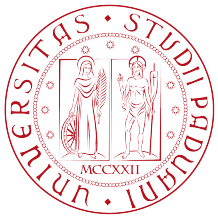

Università di Padova

Dipartimento di Matematica
Corso di Laurea Magistrale in Informatica

Corso di Intelligenza Artificiale

Marco Romanelli, [1106706]

30 marzo 2017

Capitolo 1: Introduzione

1.1 Fase iniziale

La fase iniziale del progetto consiste in una prima analisi delle principali piattaforme che offrono servizi di *Cognitive Computing*, considerando vantaggi e svantaggi di ognuna. Data la diversificazioni di proposte all'intero di ogni piattaforma, per l'analisi e la comparazione ci siamo focalizzati sul riconoscimento di immagini.

Procederemo nel seguente modo: nel Capitolo 2 verrà presentata una panoramica dei servivi presi in esame e seguirà poi nel Capitolo 3 l'analisi dettagliata per ogni servizio. Il Capitolo 4 contiene una sintesi sulle tariffe richieste per ogni servizio. Il Capitolo 5 riassume le conclusioni della squadra di lavoro. Il capitolo 6 concluderà con alcuni possibili sviluppi di questa analisi.

1.2 Snippet

In un'analisi è spesso necessario anche definire un possibile caso d'uso (obbietto) ed effettuare delle prove in relazione a tale contesto. Questo sia per approfondire l'analisi in sé e sia per poter comparare le diverse soluzioni offerte in un contesto reale (anche se limitato).

Si immagini, quindi, di dover analizzare degli scontrini fiscali con l'obbiettivo di informatizzare le informazioni contenutevi, come ad esempio il locale che ha emesso lo scontrino, le voci con i relativi prezzi, il giorno di emissione, eccetera. Questo perché, ad esempio, un'azienda potrebbe aver bisogno di un sistema che permetta l'analisi degli scontrini per stabilire se e in che misura attribuire dei rimborsi ai propri dipendenti.

Per ogni soluzione si effettueranno alcune prove in relazione al contesto appena descritto, analizzandone pregi e difetti, tenendo ovviamente in considerazione la natura limitata delle stesse.

Tutto il codice a cui si fa riferimento si può trovare nel repository online[Rom17].

Capitolo 2: Cognitive Computing

2.1 Introduzione

2.2 Servizi disponibili

Le maggiori piattaforme per il *Cognitive Computing* sono offerte da alcune fra le maggiori aziende nell'ambito informatico e tecnologico e sono:

- Microsoft Cognitive Services [Cor17a] (Microsoft Corporation),
- Watson Developer Cloud [IBM17b] (IBM: International Business Machines Corporation),
- Amazon Artificial Intelligence [Ama17] (Amazon.com, Inc)

- Google Cloud Platform (Google Inc.)

Capitolo 3: Analisi dei servizi

3.1 Microsoft Cognitive Services: Computer Vision API

3.1.1 Panoramica

Prerequisiti

- Credenziali per accedere al servizio (API key).
- Input: dati grezzi (stream application/octet) o url.
- Formati supportati: JPEG, PNG, GIF, BMP.
- Dimensione file massima: 4 MB.
- Dimensione immagine minima: 50x50 pixel.

Le API[Cor17b] sono molteplici, a seconda dello scopo finale dell'analisi visiva.

Tagging Le API ritornano un insieme di etichette (in formato JSON) che descrivono gli oggetti presenti nell'immagine, come oggetti, esseri viventi, azioni, paesaggi; per ogni etichetta viene anche fornito il livello di *confidence* (affidabilità). I tag non sono in alcun modo organizzati fra loro e non esiste nessun tipo di ereditarietà. Nel caso un tag sia ambiguo viene fornito in aggiunta un *hint* che ne spiega il contenuto. Al momento la sola lingua supportata è l'inglese.

Classificazione L'immagine viene classificata in categorie che seguono una tassonomia con ereditarietà di tipo padre-figlio. Questa tassonomia prevede 86 categorie[1] e classifica gli elementi visivi in modo più o meno specifico.

Identificazione del tipo E' possibile classificare l'immagine come in bianco o nero o a colori, se è un disegno o se è del tipo *clip-art*; in quest'ultimo caso viene fornito un livello di qualità dell'immagine, compreso fra 0 e 3.

Riconoscimento volti Riconosce i volti umani e restituisce la posizione (coordinate) di questi all'interno dell'immagine, come anche età e sesso della persona.

Contenuto personalizzato Ideato per raffinare la tassonomia a 86 categorie utilizzando informazioni specifiche sul dominio. Attualmente è supportato solamente il riconoscimento dei volti delle persone famose.

Generazione di descrizioni Genera una lista di frasi (in lingua inglese) che descrivono il contenuto dell'immagine, ordinate secondo un livello di affidabilità calcolato per ogni descrizione.

Estrazione colori Identifica i colori analizzandoli in tre contesti: di sfondo, in primo piano e d'insieme; i colori sono raggruppati in 12 colori predominanti. Classifica le immagini fra in bianco e nero e a colori.

[1] `https://www.microsoft.com/cognitive-services/en-us/Computer-Vision-API/documentation/Category-Taxonomy`

Riconoscimento contenuti non adatti ai minori Riconosce materiali pornografici e contenuti osé in generale. Può essere impostato un livello per il filtro.

Riconoscimento del testo (OCR) Rileva il testo presente nell'immagine e lo trasforma in un flusso di parole, ruota l'immagine se necessario per rendere il testo orizzontale e fornisce le coordinate per ogni parola. Al momento sono supportati 21 linguaggi, fra cui l'inglese, l'italiano, il francese, il tedesco e lo spagnolo.

L'accuratezza del riconoscimento dipende dalla qualità dell'immagine ed eventuali errori possono essere causati da immagini sfuocate, scrittura a mano, testo troppo piccolo, ecc.

Creazione anteprime Un'anteprima è una rappresentazione dell'immagine in scala ridotta. L'immagine viene prima analizzata e poi ritagliata secondo la "regione di interesse" (ROI); il rapporto dell'immagine (*aspect ratio*) può essere impostato secondo le proprie preferenze.

3.1.2 Tariffe

Due tipologie di piani:

- Gratuito: fino a 5000 chiamate al mese, massimo 20 chiamate al minuto;
- Standard: 0,015$ a chiamata, fino a 10 TPS.

3.1.3 Esecuzione

Prendendo in esame il caso d'uso descritto nell'introduzione, sono state sono state individuate due tipologie di operazioni che potrebbero risolvere il problema posto: estrazioni di caratteristiche visive (e classificazione) e l'OCR.

Estrazione *features* e classificazione Come si vede nell'immagine 3.1, il contenuto viene classificato come *text menu* con un livello di *confidence* dell'85, 5%. Vengono create anche le

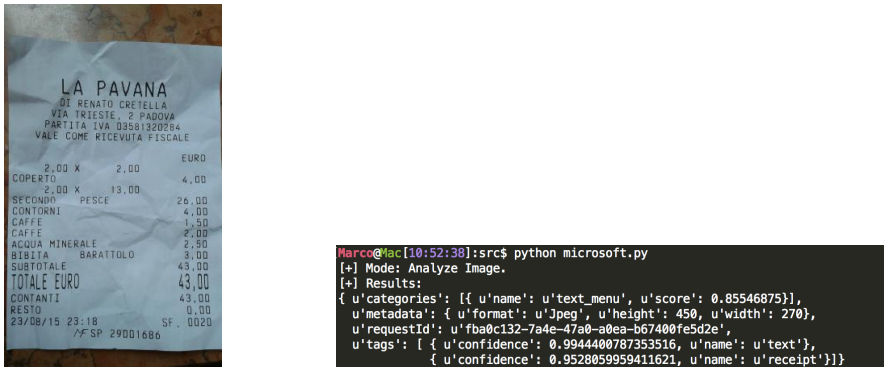

Figura 3.1: Estrazione delle caratteristiche.

etichette *text* e *receipt*, entrambe con un alto livello di affidabilità. Quindi l'algoritmo identifica correttamente il contenuto e il significato astratto dell'immagine: è uno scontrino, o comunque una lista di elementi testuali.

Tuttavia, seppur l'analisi sia corretta, non raggiunge il nostro obbiettivo di estrarre la lista di elementi dello scontrino.

Figura 3.2: Estrazione del testo.

OCR La seconda funziona che sembrerebbe essere adatta al nostro obbiettivo è il riconoscimento del testo. Dalla figura si vede come l'algoritmo sia riuscito a rilevare molte parole (ma non tutte) ma non i numeri (i prezzi); inoltre si nota come diverse parole siano state *tradotte* erroneamente, come capERTO invece che "coperto" o COMANI al posto di "contanti".

In conclusione non si ritiene il risultato sufficiente per lo scopo prefissato.

3.2 IBM Developer Cloud: Visual Recognition

3.2.1 Panoramica

Il servizio di Visual Recognition[IBM17a] utilizza tecniche e algoritmi di *deep learning* per identificare scene, oggetti, visi di persone nell'immagine che viene fornita come input al servizio. Permette, inoltre, la creazione e l'addestramento di un classificatore personalizzato per l'identificazione di elementi in base alle necessità dello sviluppatore.

Requisiti

- Credenziali per accedere al servizio (API key).
- input: dati grezzi o URL all'immagine.
- formati supportati: JPG, PNG.
- dimensioni minime: 224x224 pixel (consigliate).

Classificazione Per ogni immagine sottoposta a classificazione viene fornito in riposa una lista di coppie classe-punteggio per ogni classificatore selezionato. Il punteggio è compreso in un intervallo $0-1$, dove un valore maggiore indica una probabilità più alta che la classe descriva l'immagine; la soglia di default perché un valore sia ritornato da un classificatore è $0,5$. Le classi sono organizzate in categorie e sotto-categorie dove il livello più astratto comprende categorie quali animali, persone, cibo, sport, natura, eccetera.

Le lingue sopportate[2] nella risposta sono l'inglese, spagnolo, arabo o giapponese.

Riconoscimento dei volti Analizza i volti presenti l'immagine e ne deriva alcune informazioni, come età stimata, sesso o nome del personaggio famoso (nel caso ci sia). Anche in questo caso viene fornito un punteggio (nell'intervallo $0-1$) atto ad indicare una maggiore probabilità di correlazione.

[2] Al momento della stesura di questo documento.

Classificatore personalizzato Permette di creare un nuovo classificatore e di addestrarlo su un dato insieme di immagini. Queste sono inviate in un file compresso e devono comprendere o due immagini d'esempio positive o una positiva e una negativa. L'insieme contente le immagini d'esempio positive serve a creare le classi che definiscono il nuovo classificatore. Il complementare definisce invece quello che il classificatore *non* deve essere; le immagini d'esempio negative non devono contente i soggetti presenti nelle immagini positive.

Se, ad esempio, si volesse creare un classificatore "frutta" si potrebbe utilizzare un file compresso contente immagini di pere, uno contente immagini di mele e uno con immagini di banane. Per le immagini d'esempio negative si potrebbero utilizzare immagini di verdure.

Collezioni Questa funzione[3] permette di creare una nuova collezione, aggiungere immagini a questa e utilizzare la *Similarity Search* per cercare immagini simile all'interno della collezione.

Note per la privacy Per default, tutte le immagini e le informazioni inviate vengono salvate e utilizzate per migliorare il servizio. Per evitare questo è necessario impostare diversamente il parametro X-Watson-Learning-Opt-Out in ogni richiesta inviata.

3.2.2 Tariffe

Il piano gratuito prevede la possibilità di:

1. classificare 250 immagini al giorno,
2. addestrare un solo classificatore personalizzato con massimo 5000 immagini.

Il piano *standard* prevede:

1. per la classificazione: $0,002$ dollari a immagine,
2. per il riconoscimento volti: $0,004$ dollari a immagine,
3. per l'addestramento classificatore: $0,10$ dollari a immagine,
4. per la classificazione con classificatore personalizzato: $0,004$ dollari a immagine.

3.2.3 Esecuzione

3.3 Amazon Artificial Intelligence: Amazon Rekognition

3.3.1 Panoramica

Prerequisiti

- Credenziali per accedere al servizio.
- Input: dati grezzi.
- Formati supportati: JPEG, PNG.
- Dimensione file massima: 15 MB.
- Dimensione immagine minima: 80 pixel.

[3] Questa funzione è ancora in fase BETA

3.3.2 Tariffe

Il piano gratuito prevede, per i primi 12 mesi, di:

- analizzare 5000 immagini,
- memorizzare 1000 metadati facciali al mese.

Altrimenti:

- per il primo milione di immagini[4]: 1 dollaro ogni 1000 immagini[5];
- successivi 9 milioni di immagini: $0,80$ dollari ogni 1000 immagini;
- successivi 90 milioni di immagini: $0,60$ dollari ogni 1000 immagini;
- oltre i 100 milioni di immagini: $0,40$ dollari ogni 1000 immagini.

Inoltre utilizzando le API per il riconoscimento dei volti, il servizio memorizza ogni volta la rappresentazione vettoriale dei volti. Questo comporta dei costi pari a $0,01$ dollari per 1000 metadati memorizzati al mese.

3.4 Tabelle riassuntive

Tabella 3.1: Analisi delle funzionalità

Funzionalità	Microsoft Computer Vision	IBM Visual Recognition	Amazon Rekognition	Google ?
Riconoscimento oggetti / Tagging	✓	✓		
Classificazione	✓	✓		
Creazione di un classificatore		✓		
Riconoscimento colori	✓			
Riconoscim. tipo imm.	✓			
Riconoscimento volti	✓	✓		
Riconoscimento celebrità	✓	✓		
Generazione descrizioni	✓			
Riconoscimento nudità	✓			
Riconoscimento del testo (OCR)	✓			
Creazione anteprime	✓			
Ricerca di immagini		✓	solo volti	
Confronto fra immagini		✓		

Tabella 3.2: Analisi delle tariffe

Tipologia di piani	Microsoft Computer Vision	IBM Visual Recognition	Amazon Rekognition	Google ?
Gratuito [chiamate/mese]	5000	7500 immagini/mese[6] 1 classificatore addestrato con 5000 imm. max		
Standard [dollari/chiamata]	$0,015$	$0,002$ - $0,004$ (classificazione) $0,10$ a immagine (addestramento)	da $0,001$ a $0,0004$	

[4]Ogni API che accetta una o più messaggi di input conta come un'immagine elaborata.
[5]Al mese.

Bibliografia

[Ama17] Amazon.com, Inc. Amazon Artificial Intelligence. https://aws.amazon.com/it/amazon-ai/, 2017 (accessed March 30, 2017).

[Cor17a] Microsoft Corporation. Microsoft Cognitive Services. https://www.microsoft.com/cognitive-services/en-us/, 2017 (accessed March 22, 2017).

[Cor17b] Microsoft Corporation. Computer Vision API (Version 1.0) Documentation. https://www.microsoft.com/cognitive-services/en-us/Computer-Vision-API/documentation, 2017 (accessed March 30, 2017).

[IBM17a] IBM. IBM Visual Recognition API Reference. https://www.ibm.com/watson/developercloud/visual-recognition/api/v3/, 2017 (accessed March 30, 2017).

[IBM17b] IBM. IBM Watson Services. https://www.ibm.com/watson/developercloud/services-catalog.html, 2017 (accessed March 30, 2017).

[Rom17] Marco Romanelli. Artificial intelligence project repository. https://gitlab.com/mromanelli/ai-project, 2017.