconoscimento di immagini. Ha esperienza in Data Analysis, nell'applicazione di algoritmi di Machine Learning in ambito Forecasting e in tecniche di Computer Vision.

Mariafrancesca Guadalupi mariafrancesca.guadalupi@h-farm.com

Lavora come linguista computazionale dopo aver conseguito una laurea magistrale in Linguistica e Comunicazione Persuasiva, Tecnologia e Studi Cognitivi ed un Master di II Livello in Intelligenza Artificiale. Si occupa di progetti riguardanti: assistenti conversazionali, sentiment analysis ed opinion mining, text classification. Sta studiando per approfondire il mondo delle tecnologie assistive.

Alessandro Mazzei alessandro.mazzei@unito.it

Laurea in Fisica (indirizzo Cibernetico) presso l'Università Federico II di Napoli nel 2000. Dottore di Ricerca in Informatica presso l'Università degli Studi Torino nel 2005, dove è ricercatore presso il Dipartimento di Informatica. I suoi interessi di ricerca includono le ontologie giuridiche, la sintassi e la semantica computazionale, la generazione automatica del linguaggio, i sistemi di dialogo, il ragionamento automatico, le tecnologie assistive e il food computing. È autore di oltre 100 lavori scientifici peer reviewed.

Antonio Mazzei antonio.mazzei@h-farm.com

Si è laureato al Politecnico di Torino ed all'ENST di Parigi in ingegneria delle telecomunicazioni. Ha rivestito molti dei ruoli IT fino all'attuale posizione da Senior Manager. La sensibilità per le

peculiarità di ciascun ruolo, lo spingono a bilanciare le esigenze del team e valorizzarne le capacità al fine di raggiungere gli obiettivi di progetto.

Specializzato in sistemi di Enterprise Search ed integrazione di sistemi di NLP ed AI nei processi aziendali.

Adriano Sofia adriano.sofia@unito.it

Assegnista di ricerca presso il Laboratorio "S. Polin" dell'Università degli Studi di Torino dove si occupa di sonificazione e accessibilità. Conseguita la Laurea Magistrale in Fisica con Lode i suoi interessi scientifici riguardano due filoni: il primo incentrato sulle simulazioni di dinamica molecolare classica ed ab initio applicata alla fisica della materia e alla biochimica, e il secondo riguarda le tecnologie assistive e l'accessibilità.

Francesco Tarasconi francesco.tarasconi.h-farm.com

É Senior Data Scientist presso CELI Language Technology, ora centro di Intelligenza Artificiale di H-Farm Innovation. Laureato in Matematica presso l'Università di Torino, ha condotto un progetto di ricerca applicata sui Sistemi Complessi con il supporto di CELI e Fondazione CRT. Lavora da quasi dieci anni in progetti di Data Science, sia in ambito business che di ricerca; si interessa principalmente a Unsupervised Machine Learning, Forecasting e Deep Language Models.