

Ansökan om inrättande av masterprogram i maskininlärning

Denna ansökan behandlar inrättandet av ett masterprogram i området maskininlärning (eng. computational learning). Det kommer att ges på engelska varför vi ger bakgrund och syfte på engelska och därefter beskriver programmets struktur och målsättning på svenska

Background

The design of algorithms in computer science and engineering is traditionally based on mathematical modeling and analysis of the specific problem to be solved involving physical, geometrical etc. constraints that affect its solution. From this modeling, algorithms can be designed and optimised. This analytical approach to algorithm design is still valid in many domains and should be applied whenever problems can be understood and modeled sufficiently. For problems involving large scale data sets such as those encountered in artificial intelligence, problem analysis and modeling becomes extremely complicated. This has led to the wider use of empirical methods for algorithm design based on the consideration of exemplars of inputs and desired outputs to the algorithm. Algorithms for mapping input and output data are designed in order optimally realise the observed mappings. Algorithms are therefore “learnt” from the empirical input and output data. This data is known as training data for the algorithm and the whole process is called Computational Learning or Machine Learning.

Computational learning has come to dominate algorithmic construction in artificial intelligence fields such as text, speech and image understanding. It is also a major tool for the programming of robots for grasping and navigation. Due to the enormous empirical data sets generated in e.g. molecular biology, computational learning has also been adopted as a major tool for understanding biological processes. Due to its close connection with the way biological systems acquire skills and performance, computational learning and its relation to neuroscience is a very active field of research.

It can be expected that computational learning methodology will become increasingly important in any field involving large scale data volumes including economics, medicine, geoscience etc. It will be relevant in any domain where understanding of a system or extraction of information from empirical data is required. Automatically finding structures in large sets of data is commonly referred to as data mining or knowledge discovery. The efficient organization of large repositories of empirical data and algorithms for its access in the form of search engines is therefore a very relevant field for computational learning

Purpose of the program

It is the purpose of our proposal to organize various already existing courses relating to the field of computational learning or any of its application domains into a common master program in order to achieve maximal efficiency and synergy of education. The broad interest that exist among students, both nationally and internationally for education in artificial intelligence and its techniques will find a common ground in this program. The ambition is to become an internationally leading center for education in computational learning. A major goal of the education will be to prepare the students for research either in academics or in industry. As more and more fields in technology and society will need to analyse large sets of empirical data techniques of computational learning will become common tools in almost any professional activity concerned with decisions based on data analysis. We can therefore expect the master program in computational learning to be very relevant as an extension of bachelor programs according to the Bologna model

1 Generell information

1.1 Programnamn på svenska och engelska

Engelska: Computational Learning
Svenska: Maskininläring

1.2 Programtyp

Masterprogram 120 hp

1.3 Planerat studerandeantal, initialt och på sikt

Programmets obligatoriska kurser Artificiell Intelligens, Bildanalys och Datorseende samt Maskininläring har för närvarande 60 - 70 studenter. Detta kan tas som en utgångspunkt för en uppskattning av programmets initiala studerandeantal till c:a 30. Masterprogrammet "Systems Control and Robotics" på EES har c:a 20 internationella studenter. I framtiden kan man räkna med att Maskininlärningsprogrammet får ett motsvarande studentantal och därmed komma upp i runt 50 studenter.

1.4 Studerandemålgrupp

Studenter vid KTH med en kandidatexamen inom datalogi, tillämpad matematik fysik eller elektronik. Internationella studenter med motsvarande bakgrund.

1.5 Behörighetskrav

Allmänna regler för tillträde till masterprogram gäller. Dessutom krävs att studenterna har kunskaper i matematik på den nivå som motsvaras av kurserna: SF1906 Matematisk Statistik grundkurs , SF1600 Differential- och integralkalkyl , SF1604 Linjär algebra , DD1340 Introduktion till datalogi eller motsvarande t.ex. DD1310, DD1311, DD1312, DD1313 (Programmeringsteknik) och någon av DD1320, DD1321 (Tillämpad programmering och datalogi)

1.6 Studief orm

Programmet ges i form av heltidsstudier som bedrivs på plats på KTH.

1.7 Undervisningsspråk

Engelska (valbara kurser på svenska)

1.8 Lärarkapacitet

Den grundläggande lärarkapaciteten utgörs av 4 lärare i datalogi, speciellt datorseende och robotik, samt flera lärare från övriga avdelningar inom skolan såsom teoretisk datalogi, beräkningsbiologi

1.9 Arbetsmarknadens behov samt omvärldsanalys

Efterfrågan på tekniker och andra med praktisk och teoretisk erfarenhet av maskininläring kommer att öka allt eftersom tillämpningar av artificiell intelligens inom industri och andra samhällssektorer blir allt vanligare. Inte bara traditionella teknikområden utan även de sektorer där bearbetning av stora datavolymer är väsentligt kommer att utnyttja sig av metodik från maskininläring. Detta kommer att omfatta områden som medicin, ekonomi, geovetenskaper etc. Inom forskning baserad på behandling av sensordata kommer maskininläring att fortsätta att dominera som huvudsakligt verktyg. Detta gäller så väl akademisk som industriellt tillämpad forskning.

Liknande program som det föreslagna existerar redan vid ett flertal Europeiska universitet. Bland dessa återfinns:

Program	Universitet
Masters in Machine Learning	University College London.
Master's in Artificial Intelligence	University of Amsterdam.
Master of Mathematical Modelling and Computation	Technical University of Denmark.
Master Artificial Intelligence	Katholieke Universitet Leuven.
Neuroscience and Cognition master programme	Universiteit Utrecht.

Intresset för metodik baserad på maskininläring har ökat kraftigt de senaste 5-10 åren och om man beaktar den växande aktiviteten på forskningssidan i form av ökat antal publikationer och konferenser så kan detta intresse förväntas öka också i framtiden. Det är sannolikt att denna typ av masterprogram kommer att inrättas i allt större omfattning vid olika internationella universitet i den närmaste framtiden.

1.10 Övergripande program mål

Masterprogrammet i maskininläring ska ge teoretiska insikter såväl som praktiska färdigheter att delta i och leda arbete inom det datatekniska såväl som andra områden där extraktion av information från och bearbetning av stora datamängder är ett väsentligt inslag. Genom att använda relevanta informationskällor ska den studerande kunna behålla och utveckla denna förmåga under en yrkeskarriär, kunna följa diskussionen om områdets utveckling i samhället och själv kunna bidra till den. Utbildningen är förberedande för forskarutbildning.

Utbildningen ska ge den studerande

- fördjupad kunskap inom datateknik speciellt statistisk bearbetning av data
- god analytisk problemlösningsförmåga
- förmåga att självständigt definiera och lösa problem i anknytning till bearbetning och strukturering av stora datamängder
- förutsättningar för arbete i internationella och mångdisciplinära projektgrupper som innehåller tekniker och icke-tekniker. Hit hör färdigheter i muntlig och skriftlig framställning på svenska och engelska.

Utbildningen ska dessutom ge den studerande möjlighet till specialisering med syfte att

- ge fördjupade kunskaper inom ett eller flera av områdena datorseende, robotik, tal, språkteknologi, artificiell intelligens, informationssökning och neuroinformatik.
- ge fördjupade kunskaper om programvarukonstruktion i ett tekniskt utvecklingsområde innefattande bearbetning av stora empiriska datamängder

- använda kunskap hur man konstruerar algoritmer för klassificering av data baserade på optimeringsmetoder
- ge kunskaper om konstruktion av ett fullständigt system för sensorbaserad igenkänning av innehåll i text, tal eller bilder
- ge kunskaper för konstruktion av algoritmer för automatisk sökning av och upptäckt av information i stora datamängder t.ex. internet

Utbildningen kan dessutom ge den studerande möjlighet till studier med syfte att

- ge fördjupade kunskaper i matematik.
- ge kunskap i allmän forskningsmetodik och vetenskapsteori
- ge fördjupade kunskaper inom andra ämnesområden där KTH bedriver forskning och utbildning.

Härutöver gäller de liknande mål som definieras i högskolelagen och högskoleförordningen.

1.11 Kurslista

Programmet innehåller en obligatorisk del samt två specialiseringar som avspeglar huvudinriktningarna klassificering av sensordata samt bearbetning av och informationssökning i stora datamängder. Gemensamma obligatoriska kurser och kurser på specialnivå kan läsas parallellt.

Vi har för avsikt att ge specialiseringen mot inlärnin för informationssökning, men på grund av bland annat bemanningsosäkerhet kan vi inte garantera att det blir precis de angivna kurserna och att alla kurser i specialiseringen kommer att ges på engelska.

Gemensamt obligatoriska kurser

Course code	Course name	Credits	Period
DA2205	Research - Theory, Method, Practice	6.0 hp	p1-2
DD2380	Artificial Intelligence	6.0 hp	p1
DD2431	Machine Learning	6.0 hp	p2
DD2423	Image Analysis and Computer Vision	7.5 hp	p3

Table 1: Gemensamt Obligatoriska (25.5 hp)

Gemensamt valbara kurser Två till tre kurser väljes fritt vid KTH.

Specialisering: Perception och kognition, Kurser (45 hp)

Course code	Course name	Credits	Period
DD2401	Neuroscience	7.5 hp	p4
DD2427	Image Based Recognition and Classification	6.0 hp	p4
DD2428	Geometric Computing and Visualization	6.0 hp	p1
DD2432	Artificial Neural Networks and Other Learning Systems	6.0 hp	p3

Table 2: Obligatoriska (25.5 hp)

Course code	Course name	Credits	Period
DD2257	Visualization	6.0 hp	p4
DD2426	Robotics and Autonomous Systems	7.5 hp	p2
EL2320	Applied Estimation	7.5 hp	p3
DD2447	Statistical Methods in Applied Computer Science	6.0 hp	p3-4
DT2112	Speech Technology	7.5 hp	p3
DT2118	Speech and Speaker Recognition	7.5 hp	p4
DT2140	Multimodal Interaction and Interfaces	7.5 hp	p4
SF1811	Optimization	6.0 hp	p3
SF2945	Time Series Analysis	6.0 hp	p2
SF2950	Applied Mathematical Statistics	7.5 hp	p3
SF2960	Statistical Theory	6.0 hp	p2

Table 3: Valbara (åtminstone 3 av dessa)

Specialisering: Inlärnin g för informationssökning , Kurser (45 hp)

Course code	Course name	Credits	Period
DD2447	Statistical Methods in Applied Computer Science	6.0 hp	p3-4
DD2471	Modern Database Systems and Their Applications	7.5 hp	p4
DDXXXX	Information Retrieval	9.0 hp	new

Table 4: Obligatoriska (23.5 hp)

Course code	Course name	Credits	Period
DD2469	Database Theory	6.0 hp	p3
DH2418	Språkteknologi (Svenska)	6.0 hp	p1
DD1334	Databasteknik (Svenska)	6.0 hp	p3
DD2257	Visualization	6.0 hp	p4
DD2395	Computer Security	6.0 hp	p1
DD2427	Image Based Recognition and Classification	6.0 hp	p4
DD2432	Artificial Neural Networks and Other Learning Systems	6.0 hp	p3
SF2940	Probability Theory	7.5 hp	p1
SF2950	Applied Mathematical Statistics	7.5 hp	p3
SF2960	Statistical Theory	6.0 hp	p2

Table 5: Valbara (åtminstone 3 av dessa)

1.12 Uppfyllande av krav i examensordning

Kraven för masterexamen om 90 hp kurser på avancerad nivå varav 60 hp centrala för huvudområdet uppfylls genom de 30hp obligatoriska kurserna tillsammans med de 30hp fördjupande kurserna och 30hp examensarbete. Kraven för civilingenjörsexamen i datateknik uppfylls av den som läst tre första åren på datateknikprogrammet och uppfyller kraven för denna masterexamen. Över 45 hp matnatkurser och utöver detta över 100 hp centrala för teknikområdet ingår obligatorisk under de tre första åren. Tillsammans med minst 84 hp centrala kurser i masterprogrammet (alla ovanstående 90 hp utom Vetenskapsteori) ger det minst 180 hp

centrala kurser. Övriga krav är trivialt uppfyllda. Dessutom är målen för civilingenjörsexamen i datateknik uppfyllda.

1.13 Relation till övriga KTH-utbildningar

Helt valfria kurser kan läsas från hela KTH. Masterprogrammet i “Systems Control and Robotics” på EES innehåller kurser relaterade till robotik som är relevanta för detta program. Till skillnad från det föreslagna programmet i Maskininläring har detta program en fokusering mot metoder inom reglerteknik och signalbehandling för autonoma system.

Masterprogrammet i Systembiologi på CSC innehåller delar av neurovetenskap som behandlar inläring i biologiska system. En viss del av metodiken i beräkningsbiologi baseras på maskininläring. Den grundläggande skillnaden mellan programmen i Maskininläring och Systembiologi ligger i att biologiprogrammet är fokuserat mot biologiska vetenskaper och problem medan maskininläring är i huvudsak teknikorierat.

1.14 Relation till skolans övriga verksamhet

En handfull av kurserna är gemensamma med masterprogrammen i datalogi och MDI. Helt valfria kurser kan läsas från hela skolan.

1.15 Skiss över studerandeutbyte

Via ERASMUS programmet kommer vi i kontakt med internationella studenter som är kandidater till detta masterprogram

2 Masterprogram specifik information

2.1 Uppgift om huvudområde

Huvudområdet är Informationsteknik.

2.2 Koppling till andra program på avancerad nivå

Vi kommer att samordna kursutvecklingen och kursplaneringen med i första hand masterprogrammen i Datalogi, Computational and systems biology och “Systems control and robotics. Vi samläser och utvecklar gärna kurser med andra program på avancerad nivå .

2.3 Kurs i vetenskapsteori och forskningsmetodik

Kursmomentet kommer att erbjudas i form av en delvis nyutvecklad kurs om 6 hp i kombination med inslag i övriga obligatoriska kurser. Forskningsmetodik behandlas även i AI-kursen.

2.4 Kurser som även ges som doktorandkurs

Vetenskapsteori, AI, Bildanalys och datorseende och flertalet fördjupningskurser kan också läsas som doktorandkurser.

2.5 Tillgodoräknande vid fortsatta forskarstudier

60 hp av kurserna i en masterexamen kan tillgodoräknas i forskarutbildningen beroende på forskningens inriktning.

2.6 Programmets forskningsanknytning

Programmet är huvudsakligen knutet till skolans avdelning för datorseende och robotik (CVAP) som för närvarande bedriver forskning inom automatisk igenkänning, greppande och manipulering av objekt, autonom navigering etc. Programmets olika specialiseringar förutsätter samarbete och kursutbyte med andra avdelningar inom skolan (huvudsakligen Teoretisk datalogi och Beräkningsbiologi)

2.7 Urvalsprinciper

Om antalet behöriga sökande skulle överskrida antalet platser kommer urval att ske enligt 5 kategorier i DIAK (Bedömning av universitet, Betyg, Meritämne, Relevant arbetslivserfarenhet, Referenser) med ”Thesis proposal” samt ”Motivering till studier” strukna. Om studenter inte kan särskiljas enligt detta kriterium kommer studenter från underrepresenterat kön att lämnas företräde.