Social data science. Anna Palumbo 473634

***DATA MINING E CRITICA LETTERARIA: QUELLO CHE I TESTI DICONO E NON DICONO.***

INTRODUZIONE.

Con l’espressione Data mining si intende una nuova scienza che prevede l’estrazione di informazioni da grandi masse di dati. Nel definirla nuova scienza riprendo l’idea centrale del seminario tenuto dal professor Dino Pedreschi il giorno 24 novembre 2011: il Data mining nasce come conseguenza diretta del vivere in una società informatizzata, dove ogni nostro spostamento lascia scie luminose, onde, puntini che è ora possibile rilevare attraverso nuove tecniche di raccolta e elaborazione. Facciamo un esempio: chiunque di noi digita ogni giorno una quantità non indifferente di parole su Google, che più che parole sono richieste, interrogativi, quasi che il motore di ricerca fosse una sibilla pronta a risponderci. Si va da informazioni di carattere generale a semplici dati, fino ad arrivare alle domande personali e da un milione di dollari: “Il mio ragazzo mi ha lasciata, cosa fare?” Google e gli altri motori di ricerca diventano così il nostro specchio, una proiezione del nostro dentro, delle nostre ossessioni, dei nostri interessi, il pozzo senza fondo dal quale speriamo di ricavare soluzioni, decisioni, suggerimenti. Ora, in che modo questo specchio delle nostre debolezze- o più semplicemente della nostra curiosità e del nostro anelito alla conoscenza- può diventare a sua volta, seguendo il meccanismo delle scatole cinesi, un serbatoio di conoscenza? Attraverso il Data mining appunto. Un esempio particolarmente interessante è, in questo senso, quello fatto dal professor Pedreschi su uno studio riguardante la felicità degli abitanti degli Stati Uniti, “calcolata” sulla base del tenore dei loro messaggini diffusi su Twitter nel corso della giornata. Studio sulla felicità: questo accostamento di parole così contrastanti tra loro in retorica si chiama ossimoro: come si può studiare scientificamente la felicità, sentimento soggettivo, sfuggente, momentaneo? Questa è la sfida del Data mining, scienza che cerca di bloccare, di fissare e di interpretare le enormi masse di informazioni implicite e involontarie che diffondiamo quotidianamente e, soprattutto, singolarmente ma insieme a altre centinaia di migliaia di persone sparse in giro per il mondo. Ciò che rende accattivante e promettente lo studio di grandi masse di dati è la possibilità di poter scendere sì nel particolare e porre domande assai specifiche e originali, dimostrando così dati alla mano, grazie a calcoli, algoritmi, operazioni spesso complicatissime, tendenze fino ad ora ritenute indimostrabili, ma anche di collocare il singolo dato insieme a tanti altri uguali, facendolo ascendere fino alle vette dell’universalità. Altra caratteristica del Data mining, oltre alla capacità di ricavare l’universale dal particolare senza rimpicciolire né l’uno né l’altro, è riuscire anche a mostrare il cambiamento nel tempo e nello spazio della stessa tendenza; l’evoluzione, in poche parole, di un processo. Gli utenti americani di Twitter sono più felici la sera e durante la pausa pranzo; le ore calde del lavoro d’ufficio sono quelle dove partono scariche di messaggini che attestano insoddisfazione e tedio. In California il malessere non esplode del tutto, e i messaggi mantengono per tutto l’arco della giornata una patina di soddisfazione e di ottimismo nel confronti della vita.

A ben vedere il mio breve resoconto dello studio su Twitter e la felicità negli Stati Uniti contiene già un tentativo da parte mia di interpretare i dati emersi, di spiegare ossia, sulla base di un bagaglio di conoscenze pregresse- e magari, anche viziate dallo stereotipo- le oscillazioni spazio-temporali dello stato d’animo degli utenti presi in esame. Dopo aver ravvisato la tendenza, quindi, diviene quasi automatico descriverla e motivarla. Ed è questo il nodo della questione, che affronterò nel prossimo paragrafo.

1.1

Quando il professor Pedreschi ha parlato della fase successiva alla raccolta dei dati ha utilizzato un’espressione molto bella: la Social data science non è una semplice classificazione di tendenze, ma piuttosto un insieme di «dati resi vivi». Dati, ossia, raccolti e interpretati da studiosi appartenenti ai più vari campi del sapere: antropologia, sociologia, economia. Sta a noi, insomma, farci dire qualcosa da queste masse di dati, inventarci una domanda e saper riconoscere la risposta, approfondirla, sviscerarla. La domanda è probabilmente il momento più importante del processo creativo che può scaturire con e grazie al Data mining. Più la domanda sarà pertinente ma non scontata, più sarà produttiva, ossia in grado di ricevere una risposta illuminante e stupefacente. Il dato di per sé è una massa inerte, ma già in questo caso i dati presi in esame sono strani, nuovi, fluidi e complessi, poiché riguardano i nostri gusti, i nostri desideri, i nostri movimenti, le nostre esigenze nuove e quelle vecchissime, sempre più precise, sempre più assurde anche, se vogliamo. Ci sarà bisogno quindi di interpreti intelligenti, sensibili, colti, moderni, capaci non solo di progettare algoritmi in grado di fare emergere dal mare magnum dei dati degli isolotti tutti da esplorare ma anche, appunto, di esplorarli con la dovuta attenzione e con quella curiosità che non si ferma al primo antro suggestivo ma arriva fino agli anfratti più reconditi. Questo tipo di approccio può vivificare tutte le scienze basate sull’integrazione tra classificazione e interpretazione, anche quelle insospettabili, e consentirci, perché no, di dire cose nuove su una qualsiasi grande mole di dati. Anche un libro è un insieme complesso di informazioni, e come tale può essere indagato con l’aiuto del Data mining. Il professor Pedreschi ha illustrato, nel corso del seminario, uno studio condotto insieme ad alcuni colleghi del CNR di Pisa (Amedeo Cappelli, Michele Coscia, Fosca Giannotti, Salvo Rinzivillo) in cui viene ricostruita la rete di rapporti sociali dell’Inferno di Dante. Dante e Virgilio vengono messi al centro di un complesso sistema di scambi di informazioni- e quindi, di conoscenza. Stabilire attraverso il Data mining con quante persone i due esploratori dell’al di là parlano, chi sono queste persone e quali temi prediligono è il punto di partenza da cui possono poi scaturire tante domande collaterali, sempre più precise e sempre più interessanti (es. I Guelfi e i Ghibellini compaiono seguendo *patterns* ricorrenti?).

Queste domande-pilota, pertinenti ma ancora un po’vaghe, vanno a rispondere alla domanda fondamentale che l’equipe di studiosi impegnata innanzitutto a sperimentare le potenzialità del Data mining si pone:

 Is this useful? The challenge is to make this

 framework able to discover new knoledge or to

 confirm literary studies.

1.2

Chi ama la letteratura e vorrebbe fare di questa passione il proprio mestiere, conosce l’esigenza di cercare approcci sempre nuovi per avvicinare i testi, per capire sempre di più e sempre meglio il loro valore e i percorsi mentali che hanno seguito gli autori per strutturarli proprio in quel modo, proprio in quell’ordine. Quando Italo Calvino nel suo *Perché leggere i classici* invitava il pigro uomo della strada a leggere e rileggere i classici, facendosi garante del loro carattere di pozzi senza fondo, non si riferisce soltanto ai libri, ma anche a chi li legge. Un libro significherà sempre qualcosa in più anche perché chi legge sarà sempre diverso e collocato in una dimensione spazio-temporale alternativa rispetto a quella degli altri. E avrà a disposizione sempre nuovi strumenti di indagine e saperi ancillari cui appoggiarsi. Il Data mining in questo senso potrebbe essere un ausilio importante per fare critica letteraria per almeno due ragioni: la prima è che grazie al Data mining si possono porre domande complicate e originali, alle quali prima magari si faticava a rispondere perché richiedevano troppo tempo, sforzi e una intricata serie di confronti e raffronti nei quali si tendeva a perdersi. La seconda è che una stessa domanda può essere posta per più di un’opera e per più di un autore, anche lontanissimi per spazio e per tempo. Questo tipo di indagini, già caratteristiche dei più moderni approcci allo studio della letteratura, basati sull’intertestualità, sul confronto tra due o più autori, due o più libri, saranno velocizzate e agevolate molto dal Data mining. Gli studiosi, così, confortati dalla maggiore praticità del metodo, potranno lanciarsi in quesiti sempre più creativi e significativi, e impegnarsi a interpretarne le risposte alla luce dei loro saperi storici, letterari, artistici (è risaputo infatti che il buon critico letterario dovrebbe essere già di suo una persona con un bagaglio di saperi multiforme). Soprattutto però, il Data mining potrebbe aiutare i critici a formulare delle suggestive ipotesi e a dare vita a significativi confronti che avrebbero come fulcro la storia dell’evoluzione dell’uomo e del suo pensiero e dei suoi costumi- e quindi della sua letteratura- con lo scorrere dei secoli. Proviamo a immaginare, ad esempio, una possibile estensione dello studio di Pedreschi e colleghi o meglio, una complicazione del modello base, incentrato sull’estrazione di dati da un solo testo. Si potrebbe scegliere un grande classico rappresentativo di ogni secolo della letteratura italiana, partendo proprio dalla Divina Commedia. Per il novecento, per esempio, si potrebbero scegliere *Uno, nessuno, centomila* di Luigi Pirandello o gli *Indifferenti* di Alberto Moravia. Rammentando rapidamente il primo, apparirà subito evidente che il numero degli interlocutori del protagonista, Vitangelo Moscarda, è minimo rispetto alla corte dei miracoli al cospetto di Dante; gli argomenti futili; la capacità di ascolto e di interazione empatica tra Vitangelo e sua moglie, i colleghi e i giudici, è ridotta quasi a zero, chiusi come sono, tutti, nel proprio mondo; intenti a appiccicare etichette prestampate alla schiena dell’altro di turno. Per quanto riguarda gli *Indifferenti*, l’aspetto più produttivo del confronto con l’analisi della *Divina Commedia* fatta dal team del CNR è quello che riguarda gli argomenti trattati nel corso degli scambi di battute tra i personaggi. Dante, pressato da un tempo indefinito ma presentito come breve, ma soprattutto dall’ansia di chiarirsi i dubbi di una vita facendoli sciogliere da chi per forza di cose ormai sa, non fa domande inutili, non perde tempo in convenevoli o in formule standard. I suoi interlocutori, ancora mezzo coinvolti nelle faccende terrene e ossessionati dalla propria storia personale, gli rispondono altrettanto pienamente e onestamente. Gli argomenti degli incontri, anche di quelli fuggevoli- basti pensare al canto V, con una Francesca in pausa per miracolo e per poco dal suo folle girare in vortice nel vento- tutti importanti, vitali. L’amore, la politica, l’etica, la morte, il futuro, la storia, l’amicizia, tutto viene tirato in ballo direttamente e definito con tutta la conoscenza e la modernità possibili per l’epoca. La verità ultima è detta, i dubbi fugati, la morale stabilita una volta per tutte, seppur con sofferenza e empatia verso alcuni di coloro che non l’hanno rispettata. Il confronto con gli *Indifferenti*, grande classico del ventesimo secolo, è sconcertante: tanti personaggi che si aggirano tra salotti e camere da letto parlottando di feste o pettegolezzi o stando muti a osservare annoiati o a sorbire tazze di caffè; familiari che si rivolgono la parola solo per scopi diretti quali dare e ricevere danaro; uomini e donne che si dicono lo stretto necessario per comunicarsi come ben presto si potrà passare dalle parole ai fatti. Non si crede, probabilmente, né al potere della parola né a quello dei fatti. Manca un centro, una coscienza forte e debole al tempo stesso che abbia l’arroganza di chiedere e l’umiltà di ascoltare le risposte.

CONCLUSIONI.

Questa fantasticheria, imprecisa e improvvisata come può esserlo il lavoro di una persona che non possiede gli strumenti tecnici in grado di estrarre in modo ordinato e *simmetrico* delle informazioni da un insieme complesso come può esserlo un libro, mi serve per concludere la mia relazione dicendo che credo all’utilità del Data mining nello studio della letteratura, qualora esso sia messo a servizio della dimostrazione di intuizioni fondate e piene di senso esplicito e implicito. E soprattutto, direttamente modellate sul testo preso in esame. Qualora poi i testi fossero più di uno, anche in questo caso si dovrebbe trattare di un insieme ragionato, ossia di libri legati tra loro da qualche nesso forte. Ogni passo dell’elaborazione del criterio di selezione/estrazione dei dati prima e di interpretazione poi dovrebbe essere progettato con cura, fantasia e conoscenza approfondita della materia in modo tale che la Data mining science aiuti gli studiosi a fare emergere quei fili ancora invisibili ma esistenti che legano i libri e gli autori attraverso il tempo e lo spazio: approccio, questo, quanto mai moderno e quanto mai commovente e fondato, esattamente come il Data mining stesso, sul presupposto che oggi come mai prima d’ora abbiamo la possibilità di collegare realtà lontanissime ma misteriosamente comunicanti. Solo così si può arrivare a raggiungere «non-trivial analytical risults», per dirla con le parole degli studiosi del CNR. Risultati su larga scala che, oltre ad essere non banali, sono anche vertiginosi come un viaggio con la macchina del tempo o con il teletrasporto.

BIBLIOGRAFIA.

<http://www.spss.it/solutions/datamine/whatsdm.htm>

<http://daily.wired.it/news/internet/2012/04/16/data-mining-ecco-perche-i-tuoi-dati-personali-fanno-cosi-gola-23456.html>

Amedeo Cappelli, Michele Coscia, Fosca Giannotti, Dino Pedreschi, Salvo Rinzivillo, *The social network of Dante’s Inferno. Leonardo*, giugno 2011, Vol. 44, No. 3.