

Riassunto Tesi di Dottorato di Giacomo Scornavacca

14 novembre 2019

Questa tesi di dottorato si occupa della progettazione ed analisi di protocolli distribuiti per diversi problemi fondamentali del calcolo distribuito. I principali problemi trattati sono il Consensus (e molte sue varianti), Broadcast e Community Detection. I modelli di comunicazione studiati nella tesi corrispondono a modelli di interazione casuale tra agenti con capacità di comunicazione limitata. La tesi è fortemente incentrata sull'analisi dei processi stocastici generati dall'utilizzo di semplici protocolli distribuiti, utilizzando disparate tecniche e strumenti matematici: analysis of discrete dynamical algorithms, message-passing / label propagation algorithms, coalitions in game theory, noisy channels in information theory.

La tesi è divisa in 7 capitoli. Di questi, i primi due capitoli forniscono una introduzione generale e il necessario formalismo per la stesura dei risultati della tesi. I successivi cinque capitoli raccolgono risultati tra loro altamente complementari: ognuno di essi presenta una diversa sfaccettatura di uno dei problemi, dei protocolli o dei modelli di comunicazione considerati in questo lavoro. Ogni capitolo è composto da una discussione riguardante il particolare aspetto dell'algoritmica distribuita considerato, da un riassunto sullo stato dell'arte e da una presentazione dei risultati teorici ottenuti. Alcune prove o bozze di prova sono fornite all'interno dei capitoli, tuttavia allo scopo di una migliore leggibilità la maggior parte dei dettagli tecnici sono spostati nelle appendici.

Il capitolo 1 introduce problemi e modelli e presenta i risultati ottenuti. Per tutti i problemi studiati, si fornisce una duplice motivazione: la prima derivante dall'interesse nell'analisi di certi protocolli da parte della comunità teorica di computer science e la seconda da un interesse scientifico più ampio. Si pone infatti un'enfasi particolare su aspetti legati alla Computational and Evolutionary Biology, evidenziando come una comprensione teorica di semplici protocolli randomizzati possa migliorare la nostra comprensione di alcuni meccanismi evolutivi (queste considerazioni sono apparse sotto forma di short communication a BDA'18) e su alcune operazioni che le cellule si ritrovano a compiere per risolvere alcuni problemi inerentemente distribuiti.

Il capitolo 2 fornisce notazione e definizioni comuni che vengono successivamente utilizzate nel resto della tesi e mette in evidenza le differenze tra i diversi modelli di comunicazioni utilizzati.

Nel capitolo 3, si considera un processo distribuito noto come la Undecided-State Dynamics, definito su insieme di agenti dotati di una opinione che interagiscono in maniera casuale su un grafo completo. Questo semplice protocollo distribuito è stato presentato da Angluin et al. (Dist. Comp. 2008) come soluzione rapida per il problema dell'Approximate Majority Consensus in un modello di comunicazione asincrono. Questo processo può anche essere visto come un esempio del paradigma del "power of two choices". Viene mostrato come il processo sia estremamente sensibile al modello di comunicazione sottostante, almeno su scale di tempo molto brevi e molto lunghe, e in questo capitolo si dimostra formalmente che l'Undecided-State Dynamics risolve rapidamente il problema dell'Approximate Majority Consensus in presenza di uno scheduler sincrono (parallelo) PULL/GOSSIP. La prova di questo risultato richiede un utilizzo originale di argomenti di concentrazioni probabilistica rispetto alle tecniche utilizzate nel lavoro di Angluin. Inoltre, nel dimostrare un upper bound al problema del Consenso, viene stabilita una elegante e generale connessione tra il problema del Consenso e una variante del Broadcast, mostrando che in modelli di comunicazione noiseless i due problemi sono computazionalmente equivalenti. I risultati di questo capitolo sono apparsi in MFCS'18 e ITCS'20.

Nel capitolo 4 viene affrontato il problema della Community Detection e viene presentata una classe di algoritmi conosciuta come Label Propagation Algorithms. In questo caso il grafo di interazione studiato non è più completo, ma ha una struttura a cluster basata su una bipartizione dei nodi, con una maggiore densità di archi all'interno dei due cluster rispetto al taglio fra le due componenti. Se si inizializzano randomicamente i nodi del grafo con un valore binario, si mostra che la Two-Choice Dynamics applicata a tali valori ha una buona probabilità di entrare in una fase metastabile durante la quale i valori riflettono la comunità di appartenenza del nodo. Questo problema, in questo scenario, può essere considerato difficile per due motivi: per prima cosa, il protocollo utilizzato è costituito da una semplice regola di update che utilizza una quantità di informazione limitata. Questo significa che il protocollo è molto sensibile al random noise generato dalle interazioni casuali. Inoltre, i risultati ottenuti si applicano a grafi deterministici con certe proprietà di espansione, invece che su grafi random. Le tecniche utilizzate si basano principalmente su strumenti di probabilità e sull'algebra spettrale dei grafi. I risultati di questo capitolo sono apparsi in AAAI'19 mentre un lavoro propedeutico a questi risultati è apparso a AAMAS'18.

Il capitolo 5 considera il problema del Consensus, dove questa volta l'obiettivo è ottenere una forma di voto proporzionale in cui la probabilità che il sistema converga ad una certa opinione deve essere proporzionale al numero di agenti che supportano quella opinione nella configurazione iniziale. In questo modello

di comunicazione gli agenti hanno identificativi unici che non possono essere falsificati e che accompagnano tutti i messaggi. La comunicazione è sicura e senza errori, tuttavia un sottoinsieme di agenti può smettere di funzionare. Inoltre, gli agenti possono tentare di deviare dal protocollo fornito (profilo strategico), e l'obiettivo è quello di garantire che nessuna coalizione di agenti (fino a dimensioni leggermente sublineari) sia incentivata a deviare da esso. Qui il concetto di incentivo viene ereditato dalla game theory, e l'utilità di un agente viene massimizzata quando l'opinione del nodo vince, mentre viene minimizzata se non viene raggiunto alcun consenso. I risultati di questo capitolo sono stati pubblicati in un articolo apparso a IPDPS'17.

Gli ultimi due capitoli della tesi si presentano Upper e Lower Bounds per i problemi di Consensus e Broadcast in diversi modelli di comunicazione noisy (con una certa attenzione al modello PULL/GOSSIP). E' già noto che il problema del Broadcast è particolarmente difficile da risolvere in presenza di una comunicazione noisy visto che quanto più lunga è la distanza tra sorgente e l'agente che riceve l'informazione, tanto maggiore è la probabilità che il messaggio sia totalmente corrotto. In questo capitolo viene mostrato che in questo contesto il problema del Consensus è esponenzialmente più facile da risolvere del Broadcast, ed unendo questo risultato a quello presentato nel capitolo 3 si evince che la relazione che intercorre tra Consensus e Broadcast dipende dal modello di comunicazione considerato. I risultati di questo capitolo sono apparsi a ITCS'20.